



A Text Mining Examination of University Students' Learning Program Posters

Takehisa Kumakawa

Creative Engineering Education Center, Nagoya Institute of Technology, Nagoya, Japan

Email: kumakawa.takehisa@nitech.ac.jp

How to cite this paper: Kumakawa, T. (2017) A Text Mining Examination of University Students' Learning Program Posters. *Open Access Library Journal*, 4: e3639. <https://doi.org/10.4236/oalib.1103639>

Received: April 29, 2017

Accepted: June 6, 2017

Published: June 9, 2017

Copyright © 2017 by author and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

At present, applying text mining techniques to educational data is attracting much research attention. The present study uses text mining techniques to examine posters prepared by university freshmen in engineering fields to present their learning programs and their career goals after graduation, under the expectation that important keywords worth identifying lurked in the posters. The results showed that even though the participating students were only three months into their university education, their learning programs and career goals were already rather concrete and well adapted to the fields and courses they had chosen. Some of them had a remarkably good command of technical engineering terms.

Subject Areas

Education

Keywords

Engineering Education, Text Mining, Learning Program, Career Goal, Poster Presentation

1. Introduction

Over the last few decades, web-based learning has become more and more common and has been recognized as a potentially very effective educational method and resource. Web-based learning systems automatically collect and record a huge amount of data on students' learning behavior as students use them. To exploit this goldmine of educational data and use it to understand better how students actually proceed with learning, data-mining techniques have begun to be applied to educational data. This active research field is called *educational data mining* (Romero & Ventura [1]; Romero, Ventura, & García [2]; Baker & Yacef [3]; Romero, Ventura, Pechenizkiy, & Baker [4]; Romero & Ven-

tura [5]).

As a relatively recent development within this field, *text mining* techniques have been applied in educational research, allowing researchers to analyze text data such as formal text documents as well as informal ones like e-mails, chat messages, digital diaries, and online questions. Studies adopt a text mining approach to educational data include Hung [6], who used cluster analysis to examine extensive literature on e-learning, and Abdous and He [7] and He [8], who analyzed chat messages and online questions using text mining techniques, again including cluster analysis.

In line with these pioneering works, the present study examines university engineering students' posters describing their learning programs and career goals using text mining techniques. University freshmen prepared the posters to explain their individual learning programs and their career goals after graduation, and it can be expected that important keywords for our understanding of the students' learning status and progress worth picking out lurk in the posters.

2. Materials Used for Text Mining

2.1. Learning Program Posters

Posters were prepared by the students of Nagoya Institute of Technology. In July 2016, a “recital” was held where students presented their individual learning programs and their career goals after graduation—this was called the “C-plan.”¹ Each student prepared a poster the size of two A3 (11.7 × 16.5 inches) pages, which the present study employs as materials to which text mining was applied. Compared with the usual kinds of documents used in this approach, the amount of information in the posters might be a little limited, but it is nevertheless likely that the posters are sprinkled with important keywords worth picking out, because the students will likely have delicately considered and chosen the words they used due to space constraints. The text mining tool KH Coder was used for the analysis.

2.2. About Students

The authors of the posters were university freshmen enrolled in the Creative Engineering Education Program in the university's faculty of engineering. All students belonged to one or the other of the following two courses depending on their choice at their entrance examination: “Materials and Energy” (ME hereafter; 62 students) and “Computer and Social Engineering” (CS; 42 students) course. Specific topic areas covered by each course are listed in [Table 1](#).

3. Results

3.1. Frequently Appearing Words

Posters prepared by 104 students, pooled between the two courses, were employed for the analysis. The number of words extracted from the posters was

¹The term “C-plan” refers to the following three Cs: curriculum, career, and creativity.

Table 1. Specific areas covered by the two courses.

Materials and energy (<i>ME</i>); 62 students	Computer and social engineering (<i>CS</i>); 42 students
Life and materials chemistry	Networks
Soft materials	Computational intelligence
Advanced ceramics	Multimedia and human computers
Materials function and design	Architecture and design
Applied physics	Civil and environmental engineering
Electrical and electronic engineering	Systems management and engineering
Mechanical engineering	

17,983 in total, 2975 of which were unique. The 100 most frequently appearing words are summarized in **Table 2**. As seen in the table, overwhelmingly common words include “development” and “technology,” which seems natural in that the authors of the posters are students in the faculty of engineering. Some words such as “goal,” “study,” and “learn” would be used in a general sense to construct a learning program. It is noteworthy that even though the students were university freshmen only three months into their program, some of them had a good command of technical engineering terms such as “live body,” “sugar chain,” “catalyzer,” “macromolecule,” and “synthesis.”

