

Opinion corpus for the assessment of a study-abroad program

Katsunori Kotani · Takehiko Yoshimi · Mayumi Uchida
Kansai Gaidai University · Ryukoku University · Kansai Gaidai University / Japan

Abstract – This study compiled an opinion corpus for developing a method for automatically evaluating a study-abroad program. Evaluation should cover not only academic experience at a host institution but also intercultural experience in the dormitory and interpersonal experience with local students, which helps improve a study-abroad program. The corpus included 600 students' opinions on the satisfaction with academic, intercultural and interpersonal experiences, consisting of 40,024 words in total. Each opinion was annotated according to the opinion polarity determined by an existing sentiment classifier automatically. When automatically classified opinion polarity was compared with manually determined opinion polarity, a different distribution was observed. Because the existing classifier was not trained with a corpus that dealt with the issues related to students' opinions about a study-abroad program, this result suggested the need of a corpus for study-abroad program evaluation. The opinion classifier of this study trained with the opinion corpus demonstrated a higher accuracy (83.5 percent) than the majority class baseline (70.9 percent).

Keywords – opinion corpus, study-abroad program assessment, classification, opinion mining, degree of positiveness/happiness

1. INTRODUCTION

Making an effective study-abroad program needs evaluation regarding not only students' academic experience at a host institution but also their intercultural experience in the dormitory and interpersonal experience with local students. Previous research on the academic experience (El-Halees 2011; Leong et al. 2012; Kaewyong et al. 2015) developed evaluation methods using an opinion mining technique that analyzes students' opinions regarding the course contents, which would contribute to the improvement of the courses. On the other hand, the evaluation of intercultural and interpersonal experiences has not thoroughly been investigated, as Savicki and Brewer (2015) suggested that study-abroad program assessment was still at the early/challenging stage. Engle and Engle (2004) evaluated study-abroad programs regarding academic and intercultural experiences analyzing students' answers to multiple choice questions.

This study aims to develop a method for evaluating a study-abroad program, which predicts students' satisfaction with study-abroad experiences consisting of academic, intercultural and interpersonal experiences. In developing an evaluation method with machine learning algorithms, its accuracy depends on the quality and quantity of training data, that is, an opinion corpus of students who completed a study-abroad program. Subsequently, this study compiled an opinion corpus for study-abroad program assessment, where a data instance comprises a student's opinion for his/her satisfaction with study-abroad experiences, linguistic properties of the opinion and the opinion polarity (negative vs. positive). This study assigned the opinion polarities automatically employing a machine-learning sentiment classifier, and manually by a faculty member of an institution offering the tested study-abroad program. Since the validity of automatic classification is dubious due to the classification accuracy of a classifier trained with opinions



that were not related to a study-abroad program, this study developed and assessed a classifier based on the linguistic properties of the opinion corpus on a study-abroad program.

2. METHOD FOR CORPUS COMPILATION

The opinion corpus was compiled through course evaluations for a study-abroad program at Kansai Gaidai University in Japan. The study-abroad program offered courses on Japanese language and Asian cultures as well as business and politics in Asian countries. These courses were delivered in English, while workshops, such as Japanese pottery, were offered in Japanese. The program duration was half a year (fall semester from September to December; spring semester from January to May).

In the course evaluations, students responded to open-ended questions about their satisfaction with academic, intercultural and interpersonal experience as well as three further questions which students replied regarding their general satisfaction with the above three categories on a five-point Likert scale (1: excellent, 2: good, 3: average, 4: below average, 5: poor).

The opinion corpus was compiled from 600 university students from Europe, North America and Asian countries through course evaluations in Spring ($n=302$) and Fall ($n=298$) semesters. Among the students, 36 percent completed one year abroad by taking the program twice. More than half of the other students applied to participate in a second semester. Students were free to answer at any given time.

3. ANALYSIS OF CORPUS DATA

3.1. Opinion in answers to open-ended questions

This study analyzes opinions for intercultural experience on the basis of opinion's positiveness. The opinion polarity was evaluated manually and automatically. In the manual evaluation, the opinions were classified into three classes (negative, neutral and positive) by a faculty member at the institution of the study-abroad program. In the automatic evaluation, the opinions were classified into 100 classes of positiveness (from class 1 – absence of positiveness – to class 100 – presence of positiveness) by a classifier trained with a corpus that dealt with the general issues (uClassify 2015).

Figure 1 shows the relative frequency of manual evaluation results for the opinion polarity. The distribution did not follow the normal distribution according to the Kolmogorov-Smirnov test ($K=5.47$, $p<0.01$). Figure 2 shows the relative frequency of automatic evaluation results for the opinion polarity, where the negative and positive groups comprised the lower and higher 30 percent of the scores (negative: $0 \leq n < 30$, positive: $70 < n \leq 100$), and the neutral group comprised between the negative and positive groups. The distribution followed the normal distribution according to the Kolmogorov-Smirnov test ($K=0.73$, $p=0.33$). The automatic evaluation results demonstrated a different distribution from the manual evaluation results, which suggested that an opinion classifier should be trained with students' opinion corpus data.

