<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Typology to Improve Undergraduate Education</td>
<td>1</td>
</tr>
<tr>
<td>Soichiro Aihara</td>
<td></td>
</tr>
<tr>
<td>Comment Data Mining for Student Grade Prediction Considering Differences in Data for Two Classes</td>
<td>12</td>
</tr>
<tr>
<td>Shaymaa E. Sorour, Tsunenori Mine, Kazumasa Goda and Sachio Hirokawa</td>
<td></td>
</tr>
<tr>
<td>Detection Methods for Misplacement of Synonyms in the Japanese WorkNet</td>
<td>26</td>
</tr>
<tr>
<td>Takuya Hirao, Takahiko Suzuki, Koki Miyata and Sachio Hirokawa</td>
<td></td>
</tr>
<tr>
<td>Development of a Digital Signage and MR Contents to Promote Hand Hygiene</td>
<td>36</td>
</tr>
<tr>
<td>Akinori Kanazawa, Toshiko Asai, Akinori Minazuki and Hidehiko Hayashi</td>
<td></td>
</tr>
<tr>
<td>Online Classroom Feedback System and Their Implementation in a Foreign Language Presentation Course in Japan</td>
<td>46</td>
</tr>
<tr>
<td>Yuichi Ono, Manabu Ishihara, Sachio Hirokawa and Mitsuo Yamashiro</td>
<td></td>
</tr>
</tbody>
</table>
Student Typology to Improve Undergraduate Education

Soichiro Aihara¹
Aichi University of Education, Japan

Abstract
In this study, first I overview the evolution of student typology in higher education studies in Japan and America. Second, from JFS2013 that is a part of the JCIRP student survey, I provide two student typologies: Kaneko typology and Astin typology. The JFS has potential to construct student typologies which aim at improving undergraduate education. Student typologies have potentials to check the admission policy, curriculum policy, and degree policy of higher education institutions, and guide them toward better education practices. Third, as a case study, depending on the comprehensive I-E-O model, I execute the determinants analysis toward the alienation type student. The result indicates that higher education institutions can improve their education practices through emphasis on the student involvements, especially psychological aspect of student engagement.

Keywords: Student Survey; JCIRP; Clark=Trow typology; Astin typology; Kaneko typology; Alienation type student

1. Introduction: Studies of Student Typology

In the higher education studies of Japan, student typology named Clark=Trow typology (1966) is also the classic model which portrays student profile through their subculture [4]. The Clark=Trow typology consists of two axes: involved with ideas and identify with their college, and has four type of student subculture: Academic, Collegiate, Nonconformist, and Vocational. They assume student subculture as the intermediates that convey the affection of social structure to student learning outcomes. These student subculture from Clark=Trow typology describe the situation of the campus at that time and provide suggestions to higher education practices. Their typology, however, is constructed by conceptually. Then the task of following researchers is to verify this typology empirically. Pascarella and Terenzini (1977), for example, verified this typology by separation students by gender [18]. Table 1 shows the main researches in America from the point of evolution of student typology. In Japan, Clark=Trow typology tested by Maruyama (1983) on Japanese college students, and he indicated effectiveness of Charter theory [13]. That means where higher education institutions they attend is more important for Japanese students than what is their social origin.

Furthermore, Takeuchi (1999) investigate the determinants of comprehensive student culture which includes inner factors of an organization of higher education institutions [22]. As for using the data of JCIRP (Japanese Cooperative Institutional Research Program), Yamada (2011) [25] draw the profile of first-year student through two student typologies, those are Sugitani typology (2009) [20] and Kimura typology (2009) [10]. Maruyama focused on the factors of the higher education system as well as social background and attributes of students. Takeuchi and Yamada focused also on the inside campus and the factors about education practices and interaction among students, faculties, and staffs. On the other side, Mizokami and Hatano (2013) make student types with focused on students’ identity and factors of aggressive attitude toward classes and students learning time [5,14]. The scope of student typology in Japan has deepened according to our society changed from the industrial society to the knowledge-based society and addressed the higher education system from mass stage to universal stage. In the present scope of student typology include how important of first-year education and how to support the development of student’s identity.

As Table 1 shows the studies of student typology have also evolved in the higher education studies of America [7]. Many researchers have made their own student typology other than table 1. Furthermore, there are student profiles from Interview survey, such as Levine and Dean (2012) [12] which they named today’s students as “the first digital generation.” The student typology diversified, however, even though several categories have added, Clark=Trow typology is the classic student typology of American higher education. Their student subculture model is the underlying framework for the other studies of student typology.

In 1993, Astin made student typology from a longitudinal student survey named the CIRP (Cooperative Institutional Research Program), the study has begun a new development [3]. A lot of higher education institutions have already participated in this survey. He is not only depended on empirical continuous survey, but also clearly aims to improve higher education practices.

The effects of education differed from gender, race-ethnicity, and social origin of students. But it also differed by student type. Astin insists that his student typology contribute to improve higher education practice, because that makes to consider the “fit” between student and environment. His student typology, however, doesn’t depend on the student involvement or engagement. Then researchers, Kuh, Hu, and McCormick those are researchers of NSSE (National Survey of Student Engagement) have

---
¹ Center for Cooperative Teacher Training Development
1 Hirosawa, Iga-ya-Town, Kariya-City, Aichi, Japan, 48-8542.
aihara@aucecc.aichi-edu.ac.jp
developed student typology from student engagement [6, 11, 26]. According to Hu and McCormick, to construct student typology of student engagement is the fruitful potential data use next to measure student engagement (p.739) [7].

2. Data

The data to make student typology are JFS (Japanese Freshman Survey) 2013. This survey is conducted as a part of JCIRP (Japanese Cooperative Institutional Research Program). JCIRP has conducted continuous longitudinal student survey which is a representative survey of Japan. These surveys are consist of JCSS (Japanese College Student Survey) for senior students, JJCSS (Japanese Junior College Student Survey) for students at two-year higher education institutions, and JFS for first-year students. JFS2013 was conducted from June to July at 2013. 91 higher education institutions participated and 15,472 freshmen answered this questionnaire. There are 155 students who did not answer their school grade or who were supposed that they were not freshman.

3. Kaneko Typology

3.1 Explanation

Under the colleges and universities reform in Japan, Kaneko (2007) [8] presents student typology which consist of two axes: the development of student’s self-social recognition and the fitness with the intention of their college and university education, and makes four student types: High-conformity, Limited-conformity, Acceptance, and Alienation. According to Kaneko, higher education institutions “are requiring double task to sift their scope for the student who have been out of the education range and give them more effective impact”. Figure 1 show the overlaid view of Clark=Trow typology and Kaneko typology although they are different time and society. The feature of Kaneko typology is practical availability oriented to educational reform and emphasize of content validity to construct the framework. It could use not only “to design enjoyable lesson plan based on this typology” (Morozumi; 2013) [15] but also available to check the admission policy. On the other hand, we can construct this typology without advanced statistical methods.
3.2 Constructing Typology

In this study, I construct Kanko typology from JFS 2013. Because JFS has not the same items with which Kaneko used, strictly speaking it is not the same typology. However, referencing to variables which Sugitani [20] uses to make her typology of “positive student” and “negative student”, I try to construct Kaneko typology. She made her typology with three items: fulfillment of student life, satisfaction of overall experience, and attitude toward re-entrance. Her typology covers overall student life. Kaneko, however, focused on teaching than student life. Then I choose to use satisfaction with overall quality of teaching. Figure 2 shows Kaneko typology from JFS2013 with two items: “satisfaction with overall quality of teaching” and “attitude toward re-entrance.” These student types are as follows.

High-conformity: students who are willing to re-entrance and satisfy with their classes share 16%. This type of students have a sense of belonging to higher education institutions and their future prospects are matching with education offered by their institutions. It may be assumed that students of this type are academic-oriented who are exploring their disciplines. However, in the undergraduate education of today, I suppose that these students are standing on the comprehensive foundation. According to Ougiya (1989) [17], this foundation consist of “(a) theoretical foundation that establish the base of academic-oriented education, (b) social practice foundation that associate education with practice, (c) liberal arts foundation that is based on humanistic thought (p.44)."

Limited-conformity: students who are willing to re-entrance but neutral or not satisfy with their classes. Their share is 48%. This type of students are not entirely matching their future prospects with the intention of education offered by colleges and universities. These students are (1) they have a sense of belonging to their institutions but are not actively involved so much in class, (2) they have not so much sense of belonging to their institutions but are not dissatisfy with their classes. This type of students share the majority at the campus.

Acceptance: students who are not willing to re-entrance but neutral or not satisfy with their classes share 25%. This type of students expect to education offered by colleges and universities anyway, but their sense of belonging are not enough. It is rather off-campus that they feel their place to stay. For example, these students are who have the first choice institutions other than enrolled institutions. Unlike the United States, transfer is unusual in Japan. Therefore, these students have a tendency to live with complex feelings, such as a sense of isolation and discouragement. The higher education institutions are required to so that students could seem to be a "it was good to enter", giving not only learning support on cognitive aspect, but also impact on the emotional aspect, such as efficacy and confidence.

Alienation: Students who are neutral or not willing to re-entrance and not satisfy with their classes share 11%. This type of students are not only able to have a sense of belonging in their institutions but also cannot be satisfied with their classes. They tend to do not involve higher education practices. It is difficult for higher education institutions to recognize their actual situation.

According to Kaneko, acceptance type and alienation type are not small students group at all, and it will be plausible to increase further in the future. It is the task for higher education institutions, with the limited resources, to enhance the educational power that is to expand the range of education and to give their students impact more effectively.

3.3 Features: Students Attributes and Typology

Table 2 shows the student profile of attributes and typology. High-conformity type has features of women, high or middle-high grade at high school, attend to the first choice and high selective higher education institutions. Their share is approximately 20% at most. If we changed the cutting point, it could increase their share. Kaneko (2007) pointed, however, about 20% are validated reasonable value. Limited-conformity type has features of women, middle grade at high school, attend to the first choice institutions. This type consists of majority at the campus that is the same of student subculture named “Collegiate” of Clark=Trow typology. Acceptance type has features of man, middle-low or low grade at high school, attend to not the first choice and low selective institutions. At the end, the alienation type has features of low grade at high school and attends to not the first choice institutions. It
Eight Student Types from Astin Typology

From table 2, we can observe the hierarchical differentiation among students with the achievement principle. That is, there is a tendency for High-conformity type students that they get high grade at high school and enroll at their first choice selective higher education institutions. On the other hand, there is a tendency for acceptance type and alienation type students that they get low grade at high school and attend to not the first choice, low selective higher education institutions. It seemed that attributes factors such as family income, indirectly influence on student types. Kaneko mentioned earlier that the share of acceptance type and alienation type students will increase. It means for higher education institutions that they are facing the crisis of enlarging and stabilizing the gap of hierarchical differentiation.

3.4 Utility and Marginality

The utility of Kaneko typology is its superior performance to guide educational practices to the targets. Let me give an example one fictional teachers college. 80% of their graduates of this college have to get teaching jobs and this college decides the strategic mid-term plan to increase graduates who get teaching jobs to 85%. To increase more 5% from present score, it would be desirable more efficient education practice that will require to identify the target. Because it is important to make continuous efforts for conformity type students, but in fact, almost all of the students of this student type have already got teaching jobs. On the other hand, acceptant type and alienation type of students who could get teaching jobs are around half. Then to achieve the mid-term objectives, it is necessary educational support suitable for these student types. This typology could indicate directions for better educational practices. It does not, however, indicate the details about what kind of student engagement and educational practices are supposed to be required and how the student learning outcomes is after all.

4. Astin Typology

4.1 Explanation

One feature of Astin typology is that it was constructed through nationally representative student survey named CIRP. Therefore, if we follow the same procedure, we can reproduce his typology at their higher education institutions. Astin [3] stated as follows.

The new typology reported in this paper offers considerable potential for use in educational research and practice, for at least two reasons: (a) it is based on data from a continuing national study of American undergraduate education in which more than 1,300 different institutions have already participated; and (b) it is entirely empirical in nature, which makes it possible for investigators or practitioners in any type of institution to utilize it simply by following the procedures outline later on in the paper (p.36).

<p>| Table 3: Eight Student Types from Astin Typology |</p>
<table>
<thead>
<tr>
<th>Leader</th>
<th>Status</th>
<th>Striver</th>
<th>Artist</th>
<th>Social</th>
<th>Activist</th>
<th>Scholar</th>
<th>Privation</th>
<th>Red houson</th>
<th>Un-committed</th>
</tr>
</thead>
<tbody>
<tr>
<td>α = .85</td>
<td>α = .81</td>
<td>α = .74</td>
<td>α = .76</td>
<td>α = .77</td>
<td>α = .71</td>
<td>α = .60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Participate in community activities</td>
<td>.972</td>
<td>-.107</td>
<td>-.110</td>
<td>-.030</td>
<td>-.099</td>
<td>-.005</td>
<td>.006</td>
<td>.009</td>
<td></td>
</tr>
<tr>
<td>2. Contribute to improve awareness of human rights</td>
<td>.838</td>
<td>-.036</td>
<td>.006</td>
<td>-.003</td>
<td>-.043</td>
<td>-.046</td>
<td>.002</td>
<td>.011</td>
<td></td>
</tr>
<tr>
<td>3. Participate in activities to protect the environment</td>
<td>.686</td>
<td>.024</td>
<td>.023</td>
<td>-.026</td>
<td>-.011</td>
<td>.015</td>
<td>-.013</td>
<td>-.019</td>
<td></td>
</tr>
<tr>
<td>4. Understand the diverse society and culture</td>
<td>.582</td>
<td>-.062</td>
<td>.057</td>
<td>.050</td>
<td>.001</td>
<td>.151</td>
<td>-.052</td>
<td>.059</td>
<td></td>
</tr>
<tr>
<td>5. Become community leaders</td>
<td>.581</td>
<td>.169</td>
<td>.064</td>
<td>.021</td>
<td>.011</td>
<td>-.135</td>
<td>.030</td>
<td>-.094</td>
<td></td>
</tr>
<tr>
<td>6. Developing a meaningful philosophy of life</td>
<td>.518</td>
<td>.046</td>
<td>.105</td>
<td>-.005</td>
<td>.021</td>
<td>-.159</td>
<td>.005</td>
<td>.124</td>
<td></td>
</tr>
<tr>
<td>7. Helping others who are in difficulty</td>
<td>.477</td>
<td>.010</td>
<td>-.115</td>
<td>-.016</td>
<td>.070</td>
<td>-.262</td>
<td>-.005</td>
<td>.022</td>
<td></td>
</tr>
<tr>
<td>8. Engage in political activities</td>
<td>.347</td>
<td>.125</td>
<td>.118</td>
<td>.090</td>
<td>.169</td>
<td>-.183</td>
<td>-.012</td>
<td>-.061</td>
<td></td>
</tr>
<tr>
<td>9. Get ahead in the workplace</td>
<td>-.027</td>
<td>.951</td>
<td>-.096</td>
<td>-.020</td>
<td>-.046</td>
<td>.012</td>
<td>-.009</td>
<td>.032</td>
<td></td>
</tr>
<tr>
<td>10. Being very well off financially</td>
<td>-.165</td>
<td>.754</td>
<td>.008</td>
<td>-.012</td>
<td>.034</td>
<td>.147</td>
<td>.031</td>
<td>.025</td>
<td></td>
</tr>
<tr>
<td>11. Become administrator in companies and organizations</td>
<td>.152</td>
<td>.641</td>
<td>-.063</td>
<td>.028</td>
<td>.052</td>
<td>-.105</td>
<td>-.024</td>
<td>-.009</td>
<td></td>
</tr>
<tr>
<td>12. Succeed in business</td>
<td>.188</td>
<td>.145</td>
<td>-.024</td>
<td>-.018</td>
<td>.061</td>
<td>.001</td>
<td>.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Produce a work of art</td>
<td>-.046</td>
<td>.031</td>
<td>.912</td>
<td>-.020</td>
<td>.001</td>
<td>.054</td>
<td>-.010</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>14. Writing a literary work</td>
<td>-.021</td>
<td>-.029</td>
<td>.874</td>
<td>.004</td>
<td>-.006</td>
<td>.021</td>
<td>-.023</td>
<td>.014</td>
<td></td>
</tr>
<tr>
<td>15. Success in performing arts</td>
<td>.079</td>
<td>.039</td>
<td>.702</td>
<td>-.017</td>
<td>-.035</td>
<td>-.013</td>
<td>.040</td>
<td>-.027</td>
<td></td>
</tr>
<tr>
<td>16. Discuss about religion</td>
<td>-.077</td>
<td>.002</td>
<td>-.016</td>
<td>.005</td>
<td>-.012</td>
<td>.047</td>
<td>-.116</td>
<td>-.005</td>
<td></td>
</tr>
<tr>
<td>17. Take part in religious activities and prayer</td>
<td>-.049</td>
<td>.016</td>
<td>-.011</td>
<td>.019</td>
<td>-.028</td>
<td>.027</td>
<td>-.028</td>
<td>.006</td>
<td></td>
</tr>
<tr>
<td>18. Participate in political activities</td>
<td>.021</td>
<td>.004</td>
<td>.021</td>
<td>.547</td>
<td>.002</td>
<td>-.060</td>
<td>.211</td>
<td>-.016</td>
<td></td>
</tr>
<tr>
<td>19. Discuss the political matters</td>
<td>.054</td>
<td>-.041</td>
<td>-.012</td>
<td>.542</td>
<td>.025</td>
<td>-.006</td>
<td>-.014</td>
<td>.071</td>
<td></td>
</tr>
<tr>
<td>20. Interacted with international students</td>
<td>.099</td>
<td>.023</td>
<td>-.013</td>
<td>.347</td>
<td>-.024</td>
<td>.003</td>
<td>.086</td>
<td>.031</td>
<td></td>
</tr>
<tr>
<td>21. Becoming an authority in their field</td>
<td>.091</td>
<td>-.065</td>
<td>-.002</td>
<td>.093</td>
<td>-.031</td>
<td>.020</td>
<td>.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Obtaining recognition of contribution to specialty</td>
<td>.015</td>
<td>.004</td>
<td>-.064</td>
<td>-.030</td>
<td>.717</td>
<td>.150</td>
<td>-.003</td>
<td>.073</td>
<td></td>
</tr>
<tr>
<td>23. Contribute to the theoretical development of science</td>
<td>.133</td>
<td>.099</td>
<td>.110</td>
<td>.018</td>
<td>.358</td>
<td>-.106</td>
<td>.038</td>
<td>-.060</td>
<td></td>
</tr>
<tr>
<td>24. Be valued relationships with friends</td>
<td>.159</td>
<td>.054</td>
<td>-.018</td>
<td>.022</td>
<td>.009</td>
<td>.735</td>
<td>-.030</td>
<td>-.067</td>
<td></td>
</tr>
<tr>
<td>25. Live a life with my own way</td>
<td>.016</td>
<td>.064</td>
<td>.076</td>
<td>.034</td>
<td>.037</td>
<td>.731</td>
<td>.001</td>
<td>.039</td>
<td></td>
</tr>
<tr>
<td>26. Smoking at the time of high school seniors</td>
<td>-.008</td>
<td>.014</td>
<td>.004</td>
<td>.027</td>
<td>.019</td>
<td>-.018</td>
<td>.899</td>
<td>.018</td>
<td></td>
</tr>
<tr>
<td>27. Drinking alcohol at the time of high school seniors</td>
<td>-.032</td>
<td>.019</td>
<td>-.007</td>
<td>.003</td>
<td>-.006</td>
<td>.049</td>
<td>.055</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>28. Depressed at the time of high school seniors</td>
<td>.008</td>
<td>.036</td>
<td>-.068</td>
<td>.022</td>
<td>-.023</td>
<td>-.104</td>
<td>.005</td>
<td>.018</td>
<td></td>
</tr>
<tr>
<td>29. Overwhelmed at the time of high school seniors</td>
<td>.075</td>
<td>.061</td>
<td>-.004</td>
<td>.096</td>
<td>.056</td>
<td>.001</td>
<td>-.041</td>
<td>.541</td>
<td></td>
</tr>
</tbody>
</table>

Factor extraction method: principal factor method. Rotation method: Promax method with the normalization of the Kaiser. Use question 22 and 25. All items are 4-point scale. Delete items those are large factor loadings with multiple factors. Select the number of factors with the scree plot method. Cumulative proportion of variance explained = 52%.
Another feature is that Astin typology is not only constructed in accordance with the conventional typology but also focus on the student’s consciousness such as values, attitudes, and beliefs so on. Therefore, his type describes the student’s personality rather than hierarchical order. Astin [3] stated as follows.

