Analysis of 'Better Class' Characteristics and Patterns from College Lecture Evaluation by Longitudinal Big Data

Min-Woo Nam

Department of Education Daejeon University, Daejeon, 34520, South Korea

Eun-Soon Cho

Department of Education Mokwon University, Daejeon, 35349, South Korea

ABSTRACT

The purpose of this study was to analyze characteristics and patterns of 'better class' by using the longitudinal text mining big data analysis technique from subjective lecture evaluation comments. First, this study classified upper 30% classes to deduce certain characteristics and patterns from every five-year subjective text data for 10 years. A total of 47,177courses (100%) from spring semester 2005 to fall semester 2014 were analyzed from a university at a metropolitan city in the mid area of South Korea.

This study extracted meaningful words such as good, course, professor, appreciation, lecture, interesting, useful, know, easy, improvement, progress, teaching material, passion, and concern from the order of frequency 2005-2009. The other set of words were class, appreciation, professor, good, course, interesting, understanding, useful, help, student, effort, thinking, not difficult, explanation, lecture, hard, pleasant, easy, study, examination, like, various, fun, and knowledge 2010-2014. This study suggests that the characteristics and patterns of 'better class' at college, should be analyzed according to different academic code such as liberal arts, fine arts, social science, engineering, math and science, and etc.

Key words: Course Evaluation, Longitudinal Big Data, Text Mining, Better Class' Characteristics.

1. INTRODUCTION

College course evaluation by students has a useful meaning to the instructor who is in charge, and the college administration to maintain a course quality control. Usually the purpose of the evaluation is to find out whether course objectives are achieved, and course is delivered effectively according to the lesson plan [1].

The attention of college course evaluation has been increasing in terms of providing information for faculty appointment and promotion as well as the demand for the course improvement each year. Currently the results of students' course evaluation are a decisive formal quality control system at college to improve teaching ability for course instructors [2], [3]. The results of course evaluation are usually reported to individual faculty in the form of quantitative score base, not including the form of qualitative information in terms of explanations on what should be revised and corrected for better class management. This means that the subjective expression of

At most college, course evaluation results have been automatically accumulated by quantitative big data base because most of students should fill out the course evaluation in the form of questionnaire and essay type questions right after the semester ends. However, since the quantitative data is utilized as an important administrative purpose first, the essay type of students comments are usually used as not important reference data both to the instructor and to the administration. It is needed to provide proper guideline for 'better class' not only from the pattern analysis of quantitative big data source, but also from the interpretation of qualitative data source by text mining.

This study used text mining technique to analyze unstructured data of course evaluation response questions. Text mining extracts highly frequent words by statistical analysis from natural language type sentences. The technique is able to

students about the course is not delivered to the faculty with exactly what students want, and needs for the class. Therefore, the evaluation results cannot be utilized as useful feedback data to the instructor to make them realize what is better class for getting students interests and providing practical support to students, and how the instructor should revise the lesson plan [4].

^{*} Corresponding author, Email:echos@mokwon.ac.kr Manuscript received Mar. 04, 2019; revised May. 21, 2019; accepted May. 31, 2019

figure out overall framework of the task from big data abalysis of highly frequent words, and this study selected this data analysis technique considering of unstructured words to be used in this study.

Recently, educational research widely used text mining techniques both to reason meaningful concepts and characteristics of certain tasks, and to analyze final results and products of the tasks to match the purposes of the research [5]-[7]. This study analyzed many years of bag data on lecture response questions to find out the characteristics of 'better class' based on the students' perception.

The purpose of this study was to analyze text mining approach from course evaluation qualitative big data source to suggest what were the characteristics of 'better class' at college level. This study categorized upper 30% of course evaluation scores for 10 years at every semester, and analyzed students answer texts from qualitative evaluation items. This study used techniques such as frequency sampling, words ordering, and any pattern change for two sets of 5 year net data to identify the pattern and characteristic of 'better class' from the longitudinal text mining analyses.