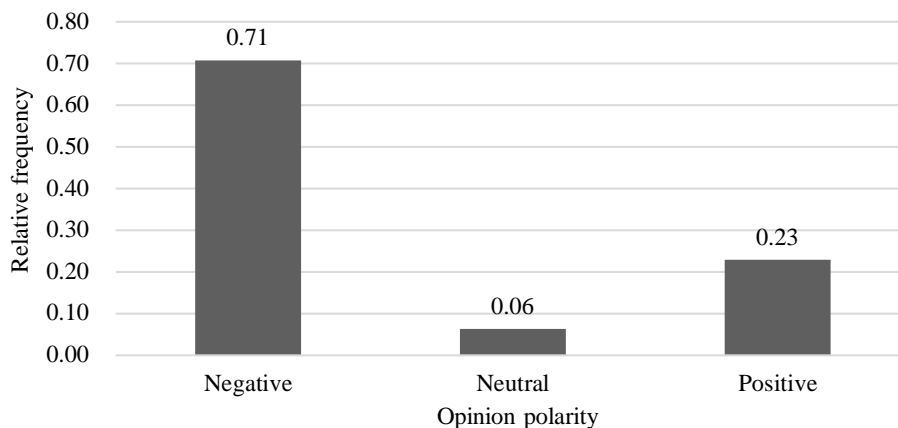


Figure 1: Relative frequency of manual evaluation results

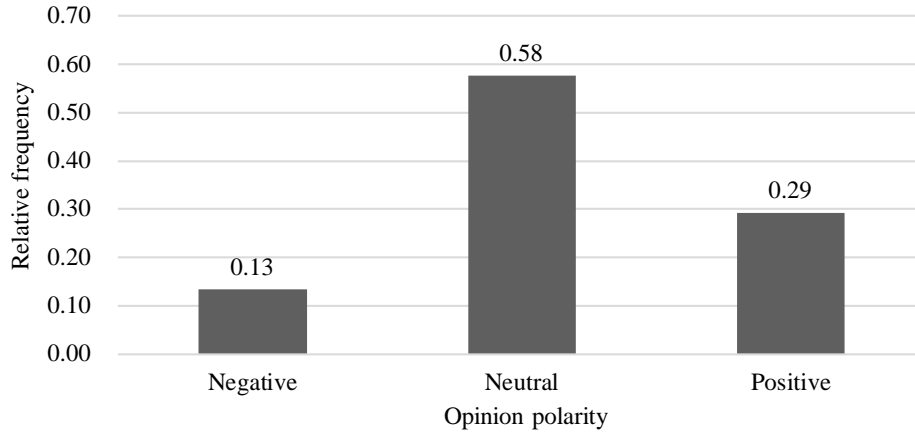


Figure 2: Relative frequency of automatic evaluation results

Table 1 shows the linguistic properties of the opinions, i.e., the answers to the open-ended questions. The mean opinion length was the highest in the opinion on academic experience. This was considered to suggest the degree of interest (positive opinions) or concern (negative opinions).

Linguistic property \ Type	Academic experience	Intercultural experience	Interpersonal experience
Number of opinions	277	351	314
Number of word-tokens	13,726	13,767	12,531
Mean opinion length (words)	49.6	39.2	39.9

Table 1: Linguistic property of students' opinion

Table 2 shows the top five tri-grams and their frequencies. The results showed frequent use of a first-person subject and that of the modal verb *would* in the academic experience opinions. A modal verb *would* expressed the presence of students' counter expectation, which should be taken as problems of the study-abroad program. In addition, the use of a first-person subject indicated students' negative opinions, because a first-person subject was infrequently used for expressing positive appreciation for their intercultural and interpersonal experiences without using a first-person subject.

Ranking \ Type	Academic experience	Intercultural experience	Interpersonal experience
1	<i>I would like</i> (34)	<i>the seminar house</i> (34)	<i>to interact with</i> (44)
2	<i>like to see</i> (27)	<i>the seminar houses</i> (25)	<i>the * students</i> (39) ¹
3	<i>would like to</i> (27)	<i>in the seminar</i> (24)	<i>a lot of</i> (27)
4	<i>it would be</i> (22)	<i>my host family</i> (16)	<i>with local students</i> (26)
5	<i>to see more</i> (19)	<i>seminar house 2</i> (15)	<i>the local students</i> (23)

Table 2: Top-five word-level tri-grams

3.2. Satisfaction from multiple-choice questions

Table 3 shows the distribution of the satisfaction with study-abroad experiences. Although the students' satisfaction was provided on the five-point Likert scale, the satisfaction was summarized into three classes: positive, neutral and negative.

Satisfaction \ Type	Academic experience	Intercultural experience	Interpersonal experience
Positive			
Excellent	511 (37.4%)	254 (42.8%)	153 (25.5%)
Good	562 (41.1%)	212 (35.7%)	227 (37.9%)
Neutral	203 (14.8%)	82 (13.8%)	140 (23.4%)
Negative			
Below average	69 (5.0%)	26 (4.4%)	54 (9.0%)
Poor	23 (1.7%)	20 (3.4%)	25 (4.2%)

Table 3: Distribution of satisfaction with the study-abroad program

¹ The names of countries have been removed (*).