The typology reported here is designed instead to capture some of the uniqueness and individuality of students as personalities by utilizing information on their values, attitudes, beliefs, self-concept, and behavior. In this sense, the typology is more stylistic and descriptive than it is hierarchical (p.36).

4.2 Constructing Typology
Astin [3] used 60 items from freshman survey of CIRP. These are student’s values (7-point scale, 15 items such as, “helping others who are in difficulty,” “developing a meaningful philosophy of life,” and “being very well off financially.”), self-concepts (5-point scale, 11 items of about personal qualities such as academic ability, popularity, and originality.), behaviors (3-point scale, 10 items of such as “drank beer,” “stayed up all night.”), attitudes (4-point scale, 10 items such as “marijuana should be legalized.”), expectation (4-point scale, 13 items such as “change career choice,” “drop out of this college temporarily.”), and academic aspiration as one additional item.

JFS2013 is designed to compatible with freshman survey of CIRP. Because I would like to construct the Astin typology more simplify. I chose to employ only 29 items from the students’ values and behaviors those are all 4-point scale. Items of student attitudes such as “marijuana should be legalized” couldn’t have translated to Japanese, because it can’t applicable to Japanese students. Items of student’s expectations also could not utilize. It is comparatively less common in Japan to drop out from colleges and universities temporarily. At last, in order to make items to simplify, I did not utilize 11 items self-concept and additional one item of academic aspiration.

4.3 Features
Table 3 shows the result of making the Astin typology from JFS2013. Although there are differences, for example the item “helping others who are in difficulty” is located in the Leader type of table 3 whereas it located at the Social Activist of the Astin typology from the American student survey by CIRP. I construct the Astin typology with eight category which nearly same in the Astin typology. The value of Uncommitted reliability coefficient, Cronbach’s α is .60. This value .60 is low, however Astin adopt even .58.

It should be noted that the “Privatism” is a category there is not in the Astin typology. This student type whose best preference is their selves and surrounding friends might be the feature of contemporary Japanese students. These two items, number 24 and 25 are the items added as Japanese unique items. The features of each student type are as follows.

**Leader**: students who have traits of participating in regional and social activities, helping others who are in difficulty, improving human rights and securing the environment. Further they also have a tendency to reflect their beliefs and spirituality on life.

**Status Striver**: Japanese definition is nearly the same that is committed to being successful for the work of others, being very well-financially. However, it does not contain that is obtaining recognition from colleagues for contribution in their special field and becoming an authority in their field. Those are traits of the Scholar type in Japan.

**Artist**: students who have traits of high estimate on producing a work of art, writing a literary work, and being successes in performing arts.

**Social Activist**: students who have traits of discussing on religion and politics, participating in religious and political activities. Furthermore, it also contains that is interacting with international students.

**Scholar**: students who have traits of becoming an authority in their field, obtaining recognition from colleagues for contribution in their special field, and contribute to the theoretical development of science.

**Privatism**: this is the original type from JCIRP. Student who has traits to be valued relationships with friends and live a life of their own way.

**Hedonist**: the definition of this type is nearly the same the definition by Astin. Those are students who have used to smoke cigarette or drinking alcohol at the time of high school seniors.

**Uncommitted**: Astin said this type of student’ connection to his or her institution is highly tenuous. Japanese type of uncommitted is a little different from Astin. In Japan, this type of student who have used to be depressed or overwhelmed at the time of high school seniors.

**4.4 Utility and Marginality**
As Astin admitted, his typology is very similar with the Holland typology which is very famous depended on the career choice theory [3]. In fact, five of eight categories are the same labelling. Therefore, it is propriety to use the Astin typology for the choice of major by students and to know the status of the interest and concern of students in departments. It is also helpful to use as a tool of counseling at the career choice and measuring personality development of students. Furthermore, it might be possible that higher education institutions to practice the education that takes into account of the “fit” between students and environments. It is also the same with the Kaneko typology, however, that this typology could indicate directions for better educational practices. It does not, however, indicate the details about
what kind of student engagement and educational practices are expected to be required and how the student learning outcomes is after all.

5. College Impact to Alienation Type Students

5.1 Analytical Framework

In order to expand the range of education and to give impacts more effectively, I examine the college impact to the alienation type student. Analytical framework used here is comprehensive I-E-O model shown in Figure 3. Based on the involvement theory and I-E-O model of Astin (1993) [2], this study extended the model through by referring to recent studies of Terenzini and Reason (2005) [24] and by incorporating the psychological aspect of engagement those are emphasized in Japan [1].

I-E-O model is widely used in the major of higher education studies. According to Astin, Inputs means the Pre-enrollment Information such as the characteristics of students. Environments means what students receive after enrollment from educational experience, peer, faculty, staffs, education programs, and education policy. Outcomes means after entering information about the characteristics of students who are affected by the environment (Astin, 1993, p.7) [2]. According to Astin, the most important factor in determining the learning outcomes is student involvements that make up the environmental factors group.

Items which use in this study are as the following.

As the Pre-enrollment Information, I consider (1) gender, (2) first-generation students, (3) high school grades, and (4) college choice order. Traditional college students who represented the college students in the United States in the 1960s, had an attribute called youth, white, male, Protestant, excellent academic grade, from a privileged economically well home. As to gender, currently many women students have been entered in college. However, there is a bias in majors, women students do not tend to attend the majors called STEM those are natural science, technology, engineering, and mathematics. The first generation students mean that students whose family has no graduates of universities and junior colleges. Student survey in the United States, they often ask about sexual orientation or ethnicity, disability, religion, and family income, etc. JFS, however, does not ask them for these personal information. Then in this study, I examine gender, the status of first-generation, academic grade in high school, and college choice order. These are all self-report of students.

Followed the Reason and Terenzini model, Learning Environment part divided into two domains: Institutional Character and Student Involvement.

Institutional Character include that College Character (control, scale, location, etc.), Financial Aid (scholarship, tuition discount, etc.), Resident Status (resident student, commuter student, etc.), Curriculum and Major. In this study, I examine (1) control and (2) selectivity of college. As to the selectivity of college, I referred to the difficulty of the entrance examination published in several exam magazines. As for Financial Aid, usually researchers had been studied the effect of the scholarship to Learning Outcomes from student surveys. There are several type of scholarship: gift types with no repayment obligation, loan type of repayment obligation, and work-study type with providing labor opportunities on campus. In addition, some colleges in the United States, every freshman stay at the student dormitory. Those policies on scholarship and residence are allowed students to concentrate on their studies and to improve the learning outcomes.

Student Involvement has five domains: (1) Faculty and Staff Support, (2) Student Peer Relations, (3) Psychological
Aspect of Student Engagement, (4) Self-Concept, and (5) Academic Involvement. It should be noted that in the student survey, in many cases, student involvement and engagement those are the central concept of the theory of involvement, are often measured by learning behavior and learning time of students. In the JFS2013, however, we ask students about learning behavior and learning time at their high school, but do not ask them at their institutions. Therefore, Student Satisfaction is a utility that results from Learning Outcomes as showed in Figure 3, but I will substitute these satisfaction items to student involvement items.

For example, I examine (1) Faculty and Staff Support of satisfaction items as emotional support of "the opportunity to talk with teachers", and as cognitive support of "individual advice and support to learn." And I examine (2) Student Peer Relations by satisfaction items of “climate that acknowledge diversity" and "a sense of belonging among students." (3) Psychological Aspect of Student Engagement is created as a domain of identity formation of students. This domain is termed as self-recognition and social-recognition in Kaneko type. I examine this area with satisfaction items of "effectiveness of teaching to career planning" and "relevance of teaching to everyday life." As for Astin model, Reason and Terenzini model, they do not have this domain. Then this is a unique area of Japan. (4) Self-Concept is a domain of self-efficacy and self-confidence. These are products from self-assessment or self-monitoring to learning process and outcomes in which students engaged among the relationship with faculty, staff, and student peer. There is a question of self-evaluation in the freshman survey and it asked students about their social confidence, academic confidence and so on. Self-efficacy and self-confidence are important. However, in this study, they should be considered as some kind of outcomes from student involvements. According to Astin, that is intermediate outcomes (Astin, 1993, p.80) [2]. Then I do not use these items. (5) Academic Involvement is the domain of behavioral aspect, such as learning activity. There are items about the learning time of devoting to courses and the distinctive education programs those are study abroad program, internship, research activity in the undergraduate program and so on. In addition, there are items about the frequency of delay, absent to courses and there are items about the frequency that could not submit a report by the deadline. In this study, I examine satisfaction of "liberal arts education and general education" and "first year student program."

It should be mentioned that in the college impact study in Japan, Takeuchi (2008) [23] explains I-E-O model in more detail from the socialization model of the students. In addition, there are some previous studies of determinants analysis of learning outcomes such as Ogata (2008) [16].

5.2 Results and Analysis

Depending on the comprehensive I-E-O model, I examine statistical analysis in order to study about college impact through the effects of student involvements toward alienation type students. Table 4 shows the list of variables that used in the analysis and their basic statistics. Dependent variable is a dummy variable defined as alienation type = 0, non-alienation type = 1. Alienation type students mean, as defined at Kaneko typology, they do not have consciousness about re-entrance and they have dissatisfaction with their courses. To escape from alienated situation, what kind of factors would have been useful for students? It tries to find from the 18 explanatory variables.

Pre-entering Information consists of 7 dummy variables. Those are gender, first-generation, high school grades (high, middle-high, middle-low, low), and order of college choice. These are the default conditions at the time when students enroll. Learning Environment part consist of 3 variables for

<table>
<thead>
<tr>
<th>Table 4: Variables and Basic Statistics</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Value</th>
<th>AV</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alienation Type dummy</td>
<td>14,927</td>
<td>Alienation Type=0, Others=1</td>
<td>.89</td>
<td>.310</td>
</tr>
<tr>
<td>Pre-entrance Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender dummy</td>
<td>14,997</td>
<td>Man=0, Woman=1</td>
<td>.50</td>
<td>.500</td>
</tr>
<tr>
<td>First generation dummy</td>
<td>14,857</td>
<td>Others=0, First Generation=1</td>
<td>.40</td>
<td>.489</td>
</tr>
<tr>
<td>High school</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>15,021</td>
<td>Others=0, High=1</td>
<td>.19</td>
<td>.392</td>
</tr>
<tr>
<td>Middle High</td>
<td>15,021</td>
<td>Others=0, Middle=1</td>
<td>.29</td>
<td>.461</td>
</tr>
<tr>
<td>Low</td>
<td>15,021</td>
<td>Others=0, Low=1</td>
<td>.15</td>
<td>.373</td>
</tr>
<tr>
<td>College choice order</td>
<td>15,462</td>
<td>Others=0, First Choice=1</td>
<td>.57</td>
<td>.490</td>
</tr>
<tr>
<td>Institutional Character</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College control dummy</td>
<td>15,519</td>
<td>National or Public=0, Private=1</td>
<td>.65</td>
<td>.477</td>
</tr>
<tr>
<td>College Selectivity</td>
<td>14,865</td>
<td>Others=0, High=1</td>
<td>.12</td>
<td>.327</td>
</tr>
<tr>
<td>Student Involvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty and Staff Support</td>
<td>13,930</td>
<td>Opportunity to Talk with Faculty</td>
<td>3.05</td>
<td>.776</td>
</tr>
<tr>
<td></td>
<td>12,793</td>
<td>Learning Advice and Support</td>
<td>3.05</td>
<td>.757</td>
</tr>
<tr>
<td>Student Peer Relations</td>
<td>13,993</td>
<td>A Sense of Belonging</td>
<td>3.19</td>
<td>.833</td>
</tr>
<tr>
<td></td>
<td>13,820</td>
<td>Climate to Admit Diversity</td>
<td>3.20</td>
<td>.791</td>
</tr>
<tr>
<td>Student Engagement</td>
<td>14,114</td>
<td>Relevancy of teaching to life</td>
<td>3.13</td>
<td>.730</td>
</tr>
<tr>
<td></td>
<td>13,573</td>
<td>Efficiency to Career Plan</td>
<td>3.22</td>
<td>.755</td>
</tr>
<tr>
<td>Academic Involvement</td>
<td>14,551</td>
<td>Career Education</td>
<td>3.31</td>
<td>.820</td>
</tr>
<tr>
<td></td>
<td>12,396</td>
<td>Liberal Education</td>
<td>3.23</td>
<td>.789</td>
</tr>
</tbody>
</table>

Sample Number after list remove = 8,796
Institutional Character those are college control and college selectivity (high, middle), 8 variables for Student Involvement those are each 2 variables for 4 domains. Valuables of Institutional Character are dummy variables and the variables of Student Involvement are five-point scale. Higher education institutions cannot easily change their control and selectivity. In this analytical framework, dependent variable is qualitative variable and explanatory variables are qualitative and quantitative variables. Then I employed logistic regression analysis as a statistical analytic method and used SPSS software.

Table 5 shows results of analysis from model1 (explanatory variables are Pre-entrance Information and Institutional Characteristics) and model2 (explanation variables are Pre-entrance Information, Institutional Character and Student Involvement). Look at model 1. The effects of statistical significance find on 5 variables. Those are gender, high school grade-high, high school grade-middle high, high school grade-low, and selection degree-high. It is difficult to interpret of the results of the logistic regression analysis than ordinary regression analysis. If the value of coefficient is greater than 1 then there is a positive effect, on the other side if the value of coefficient is less than 1 then there is a negative effect. The magnitude of that effect indicates odds ratio E(B). For example, the effect of woman is exp (0.197) = 1.218. It shows the expectations (odds) that women students are not to become alienation type about 1.2 times than men. As the same way, the odds ratio show the expectations 3.3 times of college choice-first, 1.4 times of college selectivity-high. The students who attend to the college of first choice order, who attend to highly selective college, and women students. What is interesting that the effects of high school grade. The expectation that students whose high school grade is high are not to become alienation type is 0.68 times, that is about two-thirds. It means that students who get high performance at high school are more likely to become alienation type students. In other words, students who have a sense of failure, not satisfied with the class is often the good grade at high school. Also, there was no statistically significant effect on the first generation students and control type. Therefore, if there are graduates of college or junior colleges in the family, and if college control is national, public or private, those will not affect the alienation type students. It should be noted that, from the value of 0.071 pseudo R², there is not much explanatory power of the model 1.

Then, look at model 2. Statistically significant effects are college choice-first, college selectivity-high, college selectivity-middle, and all of Student Involvement variables. From the value of odds ratio of E (B), the highly effective

<table>
<thead>
<tr>
<th>Table 5: Logistic Regression Analysis to Alienation Type Student</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Pre-entrance Information</strong></td>
</tr>
<tr>
<td>Gender dummy</td>
</tr>
<tr>
<td>First Generation dummy</td>
</tr>
<tr>
<td>High school grade dummy</td>
</tr>
<tr>
<td>Reference category: Middle-High</td>
</tr>
<tr>
<td>Reference category: Middle</td>
</tr>
<tr>
<td>Reference category: Low</td>
</tr>
<tr>
<td>College Choice dummy</td>
</tr>
<tr>
<td><strong>Institutional Character</strong></td>
</tr>
<tr>
<td>Control dummy</td>
</tr>
<tr>
<td>College Selectivity dummy</td>
</tr>
<tr>
<td>College Selectivity dummy</td>
</tr>
<tr>
<td><strong>Student Involvement</strong></td>
</tr>
<tr>
<td>Faculty and Staff Support</td>
</tr>
<tr>
<td>Student Peer Relations</td>
</tr>
<tr>
<td>Student Engagement</td>
</tr>
<tr>
<td>Academic Involvement</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Maximum of Log Likelihood (-2 log likelihood)</td>
</tr>
<tr>
<td>Pseudo R² (Nagelkerke R²)</td>
</tr>
<tr>
<td>n</td>
</tr>
</tbody>
</table>

Note: * p<.05, ** p<.01, *** p<.001
variables are 3.1 times of "relevancy of teaching to everyday life" from student engagement, 3.1 times of college choice order, 2.3 times of college selectivity-high, 1.8 times of satisfy with liberal arts education and general education. Effects of college choice order and college selectivity are still large even model 2. However, the effects of the psychological aspect of student engagement and academic involvement which is satisfaction of liberal arts education and general education are no less than those effects. If we sum up the effects of other Student Involvement: faculty staff support, student peer relations, satisfaction of "effectiveness of teaching to career planning," and satisfaction of "first year education program," the total effects of student involvements after entrance have a large effects toward the alienation type students. Therefore, in order to expand the range of college education and to give impacts more effectively, the effect of Student Involvement is substantial enough. In addition, it is not recognized statistical significance other than the variable of college choice order of Pre-entrance Information and the variables of college selectivity of Institutional Character. Further the value of 0.408 of pseudo R² indicates the explanatory power of the model 2 is significant.

The requirement of higher education policy today is require higher education institutions to enhance the educational power that is to expand the range of education and to give impact more effectively. As compare with Clark and Trow construct their student type in the 1960s, diversification of students’ body also requires higher education institutions to understand students through new student typology. Because resources of intuitions are limited, it is necessary for institutions to provide education with focused on the target students what individual student needs. In this study, I construct a student type depending on previous studies and focused on the alienation type students. The result of logistic regression analysis that relies on comprehensive I-E-O model revealed that higher education institutions are capable to widen the range of their education to use the effects of student involvements toward the alienation type student.

Among the re-examination of comprehensive higher education, Ougiya emphasize that “in the case that academic nature is to explore the truth of knowledge and its practice at the same time, we can possible to grow the learners’ engagement and their practical identity. Academic studies what engaged to construct of theoretical infrastructure of general education should stand on this natural academic perspective where design to integrate of the truth and practices” (1989, p. 44)[17].

The results of analysis on the effect of student involvements toward the alienation type students, it makes recognize that comprehensiveness of higher education is important. Student Involvement, above all Psychological Aspect of Student Engagement, can exert their effect if the curriculum of higher education has relevance about teaching to everyday life for example, and integrate them with self and social recognition. Therefore, it is important for students without a sense of alienation that higher education institutions provide their education what exploration of the truth integrates with vocational practice and students’ identity formation. It is also to change the higher education through the P-D-C-A cycle of curriculum management to reduce the share of alienation type students, and try to increase the share of high-conformity students.

At the end, I present some future task for analytical model. First, this study has not been enough scope to grasp the diversity of students. The National Survey of Student Engagement that has been revised recently in United State (NSSE2.0) is trying to wide their scope of the diversity students further, such as disabilities. Second, in the institutional character of higher education institutions, I have considered only college control and selectivity. It will need to examine the policy effects about residence and scholarship. Third, in the domain of Student Involvement, I substitute learning behavior items with satisfaction items. The questionaire of JFS2013 has not items about student learning time and behavior after their entrance. We are required to examine these items with student behavior. It is also necessary to consider the effect of the learning time and distinctive education program for each student type.

6. Discussion and Conclusion

The higher education institutions in the universal stage of the 21st century, it is required to provide effectively a good quality of education to the diverse students. Experienced skillful college teachers can distribute the attention to a wide range of students in their class. They can grasp the characteristics of students at each academic year, at the each departments and courses, and can make lesson plans for their classes. The ordinary teachers, however, usually tend to go for their eyes only to excellent students in many cases. They have a tendency to do not care so much to the majority of students. Sometimes they have some kind of stereotype views. In sometimes, they do not notice the presence of students who are alienate or disengaged. The student typology is a framework to search the range and impact of education practice. Through student typology, we can design enjoyable lesson plan which can provide education to a wide range and with high impact.