2. RESEARCH METHODS

2.1 Sampling

The data in this study was collected from the results of course evaluation between the year of 2005 and 2014 at university located in the middle area of South Korea. The starting year 2005 means especially to provide the strong administrative demand utilizing instructional strategies such as application of ICT (Information Communication Technology), constructivism oriented lesson design, and SDL (Self-Directed Learning) [8], [9] to most of courses. Also. 10 years were needed to analyze the accumulated longitudinal data to standardize pattern and characteristics of text mining analysis. The reason of grouping and analyzing from 2010 to 2014 was to reflect from diffusing of government financial support to colleges and universities to reinforce teaching and learning competencies. Considering of 10 years accumulated big data, this study divided two 5 years cycle to apply innovative changes of college and university education. The total number of the courses were 47,177.

The statistical data of this study extracted from course name, open semester & year, and evaluation score & comments by excel format. Fig. 1. shows this results.

	₩ \$ 02 \$07\$00\$ ou \$00 2x x7 Accts													
#0°	150	80	29 - 13	* 3 3	*** (** (**)	148	90			표준	ų.		보통 성명 역스	08 828 B XA604 Y A
	140	24 1	2 2 E	4 7	B · · · · · · · · · · · · · · · · · · ·	784 98		1 11 2	用版· 符		- 4			2441, 86-94-
	A1	- 1	9 B 76	7145	28		9 E4	84					수막물	4 81
	Al	R	5 A 7	0	,	F	G	Н		-			M	
-	year		course name	placemen	nt major area				conize	lecture	number	evaluation	evaluati	on comments
1	-		COMPANIAN IN	-					dassificatio		-	type	score	Name of the second seco
2	2004	- 2		11	SEGR	38	220	時音	66	이론	29	४पडे न	3602	역 있의 부분도 가르체주셔서 강의가 더 몸다로들습니다.
3	3054	- 2		11	응문대학	28	22¢(4)	時間	58	이론	29	8987	3652	발표가 이상함 조 관성매주고 개별 발표하니 왜 구지 조 관성을 했는지 이때들가 작 중 수정 됐으면
4	2004	- 2		11	인문대학	28	22¢(4)	108	55	이른	29	8987	1612	수업 책이 쉬운 책이 아니라 여러운 책이었습니다. 시간이 폭짝해서 그러신지는 요료겠는데, 내용이 다른 소약되는 것은 환기 나는데 작곡소약에서 여러나 소약용기가 힘이 됩니다.
5	3054	2		11	389 9	85	22¢	時	85	이론	29	8987	3612	수업 30분전에 표장 문자를 날리셔서 당황스러웠습니다.
6	2054	- 2		11	사용대학	85	20	時	55	이론	16	8987	1431	최근에 허당하는 현대소설들을 읽을 수 있는 기회가 되어서 좋았다.
7	3054	2		11	사용대학	88	24	粮	85	이른	16	8987	3438	현대문학을 많이 읽고, 실제로 엄용시합에도 도움이 되었다. 또한 부모님 명한 쓰기는 소는 동안 부도
8	2054	- 2		11	사회교학대학	85	220	時	66	이문	58	8987	3.223	수업시간에 학생들의 눈눈이에 맞추어서 수업해주시면 좋겠습니다. 아직 현공 지식을 배워가는 단계
9	3054	2		11	공고대학	88	조교수(비)	時	85	야흔	58	8987	3203	문의사항에 착극적으로 다용해주셨으면 하고 짜끔내와 계획적인 강의가 필요해 보입니다.
10	3054	- 2		11	사용대학	85	22¢(t)	時	88	이분	32	8987	423	강화를 듣는데 있어서 있뿐지 일찍 시작하거나 의분을 늦게 끝나거나 하도록 조번에 스케쥴을 잡으
11	3054	2		11	용약대학	88	부교수	時	83	이론	12	성약용가	4203	수업의 진행 기준을 조금 못하는 학생에게도 맞춰졌으면 즐겁습니다.
12	2054	2		11	미술·디자인대학	비찬용	시간장사	25%	25	이본+일기	19	8987	4422	콘테디자인이라는 것에 대해 선배들에게 함보면 향할 중요하다. 홈페든 기본이다. 라는 말씀을 많이 다스러워들도 잘 수 있는 제기가 되었습니다. 오세하기의 의치와 첫달 것나하면서도 조용하고 이번
13	2004	2		12	공고대학	비찬림	시간장사	独	83	이론+일기	10	성약용가	4429	교수님 수많은 형탈 배워가는것 많은 수업이라 너무 좋다요~ 4학년대 또 교수님 수업 있었으면 좋겠
14	3054	- 2		12	HERRIT	비찬용	시간장사	298	88	이론+설기	10	성덕용가	4429	많이 배워가는 수업이다. 돈이 아랍지 않은 수업이고 설기수업이 좀 더 많아지고 당전히 교수님의 다