To distinguish between general words on posters and specific words giving information on students’ learning programs and career goals and to examine how frequently given words were used by the students, a hierarchical cluster analysis was conducted it identified words that appeared in the posters at least 17 times and grouped them into five clusters, as shown in **Table 3**. The five clusters can be characterized as follows.

Cluster1 consists of the following five words: “challenge,” “present situation,” “change,” “value,” and “realization.” These words suggest that the students are highly motivated to create something new and valuable.

Cluster 2 consists of the following four words: “goal,” “career,” “study,” and “plan.” These words are commonly used among the students to construct posters.

Cluster 3 is characterized by the following typical words: “technology,” “development,” “universe,” “research,” “disaster,” and “earthquake.” These words suggest that the students are willing to work in research and development to deal with future risk or unknown territory. In particular, after the Great East Japan Earthquake in March 2011, they would have become more aware of disaster-prevention measures and the role of engineering therein.

Cluster 4 is characterized by the following typical words: “efficiency,” “method,” “power generation,” “nature,” “healthcare,” “cost,” “light,” “live body,” and “use.” These words suggest that the students are interested in improving existing technologies or saving energy and resources, in addition to creating whole new technologies/structures/concepts.

Table 2. Top 100 most frequently appearing words and their frequencies of appearance.

Word	No.	Word	No.	Word	No.	Word	No.
Development	141	Japan	25	Sugar chain	16	Action	11
Technology	126	Necessary	24	Various	16	Instrument	11
Goal	84	Solution	23	Utilization	15	Decrease	11
Study	79	Light	23	Make	15	Reality	11
Challenge	58	Power generation	23	Catalyzer	15	Contribution	11
Realization	54	Efficiency	21	Art	15	Few	11
Learn	51	Think	21	World	15	New	11
Present state	49	Design	21	Battery	15	Load	11
People	49	Earthquake	20	Frequent	14	Influence	10
Robot	46	Cost	19	Many	14	Sound	10
Career	45	Universe	19	Data	13	Home	10
Relationship	44	Nature	19	Safety	13	Strong	10
Research	44	Fuel	19	Subject	13	Macromolecule	10
Value	39	Employ	19	Slope	13	Synthesis	10
Life	38	Healthcare	17	Improvement	13	Accident	10
Change	34	Conduct	17	Structure	13	Acquisition	10
Application	32	Use	17	High	13	Treatment	10
Plan	31	Usage	17	Brain	13	Production	10
Present	31	Knowledge	17	Plastic	12	Product	10
Disaster	30	Method	17	Science	12	Occurrence	10
Automobile	30	Driving	16	Space	12	Disease	10
Problem	29	Have	16	Now	12	Drug	10
Possibility	26	Self-action	16	Especially	12	Panel	9
Human beings	26	House	16	Heat	12	Movement	9
Live body	26	New	16	Aspire	12	Medicinal product	9

Table 3. Hierarchical cluster analysis of words that appeared at least 17 times.

Cluster 1	Cluster 3	Cluster 4	Cluster 5
Challenge	Technology	Efficiency	Solution
Present situation	Development	Method	Problem
Change	Universe	Japan	Design
Value	Research	Power generation	Employ
Realization	Robot	Nature	Application
Cluster 2	People	Healthcare	Present
Goal	Disaster	Cost	Think
Career	Earthquake	Light	Knowledge
Study	Human beings	Live body	Necessary
Plan	Conduct	Learn	Automobile
		Use	Fuel
		Relationship	Usage
			Life
			Possibility

Cluster 5 is characterized by the following typical words: “solution,” “problem,” “design,” “application,” “think,” “knowledge,” “necessary,” and “possibility.” These words suggest that the students value knowledge and problem-solving thought to overcome present issues.

3.2. Differences between the Two Courses

For the analysis of differences between the two courses, their data were separated. Taking into consideration the frequently used words previously found, the following three coding rules were produced to formulate groups of words used in a similar context.