In this study, I construct two types of student typology from JFS2013 of JCIRP, those are Kaneko typology and Astin typology. Because items are not same, strictly meaning these typology from JFS2013 is not the same. However, I can say that JFS has capable to construct nearly same typology. Kaneko typology is comparable with the most famous student typology in higher education studies named Clark=Trow typology. Despite the evolution of studies, their structure nearly half-century ago has been maintained. That is, they basically make student type with
two-axes and make a four-dimensional area. Astin constructed his typology in 1993 from freshman survey of CIRP. The difference of earlier studies, he clearly intends to improve higher education using student typology from the longitudinal student survey. This tendency becomes further clear from the development of student typology by researchers of NSSE. The role of engagement indicators is to guide higher education institutions to a good education practice. If we construct student typology with student engagement, then we can make educational practices those are suitable for each students. However, this kind of typology is still under development and they do not use in practices. The development of future studies on student typology is prospected.

Note
The part of this study comes from two presentations those are June 1 at Liberal and General Education Society of Japan and June 7 at the Japanese Association of Higher Education Research.

Acknowledgment
The author appreciates kind advices from reviewers and thanks to JCIRP research team for their friendly support. This research supported by the Grant-in-Aid for Scientific Research (A) No.22243047 from Japan Society for the Promotion of Science.

References
Soichiro Aihara is a research fellow in the center for Cooperative Teacher Training Development at Aichi University of Education. Mr. Aihara's research focus is longitudinal survey to student and faculty from the international comparative viewpoint. He has investigated issues about evaluation indicators for learning process and outcomes, and faculty salaries. He has developed comprehensive I-E-O model and has made evaluation indicators of student learning for higher education institutions from student survey of Japanese Cooperative Institutional Research Program (JCIRP). He is an active member of the Association for Institutional Research (AIR), the Japanese Association of Higher Education Research (JAHER), the Japan Society of Educational Sociology (JSES) and the Liberal and General Education Society of Japan (LGESJ).
Abstract

The present study proposes methods for predicting student grades based on their comment data. Students describe their learning attitudes, tendencies and behaviors by writing their comments freely after each lesson. The main difficulty in this research is to predict student performance by separately using data for two different classes for each lesson. Although students learn the same subject, there are differences between the comments in the two classes. The proposed methods basically employ latent semantic analysis (LSA) and two types of machine learning technique: support vector machine (SVM) and artificial neural network (ANN) to predict students’ final results in terms of four grades: S, A, B and C. Moreover, an overlap method is proposed to improve prediction accuracy for stable evaluation; the method allows the acceptance of two grades for one mark to get the strong relation between LSA results and student grades. The proposed methods achieve 62.6% and 57.2% accuracy in predicting student grades from lesson 1 to 6, and 50.7% and 48.7% from lesson 7 to 15 for SVM and ANN, respectively.

The results of this study indicate models of student academic performance predictors that are valuable sources for understanding student behavior and giving feedback to them so that we can improve their learning activities.

Keywords: Comment mining, Latent semantic analysis (LSA), Artificial neural network (ANN), Support vector machine (SVM), Overlap method.

1. Introduction

Recently, many researchers have turned their attention to enhancing learners’ performance. They have contributed to the related literature. By and large, researchers in this field manage to advocate novel and smart solutions to improve performance [1]. Thus, learners’ performance assessment may not be viewed as being somewhat separate from learning process. It must be a continuous process.

Performance assessment is an integral part of learning processes [2] and, ultimately, should aim to improve the quality of student learning.

Predicting student academic performance has long been an important research topic in many academic disciplines. Based on the results of a predictive model, the instructor can take proactive measures to improve student learning, especially for those low-performance students. The results of a predictive model can be used to encourage those low-performance students to develop a better learning strategy. The prediction results might help students develop a good understanding of how well, or how poorly, they would perform in a course and therefore “force” students to rethink the way in which they have been learning [3].

In the classroom, there are many varieties of students. They have wide-ranging performance. Some are asiduous and self-motivated; others find it difficult to understand the lesson or are frustrated by the subject. Teachers can give advice based on their careful observation, but it is hard to grasp all the class members’ learning attitudes over all the periods in the semester. To control students’ learning behavior and situations, previous studies have used various regular assessment methods such as e-learning logs, test marks and questionnaires. Nevertheless, they have difficulty of interpreting them and creating good questions to evaluate each student. Although, teacher observation still has a vital role to play in improving the educational situation, teachers can only deal with some cases according to their needs mainly based on their experience in the class. Moreover, there is a difficult problem to resolve in the large number of factors or characteristics that can influence students performance, such as demographic, cultural, social, or family factors, socio-economic status, psychological profile, previous schooling, prior academic performance, interactions between students and faculty, etc. [4]. The current study proposes a method for predicting student
grades. Unlike previous studies, our method is based on students’ free-style comments collected in class. The student comments are good resources for predicting their learning situations. Each student writes his/her comments after a lesson; the student looks back upon his/her learning behavior and situation; he/she can express his/her attitudes, difficulties, and any other information that might help a teacher estimate his/her learning activities.

Goda et al. [5, 6] proposed the PCN method to estimate students’ learning situations from free-style comments written by the students. The PCN method categorizes the student comments into three items: P (Previous activity), C (Current activity), and N (Next activity). It provides data expressing students’ learning status, and reducing the task for self-observations, self-judgments, and self-reactions. Item P indicates learning activity before the class time. Item C shows the understanding and achievements of class subjects during the class time, and item N expresses the learning activity plan until the next class.

To further contribute to the understanding of student attitudes, this paper presents a study that applies text mining techniques to comments of item C (C-comments for short) from the PCN method [5, 6]. We apply the LSA technique to the comments data. LSA can construct a conceptual vector space in which each term or comment is represented as a vector in the space. It not only greatly reduces the dimensionality but also discovers the important associative relationship between terms [7].

The experiments are conducted using data from two classes (Class A & Class B) for each lesson, using one class data for training and the other for testing. Although students learn the same subject in the two classes, there are differences between the comments in the two classes. Therefore this is a challenging problem. The contributions of our work are the following:

- An artificial neural network (ANN) model is employed to predict student grades. ANN offers a number of advantages, including ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms [8].
- The SVM model is also employed to predict student grades. SVM is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates a set of objects having different class memberships [9].
- The overlap method is proposed to substitute an adjacent grade for the original grade for stable evaluation.
- The experiments are conducted to validate the proposed methods by calculating the $F$-measure and accuracy for each lesson in estimating the final grades with ANN and SVM models.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 describes the overview of our research. Section 4 introduces the methodology of the proposed method. Section 5 discusses some of the highlighted experimental results. Finally, Section 6 concludes the paper and describes our future work.

2. Related Work

The main objective of any higher educational institution is to improve the quality of managerial decisions and impart quality education. Good prediction of student’s success in a higher learning institution is one way to reach the highest level of quality in higher education systems. The implementation of Educational data mining (EDM) has been carried out in this area to predict students’ academic performance. Extensive literature reviews of the EDM research field are focused mainly on retention of students, improving institutional effectiveness, enrollment management and alumni management [10, 11]. Furthermore, several leading EDM experts [12, 13, 14] classify work in EDM into a few categories such as statistics and visualization, prediction (classification, regression, and density estimation), clustering, relationship mining, outlier detections, and text mining.

EDM can be applied to many applications that assess student learning performance, improve the learning process and guidance of student learning, provide feedback and adaptive learning recommendations based on student learning behaviors, evaluate learning materials and courseware, detect abnormal learning behaviors and problems, and achieve deeper understanding of educational phenomena [12, 15, 13].

Quite a few EDM studies have been found in the most recent literature from 2010 to 2014. For example, Gorissen et al. [16] analyzed the interactions of students with recorded lectures using educational data mining techniques. The data logged by a lecture capture system (LCS) was used and combined with collected survey data. They found discrepancies as well as similarities between students’ verbal reports and actual usage logged by the recorded lecture servers. The data suggests that students who do this have a significantly higher chance of passing their exams. They concluded that given the discrepancies between verbal reports and actual usage, research should no longer rely on verbal reports alone. Jovanovica et al. [17] applied classification models to predict student performance, and cluster models to group students based on their cognitive styles in the e-learning environment. They indicate that the classification models helped teachers, students and business people to engage early with students who are likely to become excellent on a selected topic. Parack et al. [18] used multiple data mining algorithms for student profiling and grouping. They found that data mining can be very useful in discovering valuable information which can be used to
profile students based on their academic record, such as exam scores, term work grades, attendance and practical exams. In addition, Bachtiar et al. [19] developed an estimation model to predict students’ English (listening, reading, speaking, and writing) skills and performance. They proposed a questionnaire to quantify the three major effective factors for students: motivation, attitude, and personality. The components of these factors are further identified by exploring each factor conceptually. They applied a neural network model in their experiments. The accuracy scores obtained by the model were 93.3% for listening, 94.4% for reading, 94.9% for speaking, and 93.6% for writing skills. Also, Kabakchieva [20] focused on the implementation of data mining techniques and methods for acquiring new knowledge from data collected by universities. The main goals of the research are to find out if there are any patterns in the available data that could be useful for predicting student performance at the university based on students’ personal and pre-university characteristics. The experimental study classified data by decision tree algorithm (C4.5 and J48), Bayesian classifiers (NaiveBayes and BayesNet), a the Nearest Neighbour algorithm (IBk) and two rule learners (OneR and JRip). The results indicated that the prediction rates were not remarkable (varying between 52 and 67%). Moreover, the classifiers perform differently for the five classes. The data attributes related to the students’ university admission score and the number of failures in the first-year university exams are the most influential in the classification process. Adhatrao et al. [21] built a system to predict student performance from the prior knowledge of enrolled students using concepts of data mining techniques under classification. They analyzed the data set containing information about the students, such as gender, marks scored in the board examinations, marks and rank in entrance examinations and results in the first year of the previous batch of the students. They applied ID3 and C4.5 classification algorithms, and predicted the general and individual performance of freshly admitted students in future examinations. The accuracy of the result was 75.15% for both ID3 and C4.5 algorithms. Osmanbegovi et al. [22] applied three supervised data mining algorithms (Naive Bayes, neural network, decision tree) to the preoperative assessment data, to predict students’ pass or failure in a course; they evaluated the prediction performance of the learning methods based on their predictive accuracy, ease of learning, and user-friendly characteristics. The results indicated that the Naive Bayes classifier outperforms decision tree and neural network methods in its predictive accuracy.

An emerging trend in EDM is the use of text mining, which is an extension of data mining to text data. Text mining focuses on finding and extracting useful or interesting patterns, models, directions, trends, or rules from unstructured text such as text documents, HTML files, chat messages and emails. In addition, the major applications of text mining include automatic classification (clustering), information extraction, text summarization, and link analysis [23]. As an automated technique, text mining can be used to efficiently and systematically identify, extract, manage, integrate, and exploit knowledge for research and education [24].

Currently, there are only several studies about how to use text mining techniques to analyze learning related data. For example, Tane et al. [25] used text mining (text clustering techniques) to group e-learning resources and documents according to their topics and similarities. Dringus et al. [26] conducted a research for embedding data/text mining techniques in a threaded discussion forum. They provided a strategy for assessing discussion forums in a manageable way, developed an assessment tool set that could be embedded in a threaded discussion forum, and pointed out the complexity and inconsistency inherent in a natural language text. However, intuitive measures, such as the rate of student participation and the length of discussion threads, are not necessarily good ways to judge the health of discussion forums or the quality of learning taking place [27]. Also, Lin et al. [27] proposed a genre classification system (GCS) to facilitate the automatic coding process. They treated the coding process as a document classification task via modern data mining techniques. The genre of a posting can be perceived as an announcement, a question, clarification, interpretation, conflict, assertion, etc. Also, they examined the coding coherence between GCS and experts’ judgment in terms of recall and precision. The results showed that GCS system can effectively facilitate the coding process and deal with the imbalanced distribution nature of discussion postings. Hung [28] used clustering analysis as an exploratory technique to examine e-learning literature and visualized patterns by grouping sources that share similar words and attribute values. Wu He [11] discussed methods to exploit large amounts of untapped student data including online questions and chat messages automatically collected by a live video streaming (LVS) system. The data and text mining results revealed some interesting patterns and themes in student’s interaction with the instructor and other students. Furthermore, the study revealed a disciplinary difference in online LVS participation between LVS students in different colleges. The results showed that students in Engineering and Science asked significantly less questions than students in Education and Health Science. Similarly, students in Engineering and Science also chatted much less frequently than students in Education and Health Science. A further examination concerning two educational courses found positive correlations between the number of questions students sent to the instructor and their final grades. Minami et al. [29] analyzed student attitudes towards learning, and investigated how they affect their final evaluation; they pursued a case study of lecture data analysis in which the correlations between student attitudes to learning, such as attendance and homework, as effort, and the students’ examination scores, as achievement; they analyzed the
students’ own evaluation of themselves and lectures based on a questionnaire. Through this study, they showed that a lecturer can give feedback to students who tend to over-evaluate themselves, and allow the students to recognize their real positions in the class.

Previous studies show that we need to understand individual students more deeply, and recognize students’ learning status and attitude to give feedback to them. Although applying questionnaires gave better results e.g. [19] than previous data (e.g. concerning personality, sociality, and student behavior), we need to comprehend student characteristics by letting them describe themselves, their educational situations, such as understanding of subjects, difficulties in learning, learning activities in the classroom, and their attitude toward the lesson.

Different from the above studies, Goda et al. [5] proposed the PCN method to estimate student learning situations on the basis of their freestyle comments written just after lesson. The PCN method categorizes their comments into three items: P (Previous), C (Current), and N (Next) so that it can analyze the comments from the point of view of their time-oriented learning situations. Goda et al. [6] also conducted another study on using PCN scores to determine the level of validity of assessment based on student comments and showed there are strong correlations between the PCN scores and accuracy of predicting students’ final grades. They employed multiple regression analysis to calculate PCN scores. Their results indicated that student comments with high PCN scores are considered to appropriately describe the student’s learning attitude. In addition, they applied machine learning method support vector machine (SVM) to the comments for predicting the students’ final results in terms of five grades: S, A, B, C, and D. The experimental results illustrated that as student comments get higher PCN scores, prediction of the students’ grades becomes more accurate. However, they did not discuss the accuracy of prediction of students’ final grades.

The current study is an extension of Goda et al. [6]; we focus on accuracy of prediction of students’ final grades by using C-comments from the PCN method; we try to predict their final grade in each lesson and discuss the changes in accuracy over a sequence of lessons.

3. Background

3.1. Subject of the study

In our study, we use the student comment data collected from Goda’s courses consisting of 15 lessons. The main subject from lesson 1 to 6 is computer literacy, giving information on how to use some IT tools such as word processors, spread sheets, and presentation tools. Computer literacy education is compulsory throughout senior high schools in Japan, with only a few differences in the details of course contents. From lesson 7 to 15, students begin to learn the basics of programming. The main subjects in those lessons are introductory C programming. In the 7th lesson, most students are novices at programming; it’s a new subject and not required until they enter the university [5]. Table 1 displays examples of C-comment data written by students from lesson 1-6 and 7-15.

3.2. Comment Data

In this research we use C-comments from [5, 6] to predict student grades. They have a stronger correlation with prediction accuracy than P- and N-comments [6]. In addition, they indicate the current learning activity of each student by expressing the understanding and achievements of class subjects during the class time.

Comments data were collected from 123 students in two classes: (Class A = 60 students) and (Class B = 63 students). Although we have 123 students in all lessons, some students didn’t submit their comments because they didn’t write any comments or were absent. Figure 1 displays the average real number of comments in each grade. We chose five grades (S, A, B, C and D) instead of the mark itself to predict each student’s results from his/her comments. The assessment of each student was done by considering the average mark of the student’s reports assigned three times, and attendance rate.

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Comment</th>
</tr>
</thead>
</table>
| (1-6)  | - The way of input is quite difficult; I feel that I could hardly follow all the steps. When I managed to deal with the calculation, I started to feel it’s interesting.  
- Today we created a new mail account; it's quite interesting to send email to a teacher and to people around me. I learned about several types of mail protocols, like POP, IMAP and etc. Also I learned that it is possible to save the messages to a computer. |
| (7-15) | - I was completely able to understand the subject of this lesson and have the confidence to make other functions similar to the ones I learned in this lesson.  
- I didn't finish all the exercises, because I could not understand the last two methods and the time was up. |

Figure 1 (a). From lesson 1 to 6.
Figure 1. The Correspondence relation between the comments number and the grades.

From Figure 1(a), it can be seen that grade S has the same number of comments in the class A and B data from lesson 1 to 6. There is a little difference between the number of comments of grade C and D in class A and B; the number of comments is less than 10. Figure 1(b) displays the average real number of comments from lesson 7 to 15. We can see that the number of comments with grade A, C and D in class A is smaller than that in class B. The number of comments with grade D is about 64% smaller than that of class B. In addition, more than 50% of students with grade D didn’t submit comments, whose total number is less than 7. So we combined comments data in grade C and D into one grade C and use 4-grade categories: S, A, B and C.

The number of words appearing in each comment averages 14 words, and the number of words contributed by each class is about 780 words. In addition, the number of distinct words in each class is over 310 words.

3.3 Procedures of the Proposed Method

Figure 2 displays the overall procedures of the proposed method; we have the following four phases: 1) Comment Data Collection, 2) Data Preparation, 3) Training Phase, 4) Test Phase.

1. **Comment Data Collection:** This phase focuses on collecting students’ comments after each lesson.

2. **Data Preparation:** The data preparation phase covers all the activities required to construct the data set for the training and test phases from the initial raw data, which we call the data set LSA results. Our methods analyze C-comments, extract words and parts of speech, calculate the occurrence frequencies of the words, apply a log entropy term weighting method so as to balance the effect of the occurrence frequency of words in all the comments and employ LSA to analyze patterns and relationships between the extracted words and latent concepts contained in the unstructured collection of texts (students’ comments).

3. **Training Phase:** In this phase, we employ LSA results to predict student grades by building a model for each lesson with ANN and SVM models.

4. **Test Phase:** This phase involves extracting words from a new comment, and transforming an extracted word vector of the comment into a set of $k$-dimensional vectors obtained by LSA.

4. Methodology

This section describes our methodology for predicting student performance from free-style comment data. We analyzed the data and applied the LSA technique to the data, then classified the results obtained from the ANN and SVM models to predict students’ final results as one of four grades S, A, B and C. After testing both methods, results will be compared.

4.1 Term Weighting of Comments

After choosing C-comments, we analyze comment data with Mecab program, which is a Japanese morphological analyzer designed to extract words and identify their part of speech (verb, noun, adjective, and adverb). Next we create a word-by-comment matrix with extracted words. This word-by-comment matrix, say $A$, is comprised of $m$ words $w_1$, $w_2$, ..., $w_m$ in $n$ comments $c_1, c_2, ..., c_n$, where the value of each cell $a_{ij}$ indicates the total occurrence frequency of word $w_i$ in comment $c_j$ [30].

To balance the effect of word frequencies in all the comments, log entropy term weighting is applied to the original word-by-comment matrix, which is the basis for all subsequent analyses [31]. We apply a global weighting function to each nonzero element of $A$. The global weighting function transforms each cell $a_{ij}$ of $A$ into a global term weight $g_i$ of $w_i$ for the entire collection of comments. Here $g_i$ is calculated as follows:

\[ g_i = \frac{1}{\sum_{j=1}^{n} a_{ij}} \]

\[ g_i = 1 + \sum_{j=1}^{k} (p_{ij} \log (p_{ij})/ \log (n)) \]  
\[ (1) \]
where \( p_{ij} = \frac{L_{ij}}{g_{fi}}, \) \( L_{ij} = \log (tf_{ij} + 1); \) \( tf_{ij} \) is the number of occurrences of \( w_i \) in \( c_j; \) \( g_{fi} \) is the number of occurrences of word \( w_i \) in all comments, and \( n \) is the number of all the comments in the collection.