Fig. 1. Course evaluation data sampling example

2.2 Methodology

To analyze students' comments on course evaluation, text mining techniques such as an Elastic Search system, EunJeon Korean language morpheme analysis technique, and Mysql, Apache-PHP for search engine, were used. These techniques were applied to extract evaluation comments words and to analyze frequencies for those extracted words. Fig. 2. explained the framework of text mining system developed in this study [5].

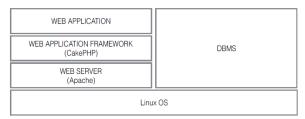


Fig. 2. Framework of course evaluation analysis system

Elastic Search is an open source search engine developed by Apache Lucene base with supporting functions of expansion, high fusibility, multi tenancy, text oriented, and strong professional search. Words extraction and frequency were analyzed by Korean language morpheme analyzer and category (or part of speech) tagging. Morpheme analysis means of figuring out language attributes such as the root of a word, prefix, suffix, and POS (part-of-speech). POS tagging is to make markup words considering of morpheme meanings and contexts of words. For example, Korean sentence 가방에 들어가신다 → 가방/NNG + 에/JKM + 들어가/VV + 시/EPH + 니다/EFN [10].

The Korean language representative open source Elastic Search 'EunJeon' has features in proper speed and qualitative morpheme analysis, and free license. Fig. 3. shows the process of EunJeon Korean language analysis with the formation of comments DB \rightarrow Eunjeaon Elastic Search \rightarrow morpheme analysis DB [5].



Fig. 3. Morpheme analyzing system for course evaluation comments analysis

From Fig. 3, the total number of comments data for 10 years was 265,939 cases, and 6.726,965 morphemes separated from 'EunJeon' Elastic Search. Fig. 4. also shows the accumulated output data from the process of searching condition input \rightarrow data extraction \rightarrow morphemic frequencies \rightarrow output [5].



Fig. 4. Output process of the data in course evaluation analysis system

Fig. 5. shows the screen of Elastic Search system and 'Eunjeaon' morpheme system to be used in this study to analyze course evaluation comments by text mining technique [10].

starting year	semester	ending year	semester
2006 -	1학기	- 2014	- 2학기
course evaluation level	class size	course classification	student year
전체 •	전체	• 전체	▼ 전체
major area	lecture type		
전체 •	전체	•	
Submit			
Coorne			
Tokens			

Fig. 5. The developed screen of course evaluation comments text analysis system

The screen from fig. 5. was designed for data analysis and extraction morphemes for the indication of searching conditions such as semester-year, lecture evaluation score levels (upper 30%, mid 40%, lower 30%), lecture size (small: 1-35 students, standard: 36-50 students, large: 51+ students), course requirements (general education courses, major education courses), student year (freshmen, sophomore, junior, senior), major areas (liberal arts, social science, teacher education, science and engineering, fine arts), and lecture type (lecture only, lecture+practical training, lecture+lab, practical training).

Words and frequencies by morpheme analysis were generated from the unstructured students evaluation comments, for example, 'submit' button with the input of upper score 30% and a certain semester-year specification, provides the most referred words orderly in a certain specified lecture.