- Human beings: “disaster,” “earthquake,” “human beings,” “robots,” “people.”
- Technology: “development,” “technology,” “research,” “efficiency,” “cost.”
- Value creation: “challenge,” “present situation,” “change,” “value,” “realization.”

For example, according to the first coding rule, if a sentence in a poster contains at least one word such as “disaster,” “earthquake,” or “human beings,” the code “human beings” is given to the sentence.

Table 4 is a cross-tabulation table that compares the appearance ratios of codes under the two courses. As we can see from the table, the appearance ratio of the code “human beings” under the ME course was lower than that under CS. In contrast, the code “technology” was under ME than under CS. These results were statistically supported by chi-squared tests; both differences were significant at the 1% level. Overall, it appeared that CS courses are more human oriented while ME courses are more technology oriented, which is natural given the content of the courses. Finally, the appearance ratio of “value creation” was not significantly different across the two courses at the 10% level; that is, students in both courses used words related to value creation with approximately the same frequency.

4. Concluding Remarks

The present study examined university students’ learning program posters using text mining techniques. It was found that even though the students were university freshmen and only three months had passed between the beginning of their university engineering education and their preparation of the posters, their learning programs and career goals were rather concrete and well adapted to

Table 4. Cross-tabulation of frequency of appearance and appearance ratio of each code for each of the two courses.

	Human beings	Technology	Value creation	No. of observations
ME course	56 (5.80%)	181 (18.76%)	95 (9.84%)	965
CS course	56 (10.65%)	67 (12.74%)	46 (8.75%)	526
Sum	112 (7.51%)	248 (16.63%)	141 (9.46%)	1491
Chi-squared	10.808*	8.465*	0.361	

Note: *denotes significance at the 1% level.

their fields and courses. This result suggests that the majority of the students had thought ahead about their future careers before admitted to university, instead of only after.

The results of the present study could be enriched by the following expansions. First, it might be interesting to apply text mining techniques to the learning programs of students majoring in fields other than engineering and compare the results to the current results. Second, it would be useful to trace how students' career plans change as their education advances. These issues should be tackled by future research.

References

- [1] Romero, C. and Ventura, S. (2007) Educational Data Mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, **33**, 135-146. <https://doi.org/10.1016/j.eswa.2006.04.005>
- [2] Romero, C., Ventura, S. and García, E. (2008) Data Mining in Course Management Systems: Moodle Case Study and Tutorial. *Computers & Education*, **51**, 368-384. <https://doi.org/10.1016/j.compedu.2007.05.016>
- [3] Baker, R.S.J.D. and Yacef, K. (2009) The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, **1**, 3-16.
- [4] Romero, C., Ventura, S., Pechenizkiy, M. and Baker, R.S.J.D. (2010) Handbook of Educational Data Mining. CRC Press, Boca Raton, FL. <https://doi.org/10.1201/b10274>
- [5] Romero, C. and Ventura, S. (2013) Data Mining in Education. *WIREs Data Mining and Knowledge Discovery*, **3**, 12-27. <https://doi.org/10.1002/widm.1075>
- [6] Hung, J.-L. (2012) Trends of E-Learning Research from 2000 to 2008: Use of Text Mining and Bibliometrics. *British Journal of Educational Technology*, **43**, 5-16. <https://doi.org/10.1111/j.1467-8535.2010.01144.x>
- [7] Abdous, M. and He, W. (2011) Using Text Mining to Uncover Students' Technology-Related Problems in Live Video Streaming. *British Journal of Educational Technology*, **42**, 40-49. <https://doi.org/10.1111/j.1467-8535.2009.00980.x>
- [8] He, W. (2013) Examining Students' Online Interaction in a Live Video Streaming Environment Using Data Mining and Text Mining. *Computers in Human Behavior*, **29**, 90-102. <https://doi.org/10.1016/j.chb.2012.07.020>



Submit or recommend next manuscript to OALib Journal and we will provide best service for you:

- Publication frequency: Monthly
- 9 [subject areas](#) of science, technology and medicine
- Fair and rigorous peer-review system
- Fast publication process
- Article promotion in various social networking sites (LinkedIn, Facebook, Twitter, etc.)
- Maximum dissemination of your research work

Submit Your Paper Online: [Click Here to Submit](#)

Or Contact service@oalib.com