### 4.2. Latent Semantic Analysis

Latent semantic analysis (LSA) is a well-known representative for semantic approaches. It identifies the hidden meaning of textual information in documents considering occurrences and co-occurrences of terms. Both terms and documents are mapped onto a semantic structure that consists of several semantic textual patterns [32]. To extract semantic patterns from comment data, we use a rank-validation procedure that is taken over from literature. The rank-validation procedure enables the identification of a maximal number of semantic patterns where each pattern can be used to represent a specific event. The rank-validation procedure is successfully evaluated by using LSA with singular value decomposition (SVD), a form of factor analysis. From the training comment data, we can get the term by comment matrix \( A \) \((m \times n)\), which means there are \( m \) distinct terms in \( n \) comments collection. The singular value decomposition of \( A \) is defined as

\[ A = USV^T \]
\[ (2) \]
where \( U \) and \( V \) are the matrices of the term vectors and document vectors. \( S = \text{diag} (r_1, \ldots, r_n) \) is the diagonal matrix of singular values [32]. To reduce the dimensions, we can simply choose the \( k \) largest singular values and the corresponding left and right singular vectors; the best approximation of \( A \) with rank-\( k \) matrix is given by:

\[ A_k = U_k S_k V_k^T \]
\[ (3) \]
where \( U_k \) is comprised of the first \( k \) columns of the matrix \( U \) and \( V_k^T \) is the first \( k \) rows of matrix \( V^T \). \( S_k = \text{diag} (r_1, \ldots, r_k) \) is the first \( k \) factors, the matrix \( A_k \) captures most of the important underlying structure in the association of terms and documents while ignoring noise due to word choice [33].

#### 4.2.1 LSA for new comments

When LSA is applied to a new comment: a query, a set of words in a query is represented as a vector in \( k \)-dimensional space. The query can be represented by

\[ q' = q^T U_k S_k^{-1} \]
\[ (4) \]
where \( q \) and \( q' \) are simply the vector of words in a new comment multiplied by the appropriate word weights and the \( k \)-dimensional vector transformed from \( q \), respectively. The sum of these \( k \) dimensional word vectors is reflected in the term \( q' U_k \) in the above equation. The right multiplication by \( S_k^{-1} \) differentially weights the separate dimensions. Thus the query vector is located at the weighted sum of its constituent term vectors [7].

To sum up, LSA has been applied to text categorization in many previous works. Zelikovitz and Hirsh [34] used LSA for text classification by adapting unlabeled data and other forms of available background text in the classification process. They used an expanded term-by-document matrix that includes both the labeled data as well as any available and relevant background text. The results showed empirically that the proposed method increased the accuracy rates in classification on a range of benchmark problems used in previous work. Yany [35] used SVD for noise reduction so as to improve the computational efficiency in text categorization. Also, Antai et al. [36] classified a set of documents according to document topic areas by using the CLUTO program with and without LSA. The results showed that the internal cluster similarity with LSA was much higher than without LSA.

In our research, we employ LSA to analyze patterns and relationships between the extracted words and latent concepts contained in the unstructured collection of texts (students’ comments). Our objective is to establish a strong relationship between analyzed comments and student grades in each lesson.

### 4.3. Artificial Neural Networks

Supervised ANNs have been widely used in areas of prediction. The wide range of applications of the ANN in many fields and sectors is due to its power to model behavior to produce an approximation of given output [37]. The advantages of neural networks over the conventional statistical analysis methods are as follows:

- Neural networks are good at modeling nonlinear relationships and interaction while conventional statistical analysis in most cases assumes a linear relationship between independent variables and dependent variables. Neural networks build their own models through a learning process, whether the relationships among variables are linear or not.
- Neural networks perform well with missing or incomplete data. A single missing value in regression analysis leads to removal of the entire observation or removal of the associated variable from all observations in the data set being analyzed. However, neural networks update weights between input, output, and intermediate nodes, so that even incomplete data can contribute to learning and produce desired output results [38, 39].

A three-layered perceptron was established in our research to estimate student grades. We constructed a network model for each lesson. The structure of the ANN is shown in Figure 3. Layer 1 of each network, which is the input layer, consists of a \( k \)-dimensional vector of LSA results that
characterize similarity between words. Layer 2 consists of one hidden layer; the number of neurons in the hidden layer depends on the number of attributes in the input layer. The number of hidden neurons is chosen heuristically because they showed the least error during the training of the data set with the class A and B data. Layer 3, the output layer, consists of 4 neurons denoting the student’s grade: S, A, B, and C.

The ANN was trained by Back Propagation (BP) which is based on the principle of gradient descent learning [8]. In our research, we used WEKA software\(^7\), to build our network models. Each of the networks was trained with more than 5,000 iterations to determine the predictive power. The convergence rate between the actual and desired output is achieved by 0.3 learning rate and 0.65 momentum coefficient. After performance was examined, iteration=5,000 showed the most predictive power in generalizing the problem for all lessons. The number of instances used in the training depends on the number of comments in each class.

4.4. Support Vector Machine

The support vector machine (SVM), a computationally powerful tool for supervised learning, is successfully and widely used in classification and regression for a variety of real-world problems like particle identification, text categorization, bioinformatics, and financial applications [40]. This technique attempts to separate two classes of data using a hyper-plane defined by support vectors, which are part of the data set. Through its training, the support vector machine technique searches for the Optimal Separating Hyper-plane (OSH); this is the optimal hyper-plane that maximizes the margin between the two classes of the training dataset. Though the standard SVM shows better generalization performance compared with many other machine learning methods, the training stage of SVM involves solving a quadratic programming problem (QPP) and is thus time-consuming. Its computational complexity is \(O(m^3)\), where \(m\) is the total size of training points. This drawback restricts its application to some real problems. Consequently, many improved variants have been proposed, e.g. Chunking, SMO, SVM Light, LSSVM, Libsvm, and Liblinear [9, 41]. In our research, we used Sequential Minimal Optimization (SMO), a fast method to solve huge quadratic programming problems and widely used to speed up the training of the SVM [42]. We employed the SMO method with RBF kernel and tolerance parameter 0.0001, to generate models to predict student grades by distinguishing four grades S, A, B and C from analyzed comment data.

![Figure 3. The structure of the ANN.](image)

\[^7\] http://www.kdd.org/explorations/issues/11-1-2009-07/p2V11n1.pdf
4.5. Methods of Predicting Student Grades

To predict student grades from their comments, 4 grade categories are used to classify students’ marks, as we mentioned in Section 3.2. The method considers that prediction is correct only if one estimated grade within the 4-grade categories is the actual grade of a student. We call this method the 4-grade prediction method. In addition to 4-grade categories, we use 7-grade categories so that we can allow the acceptance of a different grade adjacent to the original grade in 4-grade categories of a mark range, i.e., make one mark range correspond to two grades in 4-grade categories. We call this method overlap method.

Table 2 shows the correspondence relationships between the 4- and 7-grade categories and the range of student marks. For example, we assume a student’s mark is 87; the grade of the mark in 4-grade categories is A, and in 7-grade categories is AS; AS corresponds to two grades, A and S, in 4-grade categories. We adopt the overlap method because each grade in 4-grade categories corresponds to a range of 10 marks; learning status of students with the upper mark in a grade and others with the lower mark in its one upper grade are not so different from the point of view of the observing teacher. Therefore it is worth noting that handling the two adjacent grades as one grade sometimes helps teachers to grasp students’ real learning situations. For example, the mark range of grade AS is from 85 to 89, and that is closer to the lowest mark: 90 of grade S than the lowest mark: 80 of grade A.

Table 2. The correspondence between grades and the range of marks of 4- and 7-grade categories and the number of students in each grade.

(a) 4-grade categories

<table>
<thead>
<tr>
<th>Grade</th>
<th>Mark</th>
<th># Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>90-100</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>80-89</td>
<td>16</td>
</tr>
<tr>
<td>B</td>
<td>70-79</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>60-69</td>
<td>8</td>
</tr>
</tbody>
</table>

(b) 7-grade categories

<table>
<thead>
<tr>
<th>Grade</th>
<th>Mark</th>
<th># Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>90-100</td>
<td>10</td>
</tr>
<tr>
<td>AS</td>
<td>85-89</td>
<td>14</td>
</tr>
<tr>
<td>AB</td>
<td>80-84</td>
<td>2</td>
</tr>
<tr>
<td>BA</td>
<td>75-79</td>
<td>6</td>
</tr>
<tr>
<td>BC</td>
<td>70-74</td>
<td>4</td>
</tr>
<tr>
<td>CB</td>
<td>65-69</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>60-64</td>
<td>4</td>
</tr>
</tbody>
</table>

5. Experimental Results

In this section, we consider the prediction of final student grades from their comments. We evaluated prediction performance (F-measure, accuracy) by 2-fold cross validation. We constructed a model from comment data in one class. Then, as a test, we applied the model to comment data in the other class. We compared the predicted grade with the grade in the corresponding original data. The procedure was repeated in each lesson and the results were averaged. We run evaluation experiments by calculating Accuracy and F-measure as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

Table 3 shows the average overall prediction accuracy results from lesson 1 to 6 using the overlap method. To achieve highly accurate prediction results on the basis of student comments, we examined the different number of k-dimensions with two types of machine learning techniques, ANN and SVM, as shown in Figure 4. The average accuracy results from lesson 1 to 6 using the ANN model has the higher rate from k =18 to k =22. It scores 50%. The highest rate scores 54% with k =20. From lesson 7 to 15, the highest accuracy rate is taken in dimension 3, 4 and 5. It scores 47%. We also evaluate the average overall prediction accuracy results from lesson 1 to 6 using the SVM model. The higher rate scores an average 55%, with values from k =22 to k =26. The highest rate scores 57%, with k = 23. The highest prediction accuracy results from lesson 7 to 15 score 48%, with values from k =14 to k =18.

Table 3 shows the average overall prediction accuracy results obtained using ANN and SVM with 20- and 23-dimensions from lesson 1 to 6. The ANN model has the highest prediction accuracy results with 20-dimensions, scoring 55.6%. In addition, SVM has the highest prediction accuracy results with 23-dimensions. It achieved 57.4%. 

Let G be 4-grade categories (S, A, B and C) or 7-grade categories (S, AS, AB, BA, BC and C), X be a subset of G; let obs(s, X) be a function that returns 1 if the grade of student si is included in X, 0 otherwise, where 1 ≤ i ≤ n, n is the number of students; pred(si) be a function that returns a set of grade categories only including a predicted grade for student si; !pred(si) returns a complement of pred(si).

\[
\begin{align*}
\text{TP} &= \{si| \text{obs}(s, \text{pred}(si)) = 1\} \\
\text{FP} &= \{si| \text{obs}(s, \text{pred}(si)) = 0\} \\
\text{FN} &= \{si| \text{obs}(s, !\text{pred}(si)) = 1\} \\
\text{TN} &= \{si| \text{obs}(s, !\text{pred}(si)) = 0\}
\end{align*}
\]

\[
\begin{align*}
\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\
\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\
\text{F-measure} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
\text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}
\end{align*}
\]
Table 4 shows the average overall prediction results with 4- and 16-dimensions from lesson 7 to 15 obtained using ANN and SVM models, respectively. The prediction accuracy using ANN scored 47.3% with 4-dimensions. SVM has the highest prediction accuracy of 48.5% with 16-dimensions. Tables 5, 6, 7 and 8 show the average overall prediction results with the highest dimensions from lesson 1 to 15, after applying the overlap method to ANN and SVM models.

Table 3. Prediction results by number of dimensions from lesson (1-6).

<table>
<thead>
<tr>
<th># dimension</th>
<th>F-Measure ANN</th>
<th>Accuracy ANN</th>
<th>F-Measure SVM</th>
<th>Accuracy SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.503</td>
<td>0.556</td>
<td>0.523</td>
<td>0.542</td>
</tr>
<tr>
<td>23</td>
<td>0.483</td>
<td>0.523</td>
<td>0.554</td>
<td>0.574</td>
</tr>
</tbody>
</table>

Table 4. Prediction results by number of dimensions from lesson (7-15).

<table>
<thead>
<tr>
<th># dimension</th>
<th>F-Measure ANN</th>
<th>Accuracy ANN</th>
<th>F-Measure SVM</th>
<th>Accuracy SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.412</td>
<td>0.473</td>
<td>0.364</td>
<td>0.452</td>
</tr>
<tr>
<td>16</td>
<td>0.395</td>
<td>0.455</td>
<td>0.416</td>
<td>0.485</td>
</tr>
</tbody>
</table>

Table 5. Overlap prediction results by (4- Dimensions).

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>0.447</td>
<td>0.491</td>
<td>0.468</td>
<td>0.487</td>
</tr>
<tr>
<td>SVM</td>
<td>0.418</td>
<td>0.446</td>
<td>0.436</td>
<td>0.465</td>
</tr>
</tbody>
</table>

Table 6. Overlap prediction results by (16- Dimensions).

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>0.382</td>
<td>0.473</td>
<td>0.422</td>
<td>0.476</td>
</tr>
<tr>
<td>SVM</td>
<td>0.443</td>
<td>0.487</td>
<td>0.463</td>
<td>0.507</td>
</tr>
</tbody>
</table>

Table 7. Overlap prediction results by (20- Dimensions).

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>0.492</td>
<td>0.568</td>
<td>0.531</td>
<td>0.572</td>
</tr>
<tr>
<td>SVM</td>
<td>0.484</td>
<td>0.532</td>
<td>0.507</td>
<td>0.562</td>
</tr>
</tbody>
</table>

Table 8. Overlap prediction results by (23- Dimensions).

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>0.486</td>
<td>0.542</td>
<td>0.513</td>
<td>0.542</td>
</tr>
<tr>
<td>SVM</td>
<td>0.545</td>
<td>0.572</td>
<td>0.558</td>
<td>0.626</td>
</tr>
</tbody>
</table>

Figure 4 displays the average accuracy of the prediction of student grades obtained using the SVM model with 16- and 20-dimensions and the ANN model with 4- and 20-dimensions, which gave the highest accuracy after examining prediction results by varying dimension k from 2 to 30. The results from lesson 1 to 15 are evaluated using the overlap method. Figure 5, shows the accuracy results of the prediction of student grades from lessons 1 to 6, which are higher than those from lessons 7 to 15. The prediction results from lesson 1 to 6 are between 58% and 65% using the SVM model, and 52% and 63% using the ANN model. From lesson 7 to 15, the prediction accuracy results are between 37% and 57% using the SVM model, and 43% and 56% using the ANN model. The overall prediction results from lesson 1-6 are higher than those from lesson 7 to 15, and the prediction results from lesson 2, 4, 6, 9, 12 and 13 are more accurate than for other lessons.
In order to assess the contributions of interests and prestige to predicting each student’s grade, predictive discriminant analysis with different dimensions was performed. Figures 6 (a, b, c and d) display the prediction accuracy characteristics in each grade obtained by SVM and ANN models, with \( k \) dimensions varied from 2 to 30. We take S grade shown in Figure 6(a), as an example grade S has the highest top 3 prediction results using SVM with \( k \) dimension 10, 16 and 24 for lessons 1-6, and \( k \) dimension 12, 16 and 22 for lessons 7–15. Using the ANN model, for lessons 1 to 15, the prediction accuracy is highest with \( k = 22 \) for lessons 1-6 and \( k = 8 \) for lessons 7-15. According to Figure 6(b), grade A has the highest prediction results for lessons 1 to 6 using the ANN and SVM models. The top rates are between 14 and 22 dimensions. From Figure 6(c, d), it can be seen that the prediction accuracy for grades B and C are lower than for the S and A grades. The accuracy results are between 30% and 55% for lessons 1 to 15. There is no difference between the ANN and SVM models with grade B for lessons 1 to 15, but grade C achieves higher accuracy using ANN with \( k = 12 \) and SVM with \( k = 11 \) for lessons 1-6.

To sum up, the SVM model has better results for grades S, A and C than the ANN model. In addition, the results for grade A are the best among the 4 grades; a small number of comments has an impact on the prediction accuracy results. For example, grade A has the highest number of comments in all lessons and achieved the highest accuracy results using the ANN and SVM models.
Overall prediction results by grade with \( k \)-dimension from 2 to 30.

6. Conclusions and Future Work

The main contributions of this work are threefold. First, comment data were analyzed using the LSA technique; we calculated similarity between words using a comment matrix and detected noisy data by reducing the number of dimensions. Second, two methods of machine learning techniques: ANN and SVM were applied to analyze comments to predict student grades. We chose the ANN and SVM models because they are two popular strategies for supervised machine learning and classification, and it’s not clear which method is better for a particular problem. Third, the overlap method was proposed to accept two grades for one mark and establish a strong relation between LSA results and student grades. At the end, our proposed methods achieve 62.6% and 57.2% accuracy for predicting student grades from lesson 1 to 6, and 50.7% and 48.7% from lesson 7 to 15, after applying the overlap method, using SVM and ANN, respectively.

From the previous results, we can conclude that the difficulty of the subject influences the quality of the written comments; students wrote better comments while they were learning Computer Literacy from lesson 1 to 6 than while learning C programming from lesson 7 to 15; this tendency is reflected in the accuracy of the prediction of their grades.

As we mentioned previously, it was difficult to evaluate student grades by separately using one class as training data and the other class as test data: there was little similarity between the comment data in the two classes, a small number of comments badly affected prediction accuracy results, particularly with comments from the C and D grades due to students with those grades often not writing any comments or entirely absent from the lessons.

In the future, we will collect new comments from a large number of students to improve their performance. If we can collect their comments with sufficient care, they will include important information for each lesson and reflect student attitudes toward the subject. We believe these comments will help a teacher to estimate their learning situation more precisely and correctly, to give advice appropriate to the individual students, and to improve their performance as a result. In addition, student comment evaluation must be continued throughout the semester in order to continue delivering optimum advice to them.

References


Shaymaa E. Sorour She received a B.S. degree in Education Technology, at Faculty of Specific Education, Kafr Elsheikh University, Egypt and a M.S. degree in Computer Education at Department of Computer Instructor Preparation, Faculty of Specific Education, Mansoura University, Egypt, in 2004 and 2010 respectively. Since 2005 until now, she is working as assistant lecture at Department of Education Technology, Faculty of Specific Education, Kafr Elsheikh University, Egypt. She is currently a PhD student in Graduate School of Information Science and Electrical Engineering, Department of Advanced Information Technology, Kyushu University, Japan. She is a member of IEEE and IPSJ.

Tsunenori Mine He received a B.E. degree in Computer Science and Computer Engineering, in 1987, and a M.E. and D.E. degree in Information Systems, in 1989 and 1993, respectively, all from Kyushu University. He was an Assistant Professor at College of Education, Kyushu University, from 1992 to 1994 and at Department of Physics, Faculty of Science, Kyushu University from 1994 to 1996. He was a visiting researcher at DFKI, Saarbruecken, Germany from 1998 to 1999, and at Language Technology Institutes of CMU, Pittsburgh, PA, USA in 1999. Dr. Eng. Mine is currently an Associate Professor at Department of Intelligent Systems, Faculty of Information Science and Electrical Engineering, Kyushu University since 1996. His current research interests include Natural Language Processing, Information Retrieval, Information Extraction, Collaborative Filtering, Personalization and Multi-Agent Systems. He is a member of IPSJ, IEICE, JSAI, NLPSJ and ACM.

Kazumasa Goda He received a B.E. and a M.E. degree in 1994 and in 1996 from Kyushu University, respectively. He has been an Associate Professor at Kyushu Institute of Information Science since 2008. His research interests include Programming Theory, Programming Education, and Computer Education. He is a member of JSISE, JAEIS, and IPSJ.
Sachio Hirokawa  He received a B.S. and M.S. degree in Mathematics and Ph.D. degree in Interdisciplinary Graduate School of Engineering Sciences from Kyushu University in 1977, 1979, and 1992. He founded the search engine company Lafla (http://www.LafLa.co.jp) in 2008 as a university venture company. He has been working as an outside director of the company. He is currently a professor in the research institute for information technology of Kyushu University, since 1997. His research interests include search engine, text mining, and computational logic. He is a member of JSAI, IPSJ, IEICE and IEEE. He has been serving as a general chair of international conference AAI (Advanced Applied Informatics) since 2012.
Detection Methods for Misplacement of Synonyms in the Japanese WordNet

Takuya Hirao1 Takahiko Suzuki2 Koki Miyata3 Sachio Hirokawa4
Kyushu University, Japan

Abstract

Lexical database the Japanese WordNet is a useful tool in natural language processing. However, it is officially announced that the Japanese WordNet contains 5% errors. In this paper, we discuss error detection methods in the Japanese WordNet.

