Mining Concept Maps to Understand University Students' Learning

Jin Soung Yoo
Indiana University-Purdue University, Fort wayne
2101 Coliseum Blvd East,
Fort Wayne, Indiana 46805
yooj@ipfw.edu

Moon-Heum Cho
Kent State University, Stark
6000 Frank Avenue, N. W.
North Canton, Ohio 44720
mhcho@kent.edu

ABSTRACT

Concept maps, visual representations of knowledge, are used in an educational context as a way to represent students' knowledge, and identify mental models of students; however there is a limitation of using concept mapping due to its difficulty to evaluate the concept maps. A concept map has a complex structure which is composed of concepts and their relationships that often have a weighted direction. This work explores the feasibility of the analysis of concept maps using data mining methods, and investigate the possibility of using concept maps as a research tool to understand college student's learning. A total of 111 college students participated in this study. The findings from frequent concept mining and sub-concept map mining suggest that students expect a traditional way of learning. The study also shows a promising area of further study in the area of data mining in education.

1. INTRODUCTION

Understanding how students learn is an important issue in college education [14; 15]. Perhaps, the most common methods to be used to understand learning of college students were both interviews [14] and self-reporting measurements such as Learning and Study Strategies Inventory (LASSI) [20] or Motivated Strategies for Learning Questionnaire (MSLQ) [15]. Recently, however, researchers agreed upon the importance of alternative methods to understand students' learning [21; 24].

Concept maps which are visual representations of knowledge are used in educational contexts, as a way to represent students' knowledge, and identify mental models of students [13]. It includes concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line that links two concepts [12]. Figure 1 shows an example of concept map. The map depicts the way a student studies or prepares for an exam. Using nodes and links, students can visualize their ideas and understanding of the content. Because the concept maps represent students' mental models of content and their ways of problem solving, concept mapping has been widely used in teaching and learning process in academia ranging from K-12 to higher education [7; 11].

Although concept mapping is a powerful tool, there is a limitation of using concept mapping due to its difficulty to evaluate the concept maps.

The most common method to evaluate students' concept maps in educational setting might be using rubrics; however, this also has limitations. First, although educators may have a general understanding of their students' mental models from using rubrics, educators are not provided specific and detailed information about the relationships among the concepts. Second, although rubrics reduce the amount of time educators spend evaluating concept maps, evaluation is still very time-consuming using the concept maps, particularly, in a large classroom where students number over 50. Third, when using rubrics, evaluation may not be easy for the educators to determine certain patterns of the students' mental models. Each person's concept map will look differently, although they may have similar patterns. Consequently, there is a strong need of computerized methods to evaluate students' concept maps.

In educational science, Yin et al. [23] examined the equivalence of two construct-a-concept-map techniques: construct-a-map with created linking phrases and construct-a-map with selected linking phrases. Some works [10; 18] explored different scoring systems for concept maps. In educational engineering, several studies [5; 19; 4] describe the construction of concept maps. Alves et al. [2] presents a system called TextStorm, which extracts raw concept maps from text. Shen et al. [16] works on the automatic generation of concept maps through text mining techniques. These studies concern the generation of concept maps rather than the analysis of concept maps.

Data mining is the process of automatically extracting new and useful knowledge hidden in large datasets. Performing data mining on concept maps can provide useful information in understanding the thought processes which generate

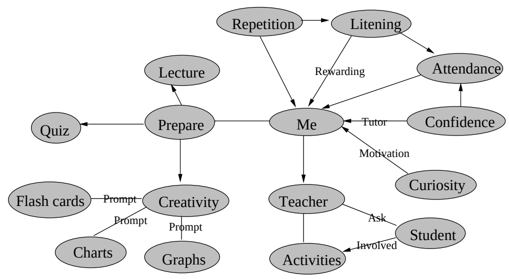


Figure 1: An example of concept map

them. In this study, we explore the feasibility of the analysis of concept maps using data mining methods, and investigate the possibility of using concept maps as a research tool to understand college student's learning.

2. DATA COLLECTION

We first describe participants in this study, concepts provided to them, and data collection procedure.

2.1 Participants

Data were collected from 10 college Critical Inquiry (CI) courses designed to help underprepared students to either acquire study skills or efficacy/motivation. The CI courses was chosen due to the large enrollment members, of approximately 300 students per semester, and they provide specific learning contexts through pairing the CI courses with other subject courses such as psychology, chemistry, biology, education, and archeology. The primary purpose of pairing the CI courses with other subjects courses is to facilitate student to learn better in a specific subject area. A total of 111 students participated in the study by drawing their concept maps. Based on the information students provided, 58 students were females, most students were Caucasian (n=83), and most students were freshmen (n=94).