The chi-square test demonstrated no statistical significance regarding the distribution of satisfaction among the academic, intercultural and interpersonal experience ($\chi^2(8) = 2.153$, $p > 0.01$). As most students judged the study-abroad experiences as positive, the study-abroad program was considered to have earned high satisfaction among the students.

4. DEVELOPMENT AND EVALUATION OF OPINION CLASSIFIER

Given the different distributions between the manual and automatic classification, this study developed an opinion classifier using the opinion corpus. Support Vector Machine, the function 'svm()' defined in the 'e1071' package of the software environment R (Meyer 2012), was employed. The parameter setting of the function 'svm()' was default. The manual evaluation results into three categories were used as classes. Features were linguistic properties demonstrating the linguistic complexity in terms of readability, and the automatically classified opinion polarity. The opinion polarity showed the degree of positiveness and that of happiness (from class 1 – absence of happiness – to class 100 – presence of happiness), which were identified with automatic classifiers (Östmar 2011; uClassify 2015). Table 4 shows the descriptive statistics for the linguistic properties of 158 opinions among the 351 opinions in the intercultural experience.

Linguistic properties		Minimum	Maximum	Mean (Standard Deviation)
Linguistic complexity	Flesch score	0.0	100.0	61.0 (18.0)
	Fog score	0.4	27.1	11.7 (4.4)
	Flesch-Kincaid score	0.0	23.7	9.0 (3.8)
	Dale-Chall score	0.1	14.6	7.5 (2.4)
Opinion polarity	Sentiment	4	100	57.1 (21.2)
	Mood	4	100	84.1 (21.6)

Table 4: Descriptive statistics for the linguistic properties of students' opinions ($n=158$)

The classifier was examined n times ($n=158$) using a leave-one-out cross validation test, considering one instance as test data and $n-1$ instances as training data. Table 5 shows the classification results where the accuracy was 83.5 percent ($((110 + 0 + 22)/158)$), which was higher than the majority class baseline (the negative, 70.9 percent, derived by $((110 + 0 + 2)/158)$).

Observed \ Identified	Negative	Neutral	Positive
Negative	110	0	2
Neutral	9	0	1
Positive	14	0	22

Table 5: Classification results

5. CONCLUSION

This study reported an opinion corpus for study-abroad program assessment and developed an evaluation method that predicts students' satisfaction by analyzing the opinions for study-abroad experiences. The classification accuracy demonstrated that corpus-based evaluation was promising for the study-abroad program assessment. The results of the opinion-polarity classification suggested that the opinion polarity was appropriately explained by the linguistic properties of readability scores and the automatic classification results of positiveness and happiness.

REFERENCES

- El-Halees, Alaa. 2011. Mining opinions in user-generated contents to improve course evaluation. In Jasni Mohamad Zain, Wan Maseri Wan Mohd and Eyas El-Qawasmeh eds. *Software engineering and computer systems. Part II, Communications in computer and information science*. Berlin: Springer-Verlag Berlin Heidelberg, 107–115.

- Engle, Lilli and John Engle. 2004. Assessing language acquisition and intercultural sensitivity development in relation to study abroad program design. *Frontiers: The Interdisciplinary Journal of Study Abroad* 10: 219–236.
- Kaewyong, Phuripoj, Anupong Sukprasert, Naomie Salim and Fatin Aliah Phang. 2015. The possibility of students' comments automatic interpret using lexicon based sentiment analysis to teacher evaluation. In *Proceeding of the 3rd International Conference on Artificial Intelligence and Computer Science*, 179–189.
- Leong, Chee Kian, Yew Haur Lee and Wai Keong Mak. 2012. Mining sentiments in SMS texts for teaching evaluation. *Expert Systems with Applications* 39: 2584–2589.
- Meyer, David. 2012. *Support Vector Machines*. The interface to `libsvm` in package `e1071`. [https://datajobs.com/data-science-repo/SVM-in-R-\[David-Meyer\].pdf](https://datajobs.com/data-science-repo/SVM-in-R-[David-Meyer].pdf) (accessed 27 July 2018)
- Östmar, Mattias. 2011. *Mood*. <https://www.uclassify.com/browse/prfekt/mood> (accessed 27 July 2018)
- Savicki, Victor and Elizabeth Brewer eds. 2015. *Assessing study abroad: theory, tools, and practice*. Sterling, VI: Stylus.
- uClassify. 2015. *The sentiment classifier*. <https://www.uclassify.com/browse/uclassify/sentiment> (accessed 27 July 2018)

Corresponding author

Katsunori Kotani
16–1 Nakamiyahigashino-cho
Hirakata, Osaka, Japan, 573–1001
e-mail: kkotani@kansai.ac.jp

received: November 2018
accepted: December 2018