Keywords: WordNet, Thesaurus, ontology mapping, ontology debugging.

1. Introduction

The Japanese WordNet [1, 2] is a lexical database based on Princeton WordNet [3]. The Japanese WordNet is a useful tool in natural language processing and is used in various researches [4, 5]. Web thesaurus services for general users are available based on the Japanese WordNet. However, the current version of the Japanese WordNet (Wn-Ja 1.1) contains about 5% errors [1, 2]. Those errors might have influence to the usefulness of the Japanese WordNet.

In this paper, we discuss error detection methods in the Japanese WordNet. The error detection is the first step for correcting errors in the Japanese WordNet. The error detection techniques are beneficial in establishing reliable large scale lexical database. We will focus on the detection of a certain kind of errors in the Japanese WordNet. We call those errors ‘misplacements of synonyms’.

There are several projects for creating non-English WordNet or building lexical databases that are similar to WordNet [6]. The Japanese WordNet and Open Chinese WordNet [7] employ semi-automatic methods to map domestic language concepts to Princeton WordNet structures in their bootstrap stages.

2. WordNet and the Japanese WordNet

2.1 Princeton WordNet

Princeton WordNet [3] is a large lexical database of English. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms named as “Synset”, each expressing a distinct concept. Each Synset is distinguished by the unique ID. Gloss, which is a short text expressing the meaning of each Synset, is linked to the Synset. Synsets are interlinked by conceptual-semantic relations and by lexical token relations. WordNet can be used as a thesaurus because words are grouped in Synsets by their meanings. A word can be a member of two or more Synsets because the word can be polysemous.

There are attempts of integrating multilingual conceptual lexical databases by extending the WordNet, such as Universal WordNet [8], BabelNet [9], and [10]. Relationships between concepts and words (or between multiple concepts) are automatically extracted from various resources such as Wikipedia or tagged corpora. The qualities of the integrated results have been evaluated by the developers themselves and by network communities.

WordNet can be viewed as a kind of ontology [2]. When developing multilingual WordNet, it is necessary to create mappings between ontologies constructed in two or more languages. There are studies about detection and correction of errors in an ontology, and a mapping between ontologies. They tried to correct inappropriate taxonomy, redundant or incorrect mapping between the ontologies, or incorrect is-a relation in the ontology [11,12,13].

The rest of the paper is organized as follows: Section 2 introduces WordNet and the Japanese WordNet. In section 3, we introduce the concept and examples of “misplacements of synonyms” in the WordNet structure. Section 4 explains our methods for detecting misplacements. Section 5 describes the main result. In Section 6, we describe other interesting features of our methods such as an application to Princeton WordNet. Section 7 is discussion and about future works.

2. WordNet and the Japanese WordNet

2.1 Princeton WordNet

Princeton WordNet [3] is a large lexical database of English. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms named as “Synset”, each expressing a distinct concept. Each Synset is distinguished by the unique ID. Gloss, which is a short text expressing the meaning of each Synset, is linked to the Synset. Synsets are interlinked by conceptual-semantic relations and by lexical token relations. WordNet can be used as a thesaurus because words are grouped in Synsets by their meanings. A word can be a member of two or more Synsets because the word can be polysemous.

---

1. Weblio, http://thesaurus.weblio.jp/content/
2. Graduate School of Information Science and Electrical Engineering Fukuoka, Japan, hirao.1.007@s.kyushu-u.ac.jp
3. Research Institute for Information Technology, Fukuoka, Japan suzuki@cc.kyushu-u.ac.jp
4. Graduated School of Information Science and Electrical Engineering Fukuoka, Japan, ke13089s@s.kyushu-u.ac.jp
5. Research Institute for Information Technology, Fukuoka, Japan hirokawa@cc.kyushu-u.ac.jp

---

2.2 The Japanese WordNet

The Japanese WordNet is a lexical database of Japanese language based on Princeton WordNet. The purpose of the Japanese WordNet project is to provide a large scale Japanese lexical database. It has been developed since 2006.

The structure of the Japanese WordNet follows that of Princeton WordNet (Fig. 1). However, because of the difference between Japanese and English language, the Japanese WordNet contains original Synsets (concepts) [1,2]. In the Japanese WordNet, more emphasis is put on the comprehensiveness rather than the accuracy [1]. The current version of the Japanese WordNet (Wn-Ja 1.1) [2] covers:

- 57,238 concepts (the number of Synsets)
- 93,834 Japanese words
- 158,058 senses (Synset-word (synonym) pairs).

Each Synset that is linked to Japanese word(s) has a Japanese gloss (short definition, including optional Japanese example usages of synonyms). Other resources such as SUMO [14], Wikipedia, GoiTaikei [15], and some corpora are used to improve coverage of the Japanese WordNet.

2.3 WordNets in other languages and extensions of WordNet

There are several projects for creating lexical databases for various languages based on Princeton WordNet. Some try to create multilingual lexical databases which integrate Princeton WordNet, Wikipedia, Wiktionaryc and other resources.

Appropriate mapping between existing resources and the new database is the key issue that guarantees the integrity of the output database. Creating a WordNet structure in a new language can be viewed as creating a mapping from an existing ontology expressed in a language to a new ontology expressed in another language.


Error corrections are necessary for newly created ontology or mapping result between ontologies. The current version of the Japanese WordNet (Wn-Ja 1.1) contains about 5% errors. The error rate of Chinese Open WordNet [7] is comparable to that of the Japanese WordNet. In the rest of this paper, we will focus on a kind of errors which we will call misplacement of synonyms.

Definition 3.1 Misplacement of a synonym

Misplacement of a synonym in the WordNet structure is a word $w_{miss}$ in synonyms of a Synset (ID $S$) which does not meet the definition in the gloss (described in the same language as the $w_{miss}$) of $S$. We use $\text{mis}(S)$ in order to represent the set of all misplacements in $S$.

In Fig. 2, Synset 02651424-v is linked to 4 Japanese Synonyms: ポン（to give shelter for), 収容 (accommodate), 宿る (take shelter), and 持ち込む (carry in).

**Fig. 2. An example of a misplacement of a synonym**

Among them, the word 持ち込む is a misplacement in 02651424-v. Although 持ち込む can be translated into lodge in some situationsd, lodge does not mean “provide housing for” in those contexts. Another word 宿る is seldom used with the meaning of “provide housing for”. If 宿る is a misplacement or not is arguable.

---

d The Synset which contains both lodge and 持ち込む as synonyms and has a proper gloss does not exist in the Japanese WordNet. This topic is out of the range of this paper.
3.1 Pilot study

We had manually checked the misplacements in the Japanese WordNet. Because it is reported that errors in a WordNet structure occur more often in verbs than in nouns [7], we chose misplacements in verbs as the main target.

We checked the misplacement by the following procedure.

1) We selected 900 verbs randomly from the intersection of the vocabulary of Japanese Language Proficiency Test (JLPT)\(^a\), and words registered as verbs in the Japanese WordNet.
2) All the Synsets containing the aforementioned verbs as synonyms are extracted from the Japanese WordNet.
3) The screening person checked all the Synset-verb pairs. He made the list of candidates of misplacement.
4) People as checkers independently evaluated the candidates. When all checkers' results coincide, we marked the Synset -word pair as a misplacement.

3.2 Patterns of Misplacements

Among 900 words, 82 words (approx. 9%) were judged as misplacements in some Synset-word pairs. Misplacements can be classified into three types.

We use \(\text{syn}(S)\) to denote the set of all Japanese synonyms in Synset \(S\).
- When \(\text{syn}(S)\) is not a singleton and \(\text{syn}(S)=\text{mis}(S)\), we call that \(S\) has an all-wrong type misplacement.
- When \(\text{syn}(S)\) \(\setminus \text{mis}(S) \neq \emptyset\) and \(\text{mis}(S) \neq \emptyset\) (there exists proper synonym(s) other than the misplacement(s)), we call that Synset \(S\) has partially wrong type misplacement(s). Fig. 2 is an example of a partially wrong type misplacement.
- When \(\text{syn}(S)\) is a singleton and \(\text{syn}(S)=\text{mis}(S)\), we call that \(S\) has a solo type misplacement.

Among 82 misplacements, there are 26 partially wrong, 27 all-wrong, and 28 solo type misplacements. We count only those misplacements that directory link the 82 words and a Synset.

4. Detection Methods for Misplacements

We tried to detect misplacements by using the information in the Japanese WordNet itself. Our methods utilize rather limited information and can be used in any WordNet structures other than the Japanese WordNet.

4.1 Detection by Synset-synonym links

The first detection method uses only the link information between Synset and synonym. We define Synset-coverage of word \(w\) as follows.

**Definition 4.1 Synset coverage**

For word \(w\) and Synset \(S\), Synset-coverage (SC) of \(w\) is

\[
\text{SC}(w) = \{S \mid w \in \text{syn}(S)\}
\]

Synset coverage \(\text{SC}(w)\) is the set of all Synset-IDs that are linked to the word \(w\). Fig. 3 is an example \(\text{SC}(売る)\).

Next we put the following hypothesis in order to use \(\text{SC}(w)\) for misplacement of synonym detection.

**Hypothesis 4.2**

Assume \(w_1, w_2 \in \text{syn}(S)\) \((w_1\) and \(w_2\) are synonyms in \(S)\).
\(\text{SC}(w_1) \cap \text{SC}(w_2)\) tends to have more elements than that of \(\text{SC}(w_1) \cap \text{SC}(w_k)\), where \(w_k\) is not a synonym in \(S\).

Fig. 1, words 商う, 売り買い, 売買, and 売る have linked to both Synset 002260362-v (English gloss: do business) and 002244956-v (offer for sale as for one's livelihood). Thus any pair of \(\text{SC}(売る)\), \(\text{SC}(売り買い)\), \(\text{SC}(売買)\) or \(\text{SC}(売る)\) has 2 or larger intersection elements.

For misplacement \(w_{\text{mis}} \subseteq \text{mis}(S)\) and \(w_k \in \text{syn}(S)\) and \(w_k \neq w_{\text{mis}}\), \(\text{SC}(w_{\text{mis}}) \cap \text{SC}(w_k)\) will tend to have smaller number of elements under the Hypothesis 4.2. Then it becomes possible to distinguish misplacement of synonyms in Synset \(S\) by comparing \(\text{SC}(w_j) \cap \text{SC}(w_k)\) \((w_j, w_k \in \text{syn}(S))\) for all \((w_j, w_k)\) pair in \(\text{syn}(S)\).

---

\(^a\) Japanese Language Proficiency Test, [http://www.jlpt.jp/e/index.html](http://www.jlpt.jp/e/index.html)
4.2 Detection by gloss-synonym link

The second method utilizes words in gloss (definition). We define gloss coverage as follows:

**Definition 4.3 Gloss coverage**

Let, the set of all Japanese words in a gloss and example sentences of Synset $S$ be $\text{gloss}(S) = \{w_1, w_2, \ldots \}$. For word $w$, gloss-coverage (GC) of $w$ is

$$GC(w) = \{u_1, u_2 \in \text{gloss}(S), S_j \in SC(w)\}$$

It is self-evident when $SC(w_1)$ and $SC(w_2)$ have common elements, $GC(w_1)$ and $GC(w_2)$ also have common elements. Gloss overlap will show similar behavior to that of Synset overlap. Are there any merit using GCs instead of SCs?

We put the following hypothesis that suggests superiority of $GC(w)$ over $SC(w)$.

**Hypothesis 4.4**

Assume $w_1, w_2 \in \text{syn}(S)$. We will call Synset $S_j \subseteq SC(w_1)$ is incompatible to $w_2$ when $S_j \notin SC(w_2)$. Even when $S_j$ is incompatible to $w_2$, $S_j$ is incompatible to $w_1$, $\text{gloss}(S_j) \cap \text{gloss}(S_i)$ tends to have more elements than $\text{gloss}(S_j) \cap \text{gloss}(S_i)$, where $S_i$ is an arbitrary Synset.

Under Hypothesis 4.4, non-common element in $SC(w_1)$ or $SC(w_2)$ have an influence on gloss overlap. For example, both words 引っ張る (approximately meaning: pull) and 延ばす (approximately elongate) are paired to Synset 00317888-v. While 01592456-v $\in SC(引っ張る)$ and 00318816-v $\notin SC(延ばす)$, while 01592456-v $\notin SC(延ばす)$, or 00318816-v $\in SC(引っ張る)$.

On the other hand, both $\text{gloss}(01592456-v) \subseteq GC(引っ張る)$ and $\text{gloss}(00318816-v) \subseteq GC(延ばす)$ contain the common word 引く (See Fig. 4).

4.3 Procedure

The followings 1) to 4) describe the detection procedure for misplacements of synonym by using $SC(w)$ (Synset-coverage).

1) Make a list of Synsets to which the detection method will be applied.
2) For each Synset $S_i$ in the list, for every pair of $w_k, w_j \in \text{Syn}(S_i)$, calculate the following Synset Overlap Ratio $SOL(w_k, w_j)$.
   $$SOL(w_k, w_j) = \frac{\#(SC(w_j) \cap SC(w_k))}{\#(SC(w_j) \cup SC(w_k))}$$
   ($\#(S)$ denotes a cardinality of a set $S$).
3) Take the minimum value of $SOL(S_i)$
   $$mSOL(S_i) = \min(SOL(w_k, w_j)) \ (\text{for all } w_k, w_j \in S_i)$$
4) Flag a Synset $S_i$ having misplacement if $mSOL(S_i) < \gamma$.
   $\gamma$ is a threshold value here ($\gamma < 1$).

The detection procedure using $GC(w)$ can be defined by replacing $SOL(w_k, w_j)$ to $GC(w_k, w_j)$ ($GOL(w_k, w_j)$) (Gloss Overlap Ratio), and $mGOL(S)$ (minimum value of gloss overlap in Synset $S$).

4.4 Applicability of the methods

To apply the procedures in 4.3, a Synset must satisfy the condition.

**Condition 4.5**

There must be 2 or more Japanese polysemous synonyms paired to the Synset.

The procedure cannot be applied to detect solo type misplacements.

It should be emphasized that our hypotheses are not always appropriate. For instance, two words 切る (cut) and 混ぜる (mix) are synonyms in Synset 01418667-v (Eng. gloss: mix so as to make a random order or arrangement; “shuffle the cards”) but not synonyms in other Synsets.

5. Result for ‘Partially Wrong’ Type in the Japanese WordNet

There are 10,321 Synsets that represent verbs in the Japanese WordNet. Among them, we find that 3,031 Synsets satisfy Condition 4.5. For those Synsets, two testers checked the misplacements in the Synsets. They also checked if the misplacement was partially-wrong or all-wrong. We classify Synsets at least one of the checkers marked as all-wrong as all-wrong, and other Synsets at least one of the checkers marked partially-wrong as partially-wrong. It resulted in 125 all-wrong misplacements and 121 partially-wrong type misplacements in 3,031 Synsets.
Fig. 5 and Fig. 6 indicate Precision, Recall, and F value for \( mSOL \) and \( mGOL \) \( (0 < \gamma < 1) \) for partially wrong type misplacements. F values become maximum when \( \gamma = 0.0526 \) for \( mSOL(S) \) and \( \gamma = 0.0455 \) for \( mGOL(S) \). TABLE 1 shows the details. Fig. 7 shows the same result as Fig. 6. The horizontal axis is changed to the ranks of Synsets ordered by \( mGOL(S) \) value. Vertical line in Fig. 7 indicates the point where F becomes maximum value.

**TABLE 1. The results when F value is max**

"Synset rank" indicates the total number of Synsets below the threshold value

<table>
<thead>
<tr>
<th></th>
<th>Synset rank</th>
<th>( \gamma )</th>
<th>Precision</th>
<th>Recall</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>( mSOL )</td>
<td>331</td>
<td>0.0526</td>
<td>0.195195</td>
<td>0.481481</td>
<td>0.277778</td>
</tr>
<tr>
<td>( mGOL )</td>
<td>313</td>
<td>0.0476</td>
<td>0.191693</td>
<td>0.495868</td>
<td>0.276498</td>
</tr>
</tbody>
</table>

6. Detailed Results

6.1 Comparison of \( mSOL(S) \) and \( mGOL(S) \) (Synset or gloss)

Although \( mGOL(S) \) shows higher precision in lower \( \gamma \), the performance difference between \( mSOL(S) \) (minimum Synset overlap) and \( mGOL(S) \) (minimum gloss overlap) is not evident from Fig. 5, 6, 7, or TABLE 1. \( mSOL(S) \) and \( mGOL(S) \) show the clear correlation as discussed in Section 5.

Fig. 8 shows \( mSOL-mGOL \) correlation. We extracted two sets of 500 Synsets which have the lowest \( mSOL \) (S) or \( mGOL \) (S) \( (S_{min500} \text{ or } G_{min500}) \). Both \( S_{min500}/G_{min500} \) and \( G_{min500}/S_{min500} \) (difference sets) have 68 elements.

In the difference set \( S_{min500}/G_{min500} \) (Synsets which have low Synset overlap, but have higher gloss overlap), there are 0 partially wrong type misplacements. On the other hand, there are 5 partially wrong type misplacements in \( G_{min500}/S_{min500} \).

We also checked \( mGOL(S)/mSOL(S) \) values in \( G_{min500} \) and \( S_{min500} \). There are several Synsets in \( G_{min500} \) and \( S_{min500} \) which have \( mGOL(S)/mSOL(S) > 1 \). However, as shown in TABLE 2, those Synsets with
mGOL(S)/mSOL(S) > 1 in Gmin500 or Smin500 are not partially wrong type misplacements (only with the single exception).

Comparison of Gmin500 and Smin500 suggests mGOL(S) is a better indicator for partially wrong type misplacements.

### TABLE 2. Number of Synsets with higher mGOL relative to mSOL in Gmin500, Smin500, and in all

<table>
<thead>
<tr>
<th></th>
<th>A: Partially wrong misplacements</th>
<th>B: mGOL/mSOL&gt;1</th>
<th>A ( \cap ) B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gmin500</td>
<td>68</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Smin500</td>
<td>63</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>All</td>
<td>135</td>
<td>431</td>
<td>4</td>
</tr>
</tbody>
</table>

#### 6.2 All-wrong type misplacement

Although all-wrong type misplacement ratio seems to decrease as mGOL(S) increases, we could not detect any significant results for all-wrong type misplacements.

We found either of the following two patterns among all-wrong type misplacements:

1) The following conditions are fulfilled simultaneously.
   - English synonym(s) in the Synset is a general verb (such as get or put).
   - The number of the English synonym(s) in the Synset is 1 or a small number.
   - The use of the synonym(s) in the sense of the English gloss is an uncommon case.
   - The Japanese synonyms for the Synset seem to be selected based on the generic meaning of the English synonym(s).

2) There is a mistranslation in the Japanese gloss. Thus, it does not match the meaning of the synonyms.

For all-wrong type misplacements with pattern 1), our methods may be extended by using the English synonyms as well as the Japanese synonyms.

Our methods will not be useful in detecting all-wrong type misplacements with pattern 2). In such cases, all the synonyms are appropriate and only the Japanese gloss is incorrect.

#### 6.3 False low mSOL (and mGOL)

There are Synsets with low mGOL(S) or mSOL(S) but don’t have a misplacement in their synonyms. For 392 Synsets in Gmin500, we could not manually detect any misplacement.

The low mGOL(S) and mSOL(S) occurs when the following condition is satisfied. For the sake of simplicity, we use mSOL(S) instead of mGOL(S).