2.2 Concepts

A total of 112 concepts were designed for this study. We reviewed the concepts with the two instructors who taught the study skill course. The 112 concepts consist of seven categories including classroom learning (e.g., listening, reading, or discussion, total 41 concepts), action for study (e.g., self-explanation, annotations, or memorization, total 18 concepts), learning tools (e.g., notes, charts, or textbooks, total 22 concepts), internal trigger (e.g., inquiry, curiosity, or creativity, total 4 concepts), motivation (e.g., will, confidence, or inspire, total 12 concepts), school facilities (e.g., library, web, or writing center, total 7 concepts), or people (e.g., teacher, classmate, or parent, total 8 concepts).

2.3 Procedures

Two weeks after mid-term exam week, we visited each class and administered the concept maps to students. In each class, we provided a 10-minute orientation about concept maps and drew one map for students on a whiteboard as a demonstration. Students were asked to respond to the question "*how do you learn in a college class?*". They drew their concept maps with pen and pencil since the class room was a regular class room that didn't have computers for individual students.

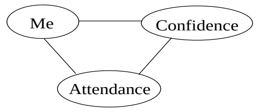
3. CONCEPT MAP MINING

Through discussions with education researchers, we listed their interesting queries to the students' concept maps. In this work, we focused on two problems: "Which concepts are frequently used in students' concept maps?" and "Which sub concept structures are commonly observed in students' concept maps?". For answering the first question, we conducted frequent item set mining task [1] to the concept map data, and for the second question, we used sub-graph mining task [9].

3.1 Data preprocess

TID	Items
1	{Repetition, Listening, Attendance, Me, Confidence, Teacher}
2	{Prepare, Lecture, Me, Quiz, Teacher, Flash cards, Charts}
3	{Assignments, Study, Friend, Teacher, Notes, Activities}
4	{Quiz, Prepare, Lecture, Creativity, Me, Teacher, Time, Note}
...	...

(a) Transaction data



t
v 0 Me
v 1 Confidence
v 2 Attendance
u 0 1 E
u 0 2 E
u 1 2 E

(b) Graph data

Figure 2: Input data formats for data mining

After collecting students' hand-drawn concept maps, we digitalized them. Concept maps can be represented by graphs, consisting of nodes(or vertices), which represent concepts, and arcs(or edges), which represent relationships between the concepts. In the graphs of concept maps, vertices should have labels associated with them. Edges may have associated labels and directions represented by arrows. Vertices and edges may have their own weighted value to designate their significance.

Although a list of predefined concept names and relationship names were provided, students misspelled some names or gave wrong concept and relationship names different from the predefined names. We also noticed, that in several instances synonyms or plurals of the same word were used as labels, and that some students did not follow a standard in the labeling of nodes and edges or in the use of arrows to denote direction on the edges. Our digitalization process checked all inconsistent concept names and relationship names in the concept maps.

3.2 Frequent association concept mining

In order to find common concepts students have used for their concept maps, we applied a methodology known as association analysis to the concept data. Association analysis task in data mining is useful for discovering interesting relationships hidden in large data sets. The uncovered relationships are represented in the form of association rules or a sets of frequent items [17; 1]. For example, a frequent item set, {Lectures, Notes} suggests that a strong relationship exists between lecture and note in learning strategy. The frequency of the associated items is often measured with support. The *support* of an item-set x is defined as the fraction of all transactions that contain x, i.e., $s(x)=\frac{sc(x)}{N}$, where N is the total number of transactions, and $sc(x)$ is the *support count* of x, $sc(x)$ $=|\{t_i|x \subseteq t_i, t_i \in T\}|$. If the support of item-set x is greater than a given support threshold, x is called a frequent item-set.

For the association analysis, data should be prepared with transaction data format. We transformed our digitalized concept map data to concept transaction data as shown in Figure 2 (a). Each row in this table corresponds to a transaction that contains a unique identifier labeled TID, and a set of concepts used by a student in drawing his/her concept map. Here, a relationship between two concepts is not included in the transaction. There are many algorithms for association analysis [1; 3; 17]. We used Apriori algorithm [1] for our analysis. Four different minimum frequency thresholds 30%, 40%, 50% and 60% were used.