3. RESULT

3.1 Analysis results of 'better class' characteristics from year 2005 to year 2009 by lecture evaluations

This study categorized 10 years accumulated lectures by three level searching conditions such as upper 30%, mid 40%, and lover 30% from the evaluation scores by students. Then, students' unstructured lecture evaluation comments for 10 years, were analyzed to extract the most referred words by students and their frequencies. The total number of analyzed morphemes used in this study was 6,726,965, and those words were reclassified by lecture traits and quality to be used as

useful morphemes. Meaningless auxiliary words and some other unidentified words were excluded.

This study analyzed 5 year periods traits and change transition of 'better class' from 10 year accumulated longitudinal data to examine evaluation word pattern recognitions and sustainable characteristics of 'better class' reflected by social demands toward college education.

Table 1. Text mining results of Upper 30% lectures for first 5 years

years						
No		word	frequency			
	1	good	32,863			
	2	class	19,391			
	3	professor	14,145			
	4	appreciation	11,423			
	5	lecture	10,280			
	6	<u>fun</u>	9,676			
	7	<u>useful</u>	9,002			
	8	know	7,645			
	9	<u>easy</u>	7,045			
	10	<u>help</u>	6,747			
	11	<u>understand</u>	5,718			
	12	student	4,932			
	13	work hard	3,854			
	14	not difficult	3,609			
	15	<u>new</u>	2,414			
Year 2005 -	16	<u>passion</u>	2,136			
Year 2009	17	study	2,134			
	18	<u>need</u>	2,009			
	19	<u>interest</u>	1,734			
	20	<u>knowledge</u>	1,713			
	21	<u>presentation</u>	1,702			
	22	<u>improvement</u>	1,646			
	23	progress	1,521			
	24	<u>text</u>	1,257			
	25	<u>passion</u>	1,173			
	26	opportunity	1,155			
	27	concern	1,151			
	28	not boring	1,086			
	29	society	1,071			
	30	<u>theory</u>	1,065			
	31	<u>confidence</u>	1,047			
	32	to reform	1,033			

Table 1 explains the most referred words with more than 1,000 frequencies out of upper 30% lectures by text mining results from year 2005 to year 2009. From table 1, the most referred words were good, class, professor, appreciation, lecture, fun, useful, know, easy, help, understand, student, work hard, not difficult, new, passion, study, need, interest, knowledge, presentation, improvement, progress, text, passion, and opportunity in order of frequency.

Except the neutral words such as good and professor, table 1 indicated that the proper words for a 'better class' were fun, useful, help, understand, not difficult, new, passion, need. Interest, knowledge, presentation, improvement, text, passion, concern, theory, and confidence. These words were underlined

on the table 1, and extracted from the students' essay type evaluation answers as followings.

'The lecture was <u>fun</u> and very <u>helpful</u> because the profession explained very well. (student 1)'

'The lecture was very <u>useful</u> and the professor explained the content very <u>easily</u> and it was <u>helpful</u> to <u>understand</u> the class. (student 2)'

'Even though the project was very difficult, the professor explained the content very <u>easily</u> and it was <u>helpful</u>. (student 3)'

'The professor explained <u>easily</u> with extra materials for students like me to make better <u>understanding</u>. (student 4)'

'The professor used <u>easy</u> examples to make me <u>understand</u>. (student 5)'

'The professor selected <u>not difficult text</u>. (student 6)'

'The lecture was **not difficult**. (student 7)'

'The lecture included \underline{new} $\underline{knowledge}$ and very \underline{useful} (student 8)'

'The professor was very <u>passionate</u>, but it was a little hard. (student 9)'

'The professor taught selected important materials, and it was very well <u>understandable</u>. (student 10)'

'The professor provide real world new information and it was *interesting*. (student 11)'

'I could improve my academic <u>knowledge</u> level. (student 12)'

'The lecture was <u>fun</u> and memorable, and <u>text</u> was <u>helpful</u>. (student 13)'

3.2 Analysis results of 'better class' characteristics from year 2010 to year 2014 by lecture evaluations

Another set of analysis was conducted with the same method as on first 5 year analyses. Following table 2 explains another set of the most referred words from upper 30% courses between year 2010 and year 2014.