**Condition 6.1**

In Synset \( S \), there are two synonyms \( w_1, w_2 \in \text{syn}(S) \). \( \text{SC}(w_1) \cup \text{SC}(w_2) \) contains a large number of Synsets (say 15, it means that \( w_1 \) and/or \( w_2 \) are highly polysemous). On the other hand, \( \text{SC}(w_1) \cap \text{SC}(w_2) \) contains one or a few elements.

The following definition is useful in order to explain the features of false low mSOL and mGOL.

**Definition 6.2**

When \( mSOL(S) = \text{SOL}(u,X) \ (u,X \subseteq S) \), we call that \( u \) appears in \( mSOL(S) \). Similarly, we call that \( v \) appears in \( mGOL(S) \) when \( mGOL(S) = \text{GOL}(v,X(v),X \subseteq S) \).

In 100 Synsets which have the lowest \( mSOL(S) \) (Smin100), same words appear repeatedly in \( mSOL(S) \). Table 3 shows most frequently appearing 4 words in Smin100.

#### 6.4 Mispacements with high mSOL (and mGOL)

There are Synsets with partially wrong type misplacements but have high mGOL(S) or mSOL(S). For the sake of simplicity again, we use mSOL(S).

There are 29 misplacements of which \( mSOL(S) \geq 0.1 \). We found that in 25 cases, for words \( w_1 \) and \( w_2 \) that appear in \( mSOL(S) \), \( \text{SC}(w_1) \cap \text{SC}(w_2) \) is the singleton. That means Hypothesis 4.2 holds in those 25 cases. The cause of high mSOL(S) is the small cardinality of the denominator \( \text{SC}(w_1) \cup \text{SC}(w_2) \) in those cases. Fig. 9 shows an example of misplacements with high mSOL(S), where one of the word 煎じる is monosemous (SC(煎じる) is a

### TABLE 3. Most frequently appearing 4 words in mSOL(S) within Smin100

<table>
<thead>
<tr>
<th>Times of appearing</th>
<th>Partially wrong</th>
<th>#SC(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>鉤る(cut)</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>掛ける(cover)</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>考える(think)</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>炊ける(become)</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

#SC(w) is the number of Synsets in SC(w)
Japanese word 茹る (boiled) can be used in the meaning of “be in an agitated emotional state” (SynID: 01767612-v) but 煮込む (boil) does not have such meaning.

Fig. 9 Sample of a misplacement with \( mSOL(S) \geq 0.1 \)

The other partially wrong type misplacement example of which \( mSOL(S) \geq 0.1 \) is 02651424-v shown in Fig. 2. The maximum cardinality of \( SC(w_j) (w_j \in syn(02651424-v)) \) is 5, \( \#(SC(w_j) \cup SC(w_i)) \) cannot be greater or equal to 10, therefore \( mSOL(02651424-v) \) cannot be below 0.1. \( SC(持ち込む) \cap SC(収容) \) (持ち込む,収容 \( \equiv \) syn(02651424-v)) is a singleton and 載込む is manually flagged as a misplacement.

### 6.5 Misplacements in Japanese Nouns

We have applied our methods to nouns in the Japanese WordNet as well. We manually checked misplacements only in \( Gmin500n \) (500 noun Synsets with lowest \( mGOL(S) \)). Although the results are not so significant compared to the results in Japanese verbs, there are correlations between the value of \( mGOL(S) \) and occurrence of partially wrong type misplacements.

There are 8,354 noun Synsets to which our methods are applicable in the Japanese WordNet. We have detected 17 partially wrong type misplacements in \( Gmin500n \). There are 2 all-wrong misplacements in \( Gmin500n \).

11 out of 17 partially wrong type misplacements are in the range of \( Gmin100n \) (100 Synsets with lowest \( mGOL(S) \)). In \( Gmin100n \), same words appear frequently in the \( mGOL(S) \). TABLE 4 shows most frequently appearing 4 words in \( Gmin100n \). As shown in TABLE 4, 4 out of 11 partially wrong type misplacements occurs in the Synsets which have either 線 or 変化 as a synonym.

In 3,224 Synsets of nouns \( mGOL(S) \) values are \( \geq 0.3333 \), and in 1,276 Synsets \( mGOL(S) \) values are 1. \( mGOL(S)=1 \) means that every synonyms in a Synset are linked to the same other Synsets.

### 6.6 Application to Princeton WordNet

We have applied our methods to Princeton WordNet 3.0 which is integrated into the Japanese WordNet. We compute \( mGOL_{eng}(S) \) using English gloss for 1,848 verb Synsets (all verb Synsets in Princeton WordNet 3.0 which satisfy Condition 4.5).

We found that several small values of \( mGOL_{eng}(S) \) are comparable to the values associated to misplacements in the Japanese WordNet. \( mGOL_{eng}(S) \) for 12 Synsets are smaller than the smallest value of the \( mGOL(S) \) in Japanese WordNet verb Synsets. We checked those 12 Synsets and found that there are rare examples of synonymous verbs in 3 Synsets among them.

The words make, get, or cover is rarely used in the sense of the glosses in TABLE 5. A native speaker who received professional education could not recognize those synonyms as correct ones unless consulting English dictionaries and the example sentences in Princeton WordNet. make in Synset 00072012-v has the smallest \( mGOL_{eng}(S) \) values in all English verb Synsets. get in Synset 00054628-v has the second smallest \( mGOL_{eng}(S) \). Our method can be used to find rare examples in the WordNet structure.

### TABLE 4. Most frequently appearing 4 words in \( mGOL(S) \) within \( Gmin100n \)

<table>
<thead>
<tr>
<th>Times of appearing</th>
<th>Partially wrong</th>
<th>#SC(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>線 (chance or relationships)</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>変化 (change or variety)</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>仲間 (party)</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>節 (season or constancy)</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

### TABLE 5. Rarely used Synonyms with low \( mGOL_{eng}(S) \) in Princeton WordNet

<table>
<thead>
<tr>
<th>Synset ID</th>
<th>Gloss and Examples</th>
<th>Rarely used synonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>00072012-v</td>
<td>eliminate urine; &quot;Again, the cat had made on the expensive rug&quot;</td>
<td>make</td>
</tr>
<tr>
<td>00054628-v</td>
<td>make children; &quot;Abraham begot Isaac&quot;; &quot;Men often father children but don't recognize them&quot;</td>
<td>get</td>
</tr>
<tr>
<td>00060185-v</td>
<td>sit on (eggs); &quot;Birds brood&quot;; &quot;The female covers the eggs&quot;</td>
<td>cover</td>
</tr>
</tbody>
</table>
The following TABLE 6 is the list of other 9 verb-Synsets that have low \textit{mGOL}_{eng}(S). It helps understanding of the features of false low \textit{mGOL}(S) discussed in Section 6.3.

<table>
<thead>
<tr>
<th>Synset ID</th>
<th>Paired words in \textit{mGOL}_{eng}(S)</th>
<th>Eng. gloss and examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>00339464-v</td>
<td>break, recrudesce</td>
<td>happen; “Report the news as it develops” “These political movements recrudesce from time to time”</td>
</tr>
<tr>
<td>02318165-v</td>
<td>break, bankrupt</td>
<td>reduce to bankruptcy; “My daughter's fancy wedding is going to break me!” “The slump in the financial markets smashed him”</td>
</tr>
<tr>
<td>02539788-v</td>
<td>raise, parent</td>
<td>bring up; “raise a family” “bring up children”</td>
</tr>
<tr>
<td>00779061-v</td>
<td>break, interrupt</td>
<td>cease an action temporarily; “We pause for station identification” “let's break for lunch”</td>
</tr>
<tr>
<td>00421691-v</td>
<td>pass, evanesce</td>
<td>disappear gradually; “The pain eventually passed off”</td>
</tr>
<tr>
<td>02072849-v</td>
<td>pass, elapse</td>
<td>pass by; “three years elapsed”</td>
</tr>
<tr>
<td>00442669-v</td>
<td>set, jell</td>
<td>become gelatinous; “the liquid jelled after we added the enzyme”</td>
</tr>
<tr>
<td>00339464-v</td>
<td>set, coiffe</td>
<td>arrange attractively; “dress my hair for the wedding”</td>
</tr>
<tr>
<td>02285392-v</td>
<td>carry, stockpile</td>
<td>have on hand; “Do you carry kerosene heaters?”</td>
</tr>
</tbody>
</table>

7. Discussion and Future Works

In this paper, we introduced two simple methods to detect “misplacement of synonyms” in WordNet structures. They gave fair results for “partially wrong” type misplaced in the Japanese WordNet. By using gloss, the result was improved. When the threshold value is optimized, about 50% of partially wrong misplacements in 313 Synsets having “minimum gloss overlap value” exist below the threshold (TABLE 1, Fig. 7). Our methods do not require any external information, thus can be applied to any WordNet structures other than the Japanese WordNet. Our methods give baseline measures for advanced studies.

We have shown various features and limitations of our methods. We have presented Condition 4.5 to which our methods are applicable. Our methods are not applicable to, for example solo type misplacements. Semantic relation such as hypernym and hyponym is useful to extend our methods. By using the set of sibling words instead of the set of synonyms in a Synset we can define sibling-Synset overlap or sibling-gloss overlap. Condition 4.5 will be relaxed by using sibling words.

All-wrong type misplacements can be divided into two categories as discussed in Section 6.1. Our methods will be of no use for all-wrong type misplacements resulting from a mistranslation of English glosses. However, all-wrong type misplacements resulting from mapping of generic meaning of English synonym(s) into Japanese synonyms should be detectable if we use the English synonym(s) in our methods with some modifications. English dictionary, Japanese Dictionary, and English-Japanese translation tool will be necessary to detect mistranslation in Japanese glosses. Links between synonyms and gloss in WordNet structure can be utilized in detection procedures.

In Section 5.4, we have pointed out the problem of misplacements with high \textit{mSOL} (or \textit{mGOL}). The main cause of the problem is the small cardinality of \textit{SC}(w_j) \cup \textit{SC}(w_k), which is the denominator of \textit{mSOL}(S). The problem may be relaxed if we use \textit{#}(\textit{SC}(w_j)) \subseteq \textit{syn}(S) as the denominator (the nominator must be modified accordingly).

The use of external corpora instead of \textit{SC}(w) or \textit{GC}(w) is promising in order to increase the cardinality of the denominator. A corpus without sense tags is sufficient for this purpose because the polysemy of the synonyms in the target Synset is limited. We already have done the preliminary experiment using an external corpus. But using the set of all nearby words of target \textit{w} occurring in the corpus as \textit{SC}(w) did not produce any significant results. Classification of words [16] or some kind of feature selection [17] is necessary.

The problem of false low \textit{mSOL}(S) (or \textit{mGOL}(S)) was shown in Section 6.3 and 6.6. The special treatment of generic words with high polysemy (\textit{#SC}(w) > 15, for example) will be effective to reduce the number of false low \textit{mSOL}(S). The result of Princeton WordNet in Section 6.6 shows that the false low \textit{mGOL}(S) (or \textit{mSOL}(S)) can be an indication of rarely used synonyms. It is reasonable to manually add an example of sentences in the gloss of the WordNet structure for each word \textit{w} with high \textit{#SC}(w) and at the same time appearing in the paired words of \textit{mGOL}(S).

The use of external corpora for the false low \textit{mSOL} (S) (and \textit{mGOL}(S)) problem is a challenging task. We cannot expect correctly sense tagged corpora for misplaced synonyms. Large untagged corpora will contain any rare usage of synonyms, but there will be the risk of overfitting [18].

Acknowledgment

We thank Mr. Brendan Flanagan for his assistance in checking the results of the experiment on Princeton WordNet in Section 6.6.

References


Takuya Hirao is a graduate student of Information Science and Electrical Engineering, Kyushu University. His interests include text mining, ontology mapping, and debugging of ontology.

Takahiko Suzuki received a B.S. degree from Kyoto University in 1981, M.S. and Dr. of Engineering degrees in Faculty of Engineering from Kyushu University in 1988 and 1991. He is an associate professor in the research institute for information technology of Kyushu University. His research interests include text mining, ontology, databases, location information, and educational technology.

Kohki Miyata is a graduate student of Information Science and Electrical Engineering, Kyushu University. His interests include natural language processing, language learning, ontology and word sense disambiguation.
Sachio Hirokawa He received a B.S. and M.S. degree in Mathematics and Ph.D. degree in Interdisciplinary Graduate School of Engineering Sciences from Kyushu University in 1977, 1979, and 1992. He founded the search engine company Lafla (http://www.Lafla.co.jp) in 2008 as a university venture company. He has been working as an outside director of the company.

He is currently a professor in the research institute for information technology of Kyushu University, since 1997. His research interests include search engine, text mining, and computational logic. He is a member of JSAI, IPSJ, IEICE and IEE. He has been serving as a general chair of international conference AAI (Advanced Applied Informatics) since 2012.
Development of a Digital Signage and MR Contents to Promote Hand Hygiene

Akinori Kanazawa\textsuperscript{a}  
Naruto University of Education, Japan

Toshiko Asai\textsuperscript{b}  
Information Workshop, co., Japan

Akinori Minazuki\textsuperscript{c}  
Kushiro Public University, Japan

Hidehiko Hayashi\textsuperscript{d}  
Naruto University of Education, Japan

Abstract

Hand hygiene is effective way of prevention of infectious disease such as influenza. The practice of hand hygiene is performed using an alcohol-based hand rub or by washing the hands with soap and water. However, even healthcare workers’ hand hygiene compliance rate is lower. Hand hygiene is important not only health care workers, even the common man. Therefore, we investigate the current status of hand hygiene by interview and questionnaire in cafeteria of Naruto University of Education. As a result, the implementation rate is low. In this study, we consider the results of the investigation, we developed a digital signage to improve hand hygiene implementation rate. Furthermore, we developed MR (Mixed Reality) contents using projection mapping to give experience the effect of hand hygiene. The MR contents inform the virus that exist around projection mapping to give experience the effect of hand hygiene. The MR contents give experience of the effect of hand hygiene by disappear the image of virus with use an alcohol-based hand rub. As a result of installing the MR system, hand hygiene implementation rate was improved by 8.5% on average.

Keywords: Hand hygiene, Mixed Reality, Digital signage, Projection mapping, Prevention of infections.

1. Introduction

Being an easy and effective way to prevent contagious infection mediated by hand contact, hand hygiene is important on a daily basis, not only during the annual influenza epidemic period\cite{1}\cite{2}. The practice of hand hygiene is performed using an alcohol-based hand rub or by washing the hands with soap and water. It is important to clearly and correctly convey such knowledge on infectious diseases prevention and promote hand hygiene practices. As such, various measures have been put in place. For example, the government and municipalities have created posters to promote hand hygiene and made available an alcohol-based hand rub at public facilities to promote public awareness of hand hygiene. In addition, for the sake of public health and to prevent the spread of infectious diseases, the government has enacted laws to prevent epidemics, whereby patients are quarantined or limitations are placed on their attendance at work or school. The importance of this system and the prevention of epidemics are widely recognized. WHO asserts that hand hygiene education is important \cite{3}\cite{4}\cite{5}\cite{6}\cite{7}\cite{8}\cite{9}\cite{10}\cite{11}\cite{12}. However, awareness of hand sanitation is low even among healthcare providers, of whom the practice of more appropriate infection control techniques is required, and the hand hygiene compliance rate of such workers is known to be low.

Hand hygiene is an important practice not only for health care providers, but also for the general public. Schools and other places where many people gather are closely linked to the spread of infections. Strict infection control is required in schools because classes may be cancelled in severe circumstances.

In this study, we conducted an opinion survey on hand hygiene by interview and questionnaire on people who are not health care workers in a cafeteria at Naruto University, and revealed the current status of hand hygiene there. We considered the results and went on to develop a digital signage to promote hand hygiene with the aim of installing it in a place such as a school where people gather. Furthermore, we developed a MR contents using projection mapping to experiential understand hand hygiene for the digital signage.

In this paper, Section 2 describes the objective of this study. Section 3 gives consideration to the results of the opinion survey on hand hygiene. Section 4 describes development of the digital signage to promote hand hygiene. Section 5 describes development of the MR contents for the digital signage. Section 6 evaluates the MR system. Section 7 gives consideration to the whole of the present study and Section 8 provides a summary.

2. Object

The present study has the following two objectives.

To develop a digital signage and the contents to promote hand hygiene based on the current status of hand hygiene. In this study, the current status of hand hygiene is the hand hygiene implementation rate, the expectation to hand hygiene, and reason of not perform hand hygiene.
3. Opinion survey on hand hygiene

An opinion survey on hand hygiene was conducted to develop a system to improve hand hygiene.

3.1 Survey method

The survey consisted of a questionnaire and interviews conducted on 93 students and faculty member of Naruto University of Education in the Naruto university cafeteria.

3.2 Survey items

The survey items were as follows.
Q1 What is your position at the university?
Q2 Have you cleaned your hands with sterilizing alcohol or washed them with soap (or liquid soap) and water since entering the cafeteria?
Q3 (Only to those who responded "3. I washed my hands with water only." in Q2) Why didn't you use soap to wash your hands? (Free response)
Q4 (Only to those who responded "3. I have not washed my hands." in Q2) Why didn't you wash your hands? (Free response)
Q5 Do you think hand-washing helps prevent infectious diseases such as influenza?
Q6 What kind of support (or education) do you think would support the promotion of hand washing? (Free response)

The results are described in the following section.

3.3 Survey results

Q1 inquired into the position of the respondents. Figure 1 shows the results.

The Naruto University of Education is composed of 1330 people with student, faculty member, and officer. There are more postgraduates than undergraduate students at the Naruto University of Education because it is positioned as a "University for teachers" and there are many teaching professionals enrolled at the university.

Q2 confirmed whether the respondents washed their hands or not. Figure 2 shows the results.

![Figure 1 Results of Q1](image1.png)

![Figure 2 Results of Q2](image2.png)
The result of Q2 was that 5.4% cleaned their hands with rubbing alcohol, 8.7% with soap and water, 10.9% with water only, and 75.0% did not clean their hands at all.

Q3 investigated the reason why the respondent washed their hands but did not use soap. The survey allowed a free response. As the result, reason for motivation such as “It is troublesome”, reason for awareness such as “I don’t do it always”, and reason for environmental facilities such as “I don’t know that there is alcohol and soap” were obtained.

Q4 investigated the reason why those respondent who did not wash their hands elected not to do so. Q4 also allowed a free response. As the result, reason for motivation such as “It is troublesome” and “I had forgotten”, reason for awareness such as “I don’t think I needed a wash” and “There was no visible soil on my hands”, and reason for environmental facilities such as “The hand washing station is too far” and “I don’t know that there is hand washing stations” were obtained. In addition, reason for method of hand hygiene such as “It takes time” was obtained.

Q5 investigated the expectations of the respondent with regards to hand hygiene. Figure 3 shows the results.

The result of Q5 was that 75.0% of the respondents have a sense of expectation concerning the effect of hand hygiene, whereas 25.0% do not.

Q6 investigated the need for support for hand hygiene. The survey allowed a free response. This question elicited responses concerning methods of promoting hand hygiene. The responses indicated methods financial compensation for hand hygiene, such as "A ¥10 discount on items on the menu," or "Money for hand hygiene"; methods of illustrating the basis of the effectiveness, such as "Show the effectiveness of hand hygiene," "Provide data on the reduction in cases of influenza due to hand washing," or "Play a video showing how much hand washing decreases germs"; methods relating to the facilities, such as "Increase the number of hand washing stations," "Increase the number of alcohol bottles," and "The hand washing stations are inconveniently located, so they need to be promoted by posters"; methods concerning education, such as "Education from a young age"; and methods of promoting the issue, such as "Posters encouraging people to wash their hands" and "large signs." These results are discussed in the following section.

3.4 Discussion of the survey

The result of the survey was that the rate of people cleaning their hands with alcohol-based hand rub or soap and water was low at 14.1%. Washing hands under water alone is not expected to eliminate bacteria, and so such respondents were included in the group that did not perform hand cleaning. Although 75% of respondents expected hand hygiene to be effective in preventing infection, the results showed that the hand hygiene implementation rate was low.