3.3 Frequent sub-concept map mining

The second analysis is performed in order to derive a set of common sub structures among the collection of concept graph data. We applied frequent sub-graph mining task [17; 6] to our concept map data. Each concept map can be represented as a graph data as shown in Figure 2 (a). A concept graph $G = (V, E)$ is composed of a concept vertex set V and a set of edges E connecting between pairs of vertices. A graph $G' = (V', E')$ is a sub-graph of another graph $G = (V, E)$ if its concept vertex set V' is a subset of V and its edge set E' is a subset of E. The frequency of a sub-graph is also measured by support. The support for a sub-graph g is defined as the fraction of all graphs that contain g as its sub-graph, i.e., $s(g)=\frac{sc(g)}{|GD|}$, where GD is a collection of graphs, and $sc(g)$ is the support count of g, i.e, $sc(g)=|\{G_i|g \subseteq G_i, G_i \in GD\}|$. Sub-graphs (sub-concept map structures) g such that $s(g) \geq minsup$ are frequent sub-graphs. There are several algorithms for frequent sub-graph mining such as FSG [9], gSpan [22] and SPIN [8]. We fed the concept graph data to the FSG algorithm [9]. Figure 2 (b) shows an example of input data format for the FSG algorithm. The required order for a valid graph data begins with 't' followed by all vertexes in the graph and finally by all the edges. Three different frequency thresholds, 10%, 20% and 30% were used for our analysis.

4. RESULTS

107 concept maps among 111 maps were analyzed. Four concepts maps were deleted because the concepts were not recognizable or students did not follow the instructions. Given the 112 concepts, a total number of 110 concepts were used. The maximum number of concepts and lines a student used were 39 and 31 respectively in each concept map, and the average concepts and lines students used were 13 and 14 respectively.

Given 112 concepts, only 15 concepts: 'teachers', 'me', 'notes', 'lectures', 'time', 'repetition', 'textbooks', 'listening', 'flashcard', 'parents', 'reading', 'annotation', 'study', 'tutor' and 'activity' concepts are used by students with 30% frequency. Among association patterns having two concepts, {teachers, me}, {teachers, notes} and {me, notes} showed very high frequency of over 60%. Among patterns having three concepts, {teachers, me, time}, {teachers, lectures, repetition} and {teachers, textbooks, repetition} showed over 50% frequency. Two concept sets having four items, {teachers, lectures, repetition, time} and {teachers, textbooks, repetition, time} showed around 30% frequency.

In the sub-graph mining, the size of the patterns is measured by the number of edges. Table 1 shows sub-concept map structures with at least 10% frequency. When the threshold was set to 30%, three types of frequent sub-concept map patterns were found such as "me-listening"(sc = 34 out of 107), "teacher-lectures"(sc = 44), and "me-notes"(sc = 49). When the threshold comes to 20%, a total of 15 frequent sub-concept map structures were observed. Among them, 12 sub-graph structures were about 'me' strategies or actions, 2 sub-graphs were about 'teacher' related patterns, and 1 sub-graph was between me and teacher. Last, when the threshold was 10%, a total of 47 simple sub-graphs were discovered where 36 sub-graphs were 'me' initiated concepts, 6 sub-graphs were 'teacher' related concepts, 5 sub-graphs explain between teacher and me, with one or more of the other elements.

5. DISCUSSION

In this study, it was found that most students learn through 'me' initiated behaviors (e.g., "me-listening" or "me-ask") or using learning tools (e.g., "me-notes" or "me-textbooks"). Learning through active interaction with teachers or peers was not observed. This indicates that early level college students focus on individual learning rather than learning through interaction with others. It may also reflect lecture oriented college courses. However, readers should consider that too many concepts given to students may have resulted in difficulty identifying patterns between students and others.

From this initial research, we found both possibilities and challenges when using concept maps in this capacity in an educational context. First, when used with data mining techniques, concept maps can be useful to interpret large sets of concept maps. As can be seen in our data analysis,

Relevant concept	Frequency (Support count)	Frequent sub-concept map patterns
Me(Students)	49	Notes - Me
	34	Listening - Me
Teachers	44	Lectures - Teachers
Me(Students)	31	Teachers - Me
	30	Textbooks - Me
	29	Flashcards - Me
	29	Lectures - Me
	29	Reading - Me
	28	Study - Me
	27	Attendance - Me
	24	Annotation - Me
	21	Pay attention - Me
	22	Flashcard - Me -Notes
Teachers	22	Notes - Teachers
Etc.	21	Teacher - Lecture - Me
Me(Students)	16	Questions - Me
	16	Review - Me
	16	Discussions - Me
	15	Assignments - Me
	14	Study - Me
	14	Memorization - Me
	13	Classmates - Me
	12	Asking - Me
	12	Class - Me
	18	Lecture - Me - Notes
	16	Listening - Me - Reading
	15	Reading - Me - Notes
	15	Attendance - Me - Notes
	14	Notes - Me - Teacher
	14	Annotation - Me - Notes
	13	Study - Me - Notes
	12	Listening - Me - Notes
	11	Textbooks - Me - Flashcard
	11	Annotation - Me - Textbook
	11	Reading - Me - Teacher
	11	Lecture - Me - Textbooks
	11	Attendance - Me - Textbooks
	11	Listening - Me - Pay attention
Teachers	15	Discussion - Teacher
	14	Questions - Teacher
	13	Teach - Teachers
	12	Listening - Teacher
Etc.	18	Teacher - Notes - Me
	13	Lectures - Teacher - Me
	11	Teacher - Questions - Me
	12	Teacher - Lecture - Me - Note