Table 2. Text mining results of Upper 30% lectures for second 5 years

No		word	frequency
	1	lesson	14,751
	2	appreciation	13,606
	3	professor	10,350
	4	good	6,354
	5	class	6,262
	6	<u>fun</u>	5,353
	7	understand	4,396
	8	<u>useful</u>	4,336
Year 2010 -	9	<u>help</u>	4,192
Year 2014	10	student	3,337
16a1 2014	11	work hard	3,057
	12	<u>thinking</u>	2,750
	13	not difficult	2,230
	14	explanation	1,519
	15	course	1,481
	16	<u>eagerly</u>	1,468
	17	<u>pleasant</u>	1,323
	18	<u>easy</u>	1,307
	19	study	1,186

No		word	frequency
	20	<u>examination</u>	1,176
	21	like	1159
	22	<u>variety</u>	1087
	23	<u>passion</u>	1,035
	24	<u>interest</u>	1,028
	25	<u>knowledge</u>	1,013

From table 2, except the neutral words such as good and professor, proper words for a 'better class' were fun, useful, help, thinking, not difficult, explanation, eagerly, pleasant, easy, examination, variety, passion, interest, and knowledge, These words were underlined on the table 2, and extracted from the students' essay type evaluation answers as followings.

'The lecture was **fun** and **interesting**. (student 1)'

'The lecture was very well <u>understandable</u> with many example cases and video materials. (student 2)'

'It was very <u>help</u> and learned a lot of <u>knowledge</u>. (student 3)'

'It was useful and helpful to me. (student 4)'

'The lecture was <u>helpful</u> to pass a license test and to master the techniques. (student 5)'

'The professor made students to <u>think</u> and share with other students' various <u>thought</u>. (student 6)'

'The professor explained very easily. (student 7)'

'The lecture was very <u>useful</u> because the professor delivered with <u>passion</u>. (student 8)'

'The examination discriminated students. (student 9)'

'The professor taught us with <u>various</u> teaching skills and with many different materials. (student 10)'

'The professor prepared the class very eagerly and passionately. (student 11)'

'It was interesting comparing to other class. (student 12)'

3.3 Word cloud results of 'better class' characteristics from year 2005 to year 2014 by lecture evaluations

Fig. 6 and 7 show the text mining results categorized by 5 year terms from 2005 to 2014 by Claude form. The bold words are the core words from word Claude visualization except vague neutral words. They are fun, understand, useful, easy, help, not difficult, passion, interest, and etc.



Fig. 6. Upper 30% courses text mining results categorized by 5 year terms from year 2005 to year 2009 by Word cloud form



Fig. 7. Upper 30% courses text mining results categorized by 5 year terms from year 2010 to year 2014 by Word cloud form

4. CONCLUSION

This study analyzed the characteristics and patterns of 'better class' by text mining analysis to find what are the most referred evaluation comments words and frequencies of those words out of upper 30% evaluation scores by college students based on 10 year longitudinal data.

First, this study found the meaningful and significant key words such as fun, easy, understand, useful, passion, and etc. from the analyses. These words can be explained that students perceive a 'better class' means learning materials by delivering easily, fun, useful, and passionate. Therefore, students are able to inspire interests and passion from the class accordingly. These words were similar results from TV show about 100 famous college lectures core elements at 2012 [11]. It emphasized 'point out the main concepts with easy, repeatedly, passionate, and with care'. These words are the decisive characteristics of 'better class (lecture)'. What are the better lecture'? This study finalized the 'better lecture' should be easily understandable. fun to learn, and delivered with professors' passion. This result supports the fidings from the previsous research such as [5], [11].

Secondly, students perceived 'better class' as easy, understandable, and useful class during first 5 years from year 2005 to year 2009, while for second 5 years from year 2010 to year 2014, there were other key words such as fun, pleasant, and interest added up to. The words, Knowledge and text from year 2005 to year 2009, were not mentioned with more than 1,000 frequencies during second 5 years. This result can be assumed that students activities have been changed from text oriented lecture type to newly recommended students' active participation in the class during that time because college education emphasized more student oriented teaching skills and learning activities to college faculty. In other perspective, college evaluation system has set most of college lectures toward criterion reference test system. This means that students are more concerned on test difficulty, discrimination, and fairness.