Analysis the reasons for not washing hands, the reason has four factors such as “Motivation” ”Awareness” ”Environmental facilities” and “Method of hand hygiene” was revealed. As the reason that includes the factor “Environmental facilities”, the hand washing station in the cafeteria of Naruto University of Education is located in a position that makes it difficult to see from the entrance. Some respondents may not be able to engage in hand cleaning because they do not know the location of the station. Disinfection by an alcohol-based hand rub is also an effective measure against infection in addition to washing with soap and water. In the case there are few hand washing stations or they are difficult to find, alcohol-based hand rub does not require any special equipment and so it should be provided and its use promoted.

The results of the questions on the need for support for hand hygiene included proposals such as "Play a video showing how much hand cleaning decreases germs," "Posters encouraging people to wash their hands," and "Large signs."

For those result, we considered need assistance that can experience an effect of hand hygiene visually to resolve two factors such as “motivation” and “awareness”. Therefore, we developed a digital signage to promote hand hygiene using AIDA model of the advertising theory.

4. Development of a digital signage to promote hand hygiene

4.1 Outline of the digital signage

The developed digital signage shows message such as promote hand hygiene on a display, draw the awareness, and promote to use an alcohol-based hand rub that is installed. In addition, the system has the function of the user interaction to motivate hand hygiene, the display switches screen when use the alcohol-based hand rub. As shown in Figure 4, such as message to promote hand hygiene in
screen is changed when use the alcohol-based hand rub and tell user disinfection complete.

4.2 Development of the system

The system is constituted by a pressure sensor (balance wii board), a display, a PC, and a control/display application. Figure 5 shows a block diagram of the system.

The pressure sensor detects to use the alcohol-based hand rub. The application shows contents to promote hand hygiene to the display.

The control/display application was developed in the C# language and Microsoft .net Framework 4.5. Communication with the application and balance wii board was used Managed Library for Nintendo’s Wiimote v1.7 (WiimoteLib.dll) made by peekb, Esq.
4.3 Development of the contents

We had developed contents for the system such as "Shows epidemic situation of infectious diseases", and "Fortune telling" that has random nature (Figure 6).

Results of evaluated of them, we was able to confirm the improve effect of hand hygiene implementation rate.

4.4 Interface design

As regards the application’s screen transition, the interface function was designed and developed based on the advertising theory to encourage disinfect. The long-established AIDA model (AIDA: Attention, Interest, Desire, and Action) theorizes the process from users’ contact with information (e.g., advertisement), processing of information, to action. AIDA model derived hypothesis that user's behavior is changed to order of "Attention", "Interest", "Desire", “Action”. In this study, we designed interface using AIDA model that was shown in Figure 7.

The system has set the use alcohol-based hand rub as the "Action". Figure 7 represents a series of operations such as “Call for attention by install the system”, “Generate some interest by show information of infectious diseases prevention”, “Draw a desire of hand hygiene”, and “Do an action of push the bottle”.

5. Development of the MR contents

5.1 Outline of the MR contents

In this study, we developed a MR contents using projection mapping to promote effects of hand hygiene more experiential. MR is technology that fusing the virtual world and the real world, the virtual world can be experienced in the real world [13][14][15]. There is a case that has been subjected to the entire building projection mapping, the promotion effect is significant in some environments. Figure 8 shows appearance of the contents.
The MR contents inform the virus that exist around us by throw the image of virus on around the alcohol-based hand rub using projector, and draw awareness of hand hygiene. Furthermore, the contents give experience of the effect of hand hygiene by disappear the image of virus with use an alcohol-based hand rub. Figure 9 shows example of the use of the MR system.

The installed display shows such as way of correct hand hygiene and information of infections. We describe development of MR system for MR contents next section.

5.2 Development of the MR system

We develop a MR system for MR contents that has been developed. The MR system is obtained by adding projector to the system that has been developed in section 4. Figure 10 shows a block diagram of the MR system.

The Projector project movie contents for projection mapping. The MR system is realized by switching movie files of the projector “For the standby state (Figure 11)” and “For the push the bottle state (Figure 12)” when the pressure sensor detects to use alcohol-based hand rub. The movie were created using Adobe After Effect CS4.

The display shows information of about hand hygiene such as “If there is visible soil in your hand, you need a hand washing with soap and water”. In addition, in response to the opinion of confusing the location of the hand washing station, the display shows the location of the hand washing station such as "Hand washing station is on the left facing the window".

![Figure 9 Example of the use of the MR system](image)

![Figure 10 A block diagram of the MR system](image)
5.3 Installation of the MR system

The developed system is installed as shown in Figure 13. As shown in Figure 13, the display, the pressure sensor, and the alcohol-based hand rub was installed in long desk. The projector was installed turn to the alcohol-based hand rub. The projector is preferably fixed to the wall or ceiling, however in this study, the projector fixed to a tripod using an attachment of the projector fixing that we created in wood. Adjustment of the angle of the projector was performed using a camera platform of the tripod. The evaluation of the MR system that has been developed is described in the next section.

6. Evaluation

6.1 Evaluation method

The constructed MR space was placed in the entrance to the Naruto University of Education cafeteria and the hand hygiene implementation rate was surveyed. The survey was performed during the busiest period of the cafeteria from 12:00 to 13:00 on the day before set the system and again after set. The survey was performed each 2 days. The implementation rate is rate of go into cafeteria people and perform hand hygiene in the time.

6.2 Result of the evaluation

Table 1 shows result of before set the system.

<table>
<thead>
<tr>
<th>Result of before set the system</th>
<th>Day of the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First day</td>
</tr>
<tr>
<td>Cafeteria number of users</td>
<td>189 people</td>
</tr>
<tr>
<td>With only water</td>
<td>9 people</td>
</tr>
<tr>
<td>With water and soap</td>
<td>6 people</td>
</tr>
<tr>
<td>With an Alcohol-based hand rub</td>
<td>2 people</td>
</tr>
<tr>
<td>With both of them</td>
<td>0 people</td>
</tr>
<tr>
<td>Not hand wash</td>
<td>172 people</td>
</tr>
<tr>
<td>Hand hygiene implementation rate</td>
<td></td>
</tr>
<tr>
<td>An alcohol-based hand rub use rate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.2%</td>
</tr>
<tr>
<td></td>
<td>1.1%</td>
</tr>
</tbody>
</table>

First day of before set the system, hand hygiene implementation rate is 4.2%. An Alcohol-based hand rub use rate is 1.1% in it.

Second day of before set the system, hand hygiene implementation rate is 10.3%. An Alcohol-based hand rub use rate is 2.5% in it.

Table 2 shows result of after set the system.

Figure 11 The frame of movie for the standby state

Figure 12 The frame of movie for the push the bottle state
First day of after set the system, hand hygiene implementation rate is 20.6%. An Alcohol - based hand rub use rate is 12.2% in it.

Second day of after set the system, hand hygiene implementation rate is 10.9%. An Alcohol - based hand rub use rate is 6.2% in it.

### 6.3 Discussion of the evaluation results

Table 3 shows result of compare before set the system and after.

As a result of installing the MR system, hand hygiene implementation rate is improved by 8.4% on average. An alcohol - based hand rub use rate is improved by 7.4% on average, we can confirm the effect. Cafeteria crowded second day of before set the system, hand hygiene implementation rate was not improved. However, an alcohol - based hand rub use rate was improved.

### 7. Consideration

We research reason of not perform hand hygiene in opinion survey of hand hygiene. As a result, the reason has four factors such as “Motivation” “Awareness” “Environmental facilities” and “Method of hand hygiene” was revealed. We developed a system and contents to resolve two factors such as “Motivation” and “Awareness”.

---

**Table 2 Result of after set the system**

<table>
<thead>
<tr>
<th>Result of after set the system</th>
<th>Day of the experiment</th>
<th>First day</th>
<th>Second day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cafeteria number of users</td>
<td></td>
<td>131 people</td>
<td>322 people</td>
</tr>
<tr>
<td>With only water</td>
<td></td>
<td>9 people</td>
<td>8 people</td>
</tr>
<tr>
<td>With water and soap</td>
<td></td>
<td>6 people</td>
<td>15 people</td>
</tr>
<tr>
<td>With an Alcohol - based hand rub</td>
<td></td>
<td>16 people</td>
<td>20 people</td>
</tr>
<tr>
<td>With both of them</td>
<td></td>
<td>5 people</td>
<td>0 people</td>
</tr>
<tr>
<td>Not hand wash</td>
<td></td>
<td>95 people</td>
<td>279 people</td>
</tr>
<tr>
<td>Hand hygiene implementation rate</td>
<td></td>
<td>20.6%</td>
<td>10.9%</td>
</tr>
<tr>
<td>An alcohol - based hand rub use rate</td>
<td></td>
<td>12.2%</td>
<td>6.2%</td>
</tr>
</tbody>
</table>

**Table 3 Compare before set the system and after**

<table>
<thead>
<tr>
<th>Compare before set the system and after</th>
<th>On the average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result of before set the system</td>
<td>Result of after set the system</td>
</tr>
<tr>
<td>Hand hygiene implementation rate</td>
<td>7.3%</td>
</tr>
<tr>
<td>An alcohol - based hand rub use rate</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

As the factor of improve the usage rate of the alcohol-based hand rub, the light that imaged virus was moved around the bottle. The alcohol-based hand rub has feature that is small and can be installed with or without sink. Thus, the difficult case of enter the field of vision is considered. Accordingly, Such use to stand out the bottle is assumed. However, in spite of showing the location of the hand washing station to display to resolution factor of “Environmental facilities”, implementation rate of hand hygiene with water and soap did not become high. As the cause, the display only shows text that guide the location of the hand washing station. There is a need to improve the display method. We can’t realization from the limit of the equipment, there is way of display a guide on the floor.

---

**Figure 13 Layout drawing of the system**
Result of the evaluation of after install the system, the implementation rate is 15.7% on average. The rate is improved by 8.4% average.

However, hand hygiene is required to be performing with such as each meal. Therefore, there is a need to habit hand hygiene. The system has interactive function that when uses an alcohol-based hand rub, virus mark is change to star mark to perform hand hygiene with an awareness of prevention of infections. However, we couldn't verification this effect in this study, it is a challenge for the future.

In this system, we extracted the MR space around near the alcohol-based hand rub, we can more widely extract by the arrangement of the projector and the brightness of the room. By the way, this is possible to display more information for more promotion, for example, it guides the hand washing station by the arrows.

Result of the evaluation of those systems, the implementation rate did not big improve like everyone performs hand hygiene. We thought this result was influenced by people who said about awareness of hand hygiene opinion in the opinion survey. Opinions of the awareness were lots of type. Furthermore, there is an opinion that downplays hand hygiene. To ask to perform hand hygiene to those people, those people has to get an importance of hand hygiene. Therefore, we considered that need study the contents again and continue to provide the knowledge of hand hygiene to innovate awareness of hand hygiene.

8. Conclusion

We perform the opinion survey and analysis reasons of not perform hand hygiene. As a result, the reason has four factors such as “Motivation” "Awareness" "Environmental facilities" and “Method of hand hygiene” was revealed.

We developed the digital signage to resolve two factors such as “Motivation” and “Awareness” to improve hand hygiene implementation rate. Furthermore, we developed MR contents and MR system that can experience visually disappearance of virus by hand hygiene using projection mapping

As a result of install the developed system, hand hygiene implementation rate is improved by 8.4% on average.

References


Akinori Kanazawa is currently a master's course student in the Department of Technology and Information Education, Naruto University of Education, Tokushima, Japan. He is also a member of Information Processing Society of Japan. His research interests are Information, Public Health, Knowledge Science, and Artificial Intelligence.

Toshiko Asai is currently a president of the Information Workshop Co. and Founder of Impact Hub Kyoto, Japan. She received a coursework completed without degree in Knowledge Science from Japan Advanced Institute of Science and Technology (JAIST) in 2008. Her research interests are Information, Education, and Innovation and Knowledge Science.

Akinori Minazuki is currently a professor and a director of Center for Information and Technology and System, Kushiro Public University, Hokkaido, Japan. He received his Ph.D degree in Knowledge Science from Japan Advanced Institute of Science and Technology, Japan (JAIST) in 2003. His research interests are Information and Knowledge Science, Information Technology for Social Systems, Information Education, Emergency Medical Service & Medical Information, and Public Health.

Hidehiko Hayashi is currently an associate professor in the Department of Technology and Information Education, Naruto University of Education, Tokushima, Japan. He received his M.S in Information Science and Ph.D degree in Knowledge Science from Japan Advanced Institute of Science and Technology (JAIST) in 2000 and 2003, respectively. His research interests are Information and Knowledge Science, Brain Computer Interface, Medical Information, Evaluation of Human Information Processing Mechanism, and Public Health.
Online Classroom Feedback System and Their Implementation in a Foreign Language Presentation Course in Japan

Yuichi Onoa
University of Tsukuba, Japan
Manabu Ishihara
Oyama National College of Technology, Japan
Sachio Hirokawa
Kyushu University, Japan
Mitsuo Yamashiro
Ashikaga Institute of Technology, Japan

Abstract
This paper is a revised and expanded version of a paper entitled ‘Implementation and evaluation of real time qualitative feedback systems in a foreign language presentation course’ presented at the ‘IIAI International Conference on Advanced Applied Informatics (AAI2014)’, Kitakyushu, Japan. This paper explores the possibility of implementing two types of text-based online-feedback system in a classroom setting. After reviewing two traditional approaches to real-time feedback, the Clicker Approach and the Forum Approach, we will suggest that neither is adequate as a motivational tool for learners on a presentation course. Instead, we propose two approaches based on a text-mining technique to compensate for these inadequacies. The first approach is a “Keyword and Frequency” system and the second is a “Mind-mapping” system. The first can be mainly used for individual real-time feedback on a presentation course, while the second can be best incorporated into the syllabus as a total feedback or teacher’s reflection offline. This paper describes the details of the two systems, and reports on their integrative implementation when carried out in the EFL presentation course of a Japanese university. It will be shown that our systems have a positive effect on learner motivation based on comparative research between the Clicker Approach and our approach. We suggest that our system has potential in a wider e-Learning environment where the audience are people from around the world.

Keywords: Online classroom feedback, Text-mining, Moodle, Presentation course, Foreign language teaching

1. Introduction

The importance of instant feedback has been demonstrated in the literature, which has emphasized its role in fostering meaningful interaction between students and instructional materials [1], its contribution to student development and retention [2], and student participation and engagement in classroom tasks and activities [3]. For example, the introduction of the audience response system had an impact on student participation and engagement in a large-size classroom lecture [4,5]. Needless to say, presentation skills are considered to be of paramount importance to students [6], but speaking in public produces high levels of anxiety [7] and this anxiety is a difficult part of the student experience, unsettling or even frightening some students [6]. This is especially true of less confident learners [8]. In these situations, it is necessary to encourage and support learners with the help of peer feedback so that they are not demotivated by fear and anxiety.

Our reason for introducing this feedback system is that we have felt the inadequacy of traditional approaches to learner motivation. Such traditional approaches include the following: (i) the quantitative approach (Clicker Approach) and (ii) the qualitative approach (Forum Approach). The former approach has the advantage of saving cost and time for the audience; they only have to click a button after a presentation (see Figure 1 below). The latter approach has the advantage that learners have an opportunity to receive comments on the presentation directly from the audience (see Figure 2 below).

We begin our discussion on the problems associated with these approaches when applied to foreign language teaching settings in terms of learner motivation. After clarifying the issues around traditional approaches, we describe the details and implementation of our proposed system in an English presentation course in a national university in Japan.

2. Backgrounds

2.1 The Role of Real-time Feedback
It is generally agreed that learning is more effective when interaction occurs between learners [9,10]. A one-way instruction style in a large classroom often makes the instruction boring for students. It is pointed out that four challenges of teacher-centered lectures: (1) large class size; (2) working memory issues (difficulty in memorizing everything said in the lecture); (3) concentration (student concentration does not last long); and (4) grasping student understanding (teachers have no way to know how well each student has understood the lecture) [11]. Strong empirical evidence suggests that active involvement in the learning process is important in two areas: for the mastery of critical thinking and problem-solving skills, and to increase the likelihood of a student completing a program [12,13,14]. The role of real-time feedback has been noted in the literature in terms of participants’ interactive engagement [4,5], and checking participants’ understanding [11]. In the case of foreign language teaching, making presentations provides students with an opportunity to reconsider their approaches to critical thinking, problem solving, collaborative learning, speaking, and writing. Instant feedback activities that engage the audience are one method of encouraging active learning for both presenters and their audience. By providing feedback immediately after a student completes his or her presentation, the experience is more authentic, and occurs precisely when he or she is the most receptive to criticism, coupled with the excitement of getting positive or sometimes negative reaction from the audience [3]. Improving the quality of instant feedback activities can motivate students to become more involved in the learning process. Various kinds of tools have been introduced to guarantee the “interactivity” and “quality” of the feedback in a large classroom. In the following subsection, we review two such major tools for real-time feedback.

2.2 Quantitative Approach: Use of Multiple-choice Questions

A commonly used device for this approach is the Clicker, which is more formally known as an “Audience-Response Analyzer” (ARA). This system involves multiple-choice questions created by the instructor before class. Even in non-wired classrooms, the analyzer makes it possible to receive responses from the audience by setting up the teacher’s PC, a response receiver, and multiple response senders. A picture of the Clicker, a sender and receiver, is shown in Figure 1, and that of its output of the system is shown in Figure 2.

The effect and usefulness of Clicker use in the classroom has been reported in the literature from various contexts [3,4,5,11,15,16]. Commonly agreed effects of the use of Clicker are enhanced engagement and participation, improved exam scores or passing rates, student comprehension, and gauging the level of understanding. Although use of Clicker has spread worldwide, including into Japanese classrooms, this approach tends to be based on a quantitative output and its descriptive statistics. The question items for instant feedback tend to be multiple-choice, voting questions among, say, five candidates, or consciousness surveys on a five-point Likert scale. In the case of feedback on a presentation course, Clicker questions tend to be 5-point Likert questions, such as “Pronunciation is good,” or “Visuals are helpful.” The more thoroughly we need to know the results based on assessment rubrics, the more detailed the question items we need to ask the audience. Students have to answer all of these questions in a short time, say, three minutes.

If we ask many analytic peer assessment questions in a presentation course, however, it raises some important issues regarding the validity of the assessment results. The issue that is raised is from the perspective of the raters. Our
earlier study [3] surveyed such peer assessment of student presentations in a Japanese EFL course on the basis of twenty items regarding “content,” “visuals,” “manners,” “source reliability,” and so on. Two groups were formed: a low-proficiency group and a high-proficiency group. It emerged that the high-proficiency group was able to evaluate these twenty items and produced four assessment factors after our factor analysis. On the other hand, the low-proficiency group was able to produce only three factors, which meant that their assessment rubrics were not clear: Specifically, two assessment standards, those of organization and English, were integrated for the low-proficiency group. The conclusion was that learners with lower proficiency had a tendency to give holistic or impressive, rather than analytic, assessments, implying that the task of peer evaluation on a quantitative basis involves a certain amount of invalidity. It might also be possible that the task becomes boring for students when they are asked to observe many presentations and answer the same question items repeatedly in a short time, resulting in less reliability of collected results owing to fatigue or boredom.

2.3 Comment List Approach: Use of Bulletin Board

In the previous section, we pointed out the issue around the invalidity of peer ratings and poor motivation using the Clicker approach. Since this approach is rooted in quantitative analysis, it is natural that “qualitative” feedback is missing, causing fatigue or boredom in students. This gap can be filled with an open-ended questionnaire. Instead of showing the statistical ratings of each question item, the presenter can recognize what the audience felt about their presentation directly from their feedback comments. In fact, this is an easy method, because the common Social Networking Sites (SNS) or bulletin boards installed in Learning Management Systems (LMS) are now available to instructors in networked classroom environments. Since the concept of “Social Interaction” and the social learning approach has been widely spread, the use of SNS (Twitter, Facebook, Line, etc.) has become much easier for everybody, including students and instructors. When we consider the effective feedback activity in a presentation course, it is reasonable to use this technology as an instant feedback tool (Hasegawa, Yasui, & Yamaguchi, 2013).