Table 1: Frequent sub concept map patterns

when using different thresholds, we were able to find certain patterns of students' learning in a college classroom. Although most of the patterns were simple such as "me-listening" or "me-annotations", this does not mean the concept mapping was an ineffective method to capture sophisticated pictures of students' learning. The simple patterns can be attributable to the large list of 112 predesigned concepts. Because students had too many choices, students' concept maps became very diverse. However, through this initial phase of study, we were able to find the most frequently referenced concepts for future study. With limiting the number of concepts, e.g., 30 concepts, we may be able to find more detailed pictures of college students' learning. In our future study, we will compare and contrast patterns of concept maps to students' self-report motivation or metacognition. For example, we can divide students into several groups based on their self-report questionnaires, and then compare how the patterns of concept maps are similar or different between groups. We are also planning to use students' final grades as a way to validate the patterns of learning. The study will be appealing to those who are interested in concept maps as an alternative tool for research as well as data mining in education.

6. REFERENCES

[1] R. Agarwal and R. Srikant. Fast algorithms for Mining association rules. In *Proc. of International Conference on Very Large Databases(VLDB)*, 1994.

[2] A. O. Alves, F. C. Pereira, and A. Cardoso. Automatic reading and learning from text. In *Proc. of International Symposium on Artificial Intelligence*, 2002.

[3] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: Generalizing association rules to correlations. In *Proc. of ACM SIGMOD International Conference Management of Data*, 1997.

[4] N.-S. Chen, P. Kinshuk, C.-W. Wei, and H.-J. Chen. Mining e-learning domain concept map from academic articles. *Computers & Education*, 50:694-698, 2008.

[5] R. B. Clariana and R. Koul. A computer-based approach for translating text into concept map-like representations. In *Proc. of International Conference on Concep Mapping*, 2004.

[6] D. Cook and L. Holder. *Mining graph data*. Wiley Publication, 2007.

[7] K. en Chang, Y.-T. Sung, and I.-D. Chen. The effect of concept mapping to enhance text comprehension and summarization. *The Journal of Experimental Education*, 71(1), 2005.

[8] J. Huan, J. Prins, W. Wang, and J. Yang. SPIN: Mining maximal frequent subgraphs from graph databases. In *Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004.

[9] G. Karypis and M. Kuramochi. Frequent subgraph discovery. In *Proc. of International Conference on Data Mining*, 2001.

[10] I. M. Kinchin. Using concept maps to reveal understanding: A Two-tier analysis. *School Science Review*, 81:41-46, 2000.

[11] J. R. McClure, B. Sonak, and H. K. Suen. Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *The Journal of Research in Science Teaching*, 36:475-492, 1999.

[12] J. D. Novak and A. J. Canas. The theory underlying concept maps and how to construct and use them. Technical report IHMC Cmap Tools 2006-01 Rev 01-2008.

[13] J. D. Novak and D. B. Gowin. *Learning how to learn*. Cambridge. UK: Cambridge University Press, 1984.

[14] W. G. Perry. *Forms of intellectual and ethical development in the college years: A scheme*. New York:Holt, Rinehart & Winston, 1970.

[15] P. R. Pintrich, D. A. Smith, T. Garcia, and W. J. McKeachie. Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53:801-813, 1993.

[16] R. Shen, R. Richardson, and E. Fox. Concept maps as visual interfaces to digial libraries: summarization, collaboration, and automatic generation. In *http://vw.indiana.edu/ivira03/shen-et-al.pdf*, 2004.

[17] P. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addision Wesley, 2005.

[18] E. M. Taricani and R. B. Clariana. A technique for automatically scoring open-ended concept maps . *Educational Technology Research & Development*, 54:65-82, 2006.

[19] A. Valerio and D. Leake. Jump-starting concept map construction with knowledge extracted from documents. In *Proc. of International Conference on Concept Mapping*, 2006.

[20] C. E. Weinstein, A. C. Schulte, and D. R. Palmer. *LASSI: Learning and study strategies inventory*. H & H Publishing Co., 1987.

[21] P. Winne and N. Perry. *Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.) Handbook of self-regulation (532-566)* . CA: Academic Press, 2000.

[22] X. Yan and J. Han. gSpan: Graph-based substructure pattern mining. In *IEEE International Conference on Data Mining*, 2001.

[23] Y. Yin, J. Vanides, M. A. Ruiz-Primo, C. C. Ayala, and R. J. Shavelson. Comparision of two concept-mapping techniques: implications for scoring, interpretation, and use. *Journal of Research in Science Teaching*, 2:166-184, 2005.

[24] B. Zimmerman. Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45:166-183, 2008.