Thirdly, this study also indicated that a 'better class' at college is able to be controlled by faculty members in all times. At most of colleges, instructors are domain experts, not

teaching expert. Although teaching and learning centers at college provides various teaching strategies and student oriented teaching skills, faculty members passively participate, and are not interested. Therefore, reflective feedback from students should be delivered to faculty members and provide proper class consulting to improve their classes.

Finally, this study suggests that the characteristics and patterns of 'better class' at college, should be analyzed according to different academic code such as liberal arts, fine arts, social science, engineering, math and science, and etc. The 'better class' can be interpreted in different way based on the academic areas, and provide different delusion to the faculty when they have student evaluation comments on the specific lecture. However, this research can provide very important implications to college education what and how it moves to provide better qualified education to students. Also, when considering Korean college environment in terms of future population cliff, college education should focus on the fundamental teaching and learning basic strategies, and therefore increasing student satisfaction.

REFERENCES

- [1] Y. J. Jang, Analysis of the Faculty and Student's Perspectives on Student Course Evaluation, master's dissertation, Korea University, Seoul, Korea, 2012.
- [2] H. R. Min, B. M. Yu, and M. J. Shim, "A study on the factors of best Lecture for improving the educational quality of University Teaching," The Journal of Educational Information and Media, vol. 11, no. 3, 2005, pp. 41-65.
- [3] S. I. Han, H. J. Kim, and J. Y. Lee, "A Comprehensive Study of Korean Students' Evaluations of University Teaching," The Journal of Educational Administration, vol. 23, no. 3, 2005, pp. 379-403.
- [4] M. W. Nam and E. S. Cho, "A Study of College Semesterend Lecture Evaluation Weight Control and Class Size Difference Verification," Korean Journal of the Learning Sciences, vol. 8, no. 2, 2014, pp. 153-168.
- [5] H. D. Lee and M. W. Nam, "Better Class's Characteristics by Major Field based on the Analysis of Text Mining from College Course Evaluation Subjective Results," Korean Journal of Early Childhood Education, vol. 20, no. 2, 2018, pp. 21-41. doi: 10.15409/riece.2018.20.2.2
- [6] E. J. Baik, "Pre-service Early Childhood Teachers' Science-Teaching Efficacy and Attitude during a Course on Nuri Curriculum Content's PBL 'Early Childhood Science Education," Korean Journal of Children's Media, vol. 15, no. 2, 2016, pp. 1-20. doi: 10.21183/kjcm.2016.06.15.2.1
- [7] S. K. Lee, "Exploring the Possibility of Using Text Mining in Writing Assessment," Research on Writing, vol. 35, 2017, pp. 99-131. doi: 10.31565/korrow.2017.35.004
- [8] J. H. Leem, "Teaching & Learning Strategies for increasing the effectiveness of instruction using ICT," Journal of elementary education studies, vol. 9, no. 2, 2002, pp. 19-42.

- [9] Korea Education and Research Information Service, *Edunet Policy Research*, 2002.
- [10] KoNLPy, Morphological analysis and POS tagging, Retrieved Jun. 28, 2018. from http://konlpy.org/ko/ v0.4.3/morph, 2014.
- [11] Korean Educational Development Institute, *Best lectures:* SBS TV special program about 100 famous college lectures, Hakjisa, Seoul, 2014.



Min-Woo Nam

He received his doctoral degree in Educational Technology in Konkuk University, Seoul(2010). He is a professor at the Center for Teaching & Learning, Daejeon University in Daejeon, South Korea. His current research interests are in e-Learning,

Learning Analytics and Educational Data Mining.



Eun-Soon Cho

She received her doctoral degree from University of Connecticut in 1992. She worked for Aetna Group in U.S. from 1992-1993 and POSCO Research Institute as a senior researcher from 1994-1999. Currently, she is a professor at the department of education, Mokwon

University in Daejeon, South Korea. She is also a Korean university accreditation screening & review committee member. Her interests are in instructional design and implement, and HRD.