This approach has some other merits for learners. Students try harder to understand the presentation, because they have to give free comments that appear on the projector shortly after they write them. In other words, students’ active engagement seems to be brought about by this task. In addition, low-proficiency students prefer free comments to scoring, probably because they can choose what point to address; they do not need to be analytic and check everything during the presentation for the peer assessment session. In particular, low-proficiency students prefer holistic assessment to analytic assessment. As noted above, the comment writing approach seems to be accepted more readily.

Thus far, it seems that this approach is fine as long as it is not stressful for raters, and presenters can immediately read audience feedback comments. However, if this is a large class, say, more than 100 students, conducted in a large classroom, the presenter may have to look at 100 feedback comments one by one. It might be difficult to understand the general trend of opinion about the presentation from a list of raw feedback comments. Uncontrolled feedback comments might automatically appear on the screen. The instructor must give sufficient instruction regarding what kinds of comment should (not) be posted for feedback activity. The use of LMS can remove this risk, but it is not certain that the system can draw only acceptable text data from the audience. The point here is that we need to provide a certain level of analysis so that all comments can encourage the presenter to prepare for the next presentation.

2.4 Summary of Issues in Traditional Approaches

The aim of instant feedback in a foreign language presentation course is to encourage the presenter to become aware of the good/bad points of their achievements through the real-time reaction of the audience. In this sense, an issue arises when using the Clicker approach from the viewpoint of motivating presenters from feedback. The traditional quantitative approach might be helpful in understanding the trend of the audience based on certain standards, like pronunciation, organization, and visual aids. This approach, however, does not provide information about what the presenter really wants to know, unless s/he questions the audience. Moreover, this simple approach, which appears to be easy and not time-consuming, in fact takes time if the number of questions is large enough to be analytic. The questions in a traditional approach tend to be 5-point Likert questions, such as “Pronunciation is good,” or “Visuals are
helpful.” The more thoroughly we need to understand the results based on assessment rubrics, the more question items we need to ask the audience. For their part, the audience has to answer each of, say, 20 questions within, for example, three minutes. In this sense, the Clicker approach is not practical for use with a large number of questions when we consider the need to pick up encouraging or motivating factors from the given statistical description.

On the other hand, the list of comments from the audience using the Bulletin Board approach provides qualitative feedback, compensating for the inadequacies of the Clicker Approach. However, the issue of how to control the board remains. Moreover, it is highly possible that the presenter will not perceive any trends, other than there are many comments about their presentation.

Our proposal is to compensate for the latter approach, introducing the concepts of “Summary of free comments” by using keywords and frequency indices. In the following section, we would like to introduce our system design.

2.4 Proposal: Keywords and Frequency Approach

The basic concepts of our proposal are given below [18,19]:

- Real-time analysis of collected raw text data
- Production of tendency graphs according to keywords or categories previously prepared by the instructor
- Use of the natural language processing (text-mining) method to pick up keywords
- Reference to the original feedback comments by presenters

The first trial is to produce graphs of results based on frequency of the given keywords. The schematic outline of the proposed system is shown in Figure 4 below.

In a typical text-mining approach, the essential procedure is how to process the real raw data; namely, the process of natural language processing of Japanese in our research. In the case of the Japanese language, the process of text mining is usually composed of the following four stages:

- Analysis of morphology
- Analysis of structure
- Analysis of meaning
- Analysis of discourse

The first stage is unique to the Japanese language, because unlike the English language, there is no space between words in Japanese. The analyzer first has to divide the whole text based on Japanese morphological rules. In our selection of free software, we employed MeCab for its high versatility. In the next stage, the analyzer has to recognize the dependency relations between words. CaboCha is a high-level free Japanese syntactic parser, which enables us to analyze syntactic dependency of Japanese sentences. CaboCha was employed in our system to analyze the dependency relations of a sentence. The process is driven by proc_open() function and the package is installed in the activity module of our LMS - Moodle. This is shown in Figure 5.
Usually text mining is conducted off-line; a certain amount of time is spent analyzing and deriving important clusters of meaningful concepts from text data through statistical processes. Since our system has been constructed so that the feedback is provided soon after the audience input is complete, we concentrated on the use of descriptive aspects of statistics as an output of the system. The system is implemented on Moodle as one activity session. This makes the students work on this activity even when outside the classroom. The students can refer back to their text-mining results since they are stored on a database. In addition, they can refer to original text data that the audience input to the database to make their points clearer.

In order to process text data for our purposes, we created a dictionary for evaluation. To select lexical items for this dictionary, we used actual text data, collected during the pilot study. Fifty students participated and made open-ended comments on each of ten presentations. After analyzing these data using CaboCha, around 410 lexical items were selected, and two kinds of property were ascribed to each item. These properties were: (i) properties of impression, and (ii) properties of semantic category, such as “design/layout,” “interest,” “English,” “pronunciation,” “citation,” and “others.” Our evaluation dictionary file is shown in Figure 6. The audience simply had to write down
comments. The presenter was given the feedback output soon after the presentation was finished, as shown in Figure 7 where the results are given in a graph and radar chart. The results of Figure 7 include the ratio of positive/negative word appearances and frequency of each semantic category. In addition to this, the student (presenter) can make refer to the original text data by clicking the key words. This is shown in Figure 7.

2.5 Mind-mapping Approach

The proposed keyword and frequency approach has the advantage of providing information regarding what kinds of keyword the audience picked up about the presentation, not just as an average score of a question item. If the student was enthusiastic in creating layouts of the slides and the audience made comments on this point, then this is very encouraging feedback. However, this system does not consider the relatedness among keywords or frequency of co-occurrence between words; our system simply shows the frequency of the keywords themselves. The collected text data should be shown so that the concepts or ideas of the audience can be visualized schematically as a mind map. This might help instructors to gain insight into the audience’s conceptual models from plain text comments.

According to [20], drawing a mind map can be thought of as a process of searching one-by-one for related keywords, starting in the center with the keyword or image that is central to that concept. The related keywords or images are reiterated and expand in a radial pattern, linking back to the central concept through contextual relations. Our mind map of a word was constructed through the depth-first-search procedure. Given a word w, firstly we obtain the set of documents that contain the word w. Then, we extract the characteristic words of the documents, set according to the relevance degree. In the present paper, we adopted the SMART measure [21]. The top K words were selected according to the measure. Each word $u_i$ in the feature words is then used to “AND” search by “w $u_i$”
and a new node with the label “ui” is linked from the root whose label is “w.” We repeat this process until the depth reaches the fixed parameter D. Note that a word is checked once it is used in a search keyword. Thus, a word appears only once, and the graph becomes a tree. A sample map is shown in Figure 8 above.

3 Implementation in a Presentation Course

Figure 9 illustrates how this system was implemented in our class. Soon after each presentation was completed, a peer evaluation session followed. Evaluation results were shown on the projector screen. They were also updated in the course LMS so that the students could reflect afterwards. At the end of the presentation, the visualized result of the mind map of all the collected free comments was shown on the projector and the instructor made some comments on it in a summary session. In our study, two classes of a Japanese national university participated and they worked on two types of feedback system: (i) Clicker Feedback and (ii) Keywords and Frequency Feedback. The presentation topic was “The Introduction of Foreign Culture: Discovering things unknown to Japan.” In a twenty multiple-choice question feedback, all the students have to do is to click the button, as shown in Figure 10, for 20 questions. In a text-mining approach, the students write their comments in an open box on the LMS.
4 Evaluation

4.1 Comparison between Quantitative Approach and Keyword and Frequency Approach

The experimental study was conducted in an EFL presentation course in a Japanese National University. Two classes of 59 first-year students participated in the study. The proficiency level of these two classes varied greatly. Class A was a high-level class in which there were 24 students, 12 of whom had studied abroad for more than one year. Class B consisted of 35 students, only 1 of whom had been abroad. All first-year students in the university took a placement test at the beginning of the academic year and were streamed according to their placement scores. The placement test consisted of listening and reading questions, and was similar to TOEFL. The average scores of Classes A and B were 71.0 and 41.0, one of the highest and the lowest in our university, respectively. Two feedback systems were utilized: (i) Twenty 5-point Likert scale questions for analytic evaluation, and (ii) Open-ended feedback. After the course, a questionnaire was carried out to see if there was any difference between the two instant feedback activities. In conducting the survey, we established two research questions: (i) Is there any difference in usability?, and (ii) Is there any effect on students’ self-awareness or motivation? The survey was carried out after the presentation project had been completed for each topic. The survey contained ten questions concerning rater difficulty or degree of satisfaction, and how receiving feedback was motivating as a learner. We compared the results between the two instant feedback systems.

The statistical results are given in Table 1, where the result of a t-test is also shown. Three items were statistically significant or tended towards significance in each class. In our early studies, students were encouraged to think deeply about each presentation so as to express their opinions concretely. This encouraged students to enhance their analytic and critical thinking skills. When it came to rater difficulty or troublesomeness, students did not perceive a big difference between the two feedback systems. However, when considering proficiency level as a variable, the structure seemed to change: for students with a lower proficiency level, writing free comments reduced the level of difficulty or troublesomeness. This result seems to be related to the task difficulty of peer evaluation and student proficiency. As our earlier study suggested, low-level learners tend to integrate analytic standards into one abstract evaluation. In other words, it is difficult and troublesome for such learners to pay attention to every standard in the rubrics during the presentation and their assessments tend to be holistic rather than analytic, reducing the reliability of the peer evaluation. On the other hand, the high-proficiency students seemed to feel that they were more motivated in our text-based feedback, as was expected.

The survey included an open-ended question. Most of the students answered positively about our system. Some of the comments are listed in Table 2, where the sentences were translated into English from the original Japanese ones.

Table 1. Descriptive Comparison between the Two Feedback Systems

<table>
<thead>
<tr>
<th></th>
<th>Quantitative</th>
<th>Free-comments</th>
<th>t-value (two-tailed)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>High Proficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was difficult to give evaluations to others.</td>
<td>3.167</td>
<td>1.090</td>
<td>2.958</td>
<td>1.197</td>
</tr>
<tr>
<td>It took a long time to give evaluations.</td>
<td>3.542</td>
<td>1.141</td>
<td>2.958</td>
<td>1.197</td>
</tr>
<tr>
<td>It was troublesome to give evaluations.</td>
<td>2.583</td>
<td>1.060</td>
<td>2.667</td>
<td>0.816</td>
</tr>
<tr>
<td>Giving evaluations was beneficial for my study.</td>
<td>3.208</td>
<td>1.103</td>
<td>3.750</td>
<td>1.032</td>
</tr>
<tr>
<td>I was satisfied with my evaluation to others.</td>
<td>3.583</td>
<td>0.974</td>
<td>4.000</td>
<td>0.933</td>
</tr>
<tr>
<td>Their evaluations were helpful.</td>
<td>3.792</td>
<td>1.103</td>
<td>3.792</td>
<td>1.103</td>
</tr>
<tr>
<td>I realized how the audience feel about my presentation.</td>
<td>3.917</td>
<td>1.283</td>
<td>3.958</td>
<td>1.160</td>
</tr>
<tr>
<td>The content of their feedback was convincing.</td>
<td>3.792</td>
<td>1.021</td>
<td>3.833</td>
<td>1.129</td>
</tr>
<tr>
<td>Their feedback gave me hints for the next opportunity to present.</td>
<td>3.792</td>
<td>0.932</td>
<td>4.167</td>
<td>1.129</td>
</tr>
<tr>
<td>I became confident as a result of their feedback.</td>
<td>3.667</td>
<td>1.049</td>
<td>3.542</td>
<td>0.977</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean SD</th>
<th>Mean SD</th>
<th>Mean SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Proficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was difficult to give evaluations to others.</td>
<td>3.806</td>
<td>0.980</td>
</tr>
<tr>
<td>It took a long time to give evaluations.</td>
<td>4.129</td>
<td>1.088</td>
</tr>
<tr>
<td>It was troublesome to give evaluations.</td>
<td>3.613</td>
<td>1.283</td>
</tr>
<tr>
<td>Giving evaluations was beneficial for my study.</td>
<td>3.484</td>
<td>0.890</td>
</tr>
<tr>
<td>I was satisfied with my evaluation to others.</td>
<td>3.581</td>
<td>0.848</td>
</tr>
<tr>
<td>Their evaluations were helpful.</td>
<td>3.645</td>
<td>0.915</td>
</tr>
<tr>
<td>I realized how the audience feel about my presentation.</td>
<td>3.935</td>
<td>0.854</td>
</tr>
<tr>
<td>The content of their feedback was convincing.</td>
<td>3.968</td>
<td>0.752</td>
</tr>
<tr>
<td>Their feedback gave me hints for the next opportunity to present.</td>
<td>3.806</td>
<td>0.873</td>
</tr>
<tr>
<td>I became confident as a result of their feedback.</td>
<td>3.484</td>
<td>1.092</td>
</tr>
</tbody>
</table>

***: p<.001, **: p<.01, *: p<.05, +: p<0.1
The text data collected in this survey were divided into the following four points: (i) much simpler interface in comparison with the traditional quantitative feedback; (ii) deeper understanding of good/bad points of the presentation; (iii) usefulness of keyword reference and graphic outputs; and (iv) made the class exciting.

Summarizing the effects of keywords and frequency approach, the system reduces the difficulty or trouble of learners with lower proficiency in tackling on analytic assessment that must be done in a short time. It also has an effect on high-level learners in terms of deeper reflection on the contents of their presentation. In any way, it can be said that the new system serves as a tool compensating for the insufficiency of Quantitative approach to real time feedback.

4.1 Use of Mind Maps as a Summary of the Class

The sample map is shown in Figure 8 above. In this situation, a presenter is interested in how his “slide,” which is located at the center of the mind map, appeared in the feedback from the audience. The mind map output shows that the word “slide” is related to three words: “letters,” “easy to see,” and “speak.” The last keyword “speak” might be a new finding for learners, since “slide” has a relationship with “speaking.” This might be a striking suggestion because students might not be aware that the speed of speech helps the audience focus on their slide. Conversely, the new finding might include that speaking too fast does not allow the audience to see their slides comfortably. This type of analysis would not be brought about from the simple list of free comments unless several participants directly made comments in this way.

A pilot survey was conducted to see how participants felt the system. Since the system was completed one month after the course, the author asked 10 volunteer participants to try this system and they gave free comments on the system. Most of their response was very positive. Some answered, “This is what I was thinking about” and “I was made to realize important factors to improve my presentation by this mind map”.

On the part of the instructors, the result of mind map gives us the opportunities. One of the instructor I interviewed answered the following:

This mind map is interesting and useful as a teaching portfolio. By comparing maps, I can observe how the change of students’ comments is visualized. I can also observe the topic effect if we compare maps among different topics in the same class.

How to incorporate this tool into the course needs to be studied further in the future, of course. However, it can be said that our mind map approach provides implications for deepening the quality of real time feedback tasks in a presentation course.

5 Concluding Remarks

This paper proposed a new online qualitative feedback system based on text data collected on LMS for foreign language presentation course and validated the motivational effects on learners with various proficiency levels. This paper further introduced a mind map approach that can be used online or offline as a visualization of relatedness among clusters. This system is not, of course, unique to presentation course, since the need for instant feedback is more universal in other areas as well. For teachers and instructors, the system might be incorporated in the class for improving the course contents. Students might benefit in other courses than foreign language teaching. Given a network environment, the students can get access to the system in distant classroom settings. Anyway, further research is required to apply this system to various kinds of educational settings, including the improvement of user interface, creation of more sophisticated evaluation dictionary for each setting and proper incorporation into instructional design.

Acknowledgement

This study was partially supported by Grant-in-Aid for Scientific Research No. 26370656, No. 24500176, and No. 25350369 from the Japan Society for the Promotion of Science (JSPS).

References


Yuichi Ono is an Assistant Professor at the Faculty of Modern Language and Cultures of Tsukuba University. He holds a Master Degree in Education from Tsukuba University. After working on e-learning staff at Oyama National College of Technology, he was employed in Foreign Language Center in 2011. He was appointed as a chief member of CALL management committee of Foreign Language Center. His recent research interests focus on design, evaluation and implementation of CALL and mobile-integrated blended learning in the settings of foreign language education.He is a member of the IEEE, the IEE of Japan, and LET.

Manabu Ishihara is a Professor in the Department of Electrical and Computer Engineering, Director of Education and Research Center of Information Science, Oyama National College of Technology. He received his PhD degrees from Meisei University, Tokyo, Japan, in 1987. His research interests include
multimedia, computer network, human engineering and virtual reality technology. He is a member of the IEEE, the IEICE, the IEE of Japan and ASJ.

Sachio Hirokawa received a BS and MS degree in Mathematics and PhD degree in Interdisciplinary Graduate School of Engineering Sciences from Kyushu University in 1977, 1979, and 1992. Since 1997, he has been a professor in the Research Institute for Information Technology of Kyushu University. His research interest includes search engine, text mining, and computational logic.

Mistuo Yamashiro is a professor in the department of systems and information engineering of Ashikaga Institute of Technology. He received PhD degree in engineering. His major is manufacturing systems and powder technology.
Information for Authors

Scope

The aim of the International Journal of Computer & Information Science is to present information on well defined theoretical results and empirical studies, to describe current and prospective practices and techniques, to provide critical analysis of the use of CASE tools and to present broad issues in the fields of Computer Science and Engineering, Software Engineering and Information Science and Engineering. A primary aim of the journal will be to make the research accessible and useful to practitioners through carefully written methodology section and discussions of research results in ways that relates to practitioners.

Submission Policy

1. The IJCIS does not accept manuscripts or works that have been published in a journal, but accepts substantially revised and expanded versions of papers that were previously presented at conferences. In this case, the author should inform the Editor-in-Chief of the previous publication when the paper is submitted.

2. Submissions, referencing and all the correspondence will be conducted by e-mail.

3. Manuscripts should be in English and are subject to peer review before selection.

Procedure for Submission for Review

1. Send your complete text of manuscript to the Editor-in-Chief (e-mail: lee@cps.cmich.edu) via e-mail as an attachment document either in Word, PDF or PostScript format.

2. It may be useful for non-English speakers to have a native English speaker read through the article before submission.

3. Your e-mail message (to the Editor-in-Chief) should include the title of your paper, author(s) information (department/section, university/company, phone/fax numbers, e-mail address), the corresponding author’s complete address (mailing address), and area(s) of the paper.

Style of Manuscript to be Submitted for Review

1. The text should be double-spaced, two columned, single-sided in Times 12-point type on 8.5-inch x 11-inch paper with margins of 1 inch on all four sides. It should not exceed 30 double-spaced pages. All pages must be numbered sequentially.

2. Provide a 200-250 word abstract, 5-7 keywords. Your abstract should briefly summarize the essential contents.

3. All figures/tables must be captioned, numbered and referenced in the text.

4. An acknowledgement section may be presented after the conclusion section.

5. References should appear at the end of the paper with items referred to by numbers in square brackets.

Journal Papers: author’s first initials, followed by last name, title in quotation marks, journal (in Italics), volume and issue number, inclusive pages, month and year. For e.g.:


Papers of Proceedings: author’s first initials, followed by last name, title in quotation marks, proceedings / symposium / workshop name (in Italics), location, month, and year, and inclusive page numbers. For e.g.:


Books: author’s first initials, followed by last name, title (in Italics), edition, location: publisher, year. For e.g.:


Accepted Manuscript Submission

All accepted papers are edited electronically. You are required to supply an electronic version of the accepted paper. The detailed guidelines for preparing your final manuscript will be sent upon article acceptance. It is also available on the ACIS website at http://acisinternational.org.

Page Charges

After a manuscript has been accepted for publication, the author will be invoiced a page charge of $400.00 USD for up to 10 pages if he/she is not a member of ACIS and thereafter an additional $30 per printed page to cover part of publication after the receipt of your final manuscript (camera ready).