



International Journal of Intellectual Advancements and Research in Engineering Computations

Effective online discussion data for teacher's reflective thinking

¹Mrs. K.E. Eswari, M.C.A., M.Phil.,M.E., Mr. B.Thayanandhan²

¹Assistant Professor, ²Final MCA Students

Department of Mechanical Engineering, Nandha Engineering College, Erode - 52.
TamilNadu. India.

ABSTRACT

In this paper analysis automatic coding method by integrating the inductive content analysis and text classification techniques. In existing model acquire the reflective thinking categories by conducting an inductive content analysis and base our text classification algorithm on the categories, so we augment the manual method of coding. We apply the trained classification model to a large-scale and unexplored online discussion data set, so we can have a comprehensive understanding of teachers' reflection. This paper also provides six types of visualizations of the text classification results: the visualization of teachers' reflection level and the visualization of teachers' reflection evolution. By using the categories gained from inductive content analysis to create a radar map, we visually represent teachers' reflection level after obtain the results of text classification future classification model.

Keywords: Data Mining, Teacher Reflection Model, TF-IDF Classification, Visualization Learning

INTRODUCTION

Data mining is that the method of extracting patterns from knowledge. Data mining is seen as a progressively vital tool by fashionable business to transform knowledge into an informational advantage. It's presently utilized in a wide range of identification practices, like selling, surveillance, fraud detection, and scientific discovery.

The connected terms knowledge dredging, knowledge fishing and knowledge snooping confer with the utilization of knowledge mining techniques to sample parts of the larger population data set that are (or might be) too tiny for reliable statistical inferences to be created regarding the validity of any patterns discovered (see additionally data-snooping bias). These techniques will but, be utilized in the creation of latest hypothesizes to check against the larger knowledge populations.

Data mining is that the method of applying these strategies to knowledge with the intention of

uncovering hidden patterns. It's been used for several years by businesses, scientists and governments to sift through volumes of information like airline passenger trip records, census knowledge and food market scanner knowledge to supply research reports. (Note, however, that coverage isn't perpetually thought-about to be data processing.)

In this study, we encoded and visualized the large-scale unstructured text data in teachers' online collaborative learning activities in order to understanding teachers' reflection levels and evolution, integrating both the qualitative content analysis method and educational data mining techniques. The qualitative content analysis method revealed that teachers' reflective thinking included Technical-Description, Technical-Analysis, Technical-Critique, Personalistic Description, Personalistic-Analysis, and Personalistic-Critique. Based on the results of inductive content analysis, we implemented a

single-label text classification algorithm to classify the sample data. Then we applied the trained classification model on a large-scale and unexplored online discussion text data set. After the online discussion text data being classified, two types of visualizations of the results were provided. By using the categories gained from inductive content analysis to create a radar map, teachers' reflection level had been represented. In addition, by defining the variables of the coding scheme as the nodes and the co-occurrences of nodes within a discussion post as the connections, a cumulative adjacency matrix was created to characterize the evolution of teachers' reflective thinking [1-5].

RELATED WORKS

Weize Kong and James Allan [1] evaluated Faceted Web Search systems by their utility in assisting users to clarify search intent and subtopic information. The authors described how to build reusable test collections for such tasks, and propose an evaluation method that considers both gain and cost for users. Faceted search enables users to navigate a multi-faceted information space by combining text search with drill-down options in each facet. For example, when searching "computer monitor" in an e-commerce site, users can select brands and monitor types from the provided facets: Samsung, Dell, Acer... g and f LET-Lit, LCD, OLEDg. This technique has been used successfully for many vertical applications, including e-commerce and digital libraries.

Krisztian Balog et al., [2] consider the task of entity search and examine to which extent state-of-art information retrieval (IR) and semantic web (SW) technologies are capable of answering information needs that focus on entities. We also explore the potential of combining IR with SW technologies to improve the end-to-end performance on a specific entity search task. We arrive at and motivate a proposal to combine text-based entity models with semantic information from the Linked Open Data cloud. The problem of entity search has been and is being looked at by both the Information Retrieval (IR) and Semantic Web (SW) communities and is, in fact, ranked high on the research agendas of the two communities.

The entity search task comes in several flavors. One is known as entity ranking (given a query and target category, return a ranked list of relevant entities), another is list completion (given a query and example entities, return similar entities), and a third is related entity finding (given a source entity, a relation and a target type, identify target entities that enjoy the specified relation with the source entity and that satisfy the target type constraint).

Chengkai Li, Ning Yan et al [3] focused on automatic and dynamic faceted interfaces. The facets could not be pre-computed due to the query-dependent nature of the system. In applications where faceted interfaces are deployed for relational tuples or schema-available objects, the tuples/objects are captured by prescribed schemata with clearly defined dimensions (attributes), therefore a query-independent static faceted interface (either manually or automatically generated) may suffice. By contrast, the articles in Wikipedia are lacking such predetermined dimensions that could fit all possible dynamic query results. Therefore efforts on static facets would be futile.

Wisam Dakka et al., [4] presented a set of techniques for automatically identifying terms that are useful for building faceted hierarchies. The techniques build on the idea that external resources, when queried with the appropriate terms, provide useful context that is valuable for locating the facets that appear in a database of text documents. It is demonstrated the usefulness of Wikipedia, WordNet, and Google as external resources. Experimental results, validated by an extensive study using human subjects, indicate that our techniques generate facets of high quality that can improve the browsing experience for users. If efficiency is not a major concern, can incorporate multiple such resources in this framework, for a variety of topics, and use all of them, irrespectively of the topics that appear in the underlying collection. The distributional analysis step of our technique automatically identifies which concepts are important for the underlying database and generates the appropriate facet terms.

Amaç Herdagdelen et al., [5] presents a novel approach to query reformulation which combines syntactic and semantic information by means of

generalized Levenshtein distance algorithms where the substitution operation costs are based on probabilistic term rewrite functions. We investigate unsupervised, compact and efficient models, and provide empirical evidence of their effectiveness. Further it explores a generative model of query reformulation and supervised combination methods providing improved performance at variable computational costs. Among other desirable properties, our similarity measures incorporate information-theoretic interpretations of taxonomic relations such as specification and generalization.

X. Xue and W. B. Croft [6] propose a novel framework where the original query is transformed into a distribution of reformulated queries. A reformulated query is generated by applying different operations including adding or replacing query words, detecting phrase structures, and so on. Since the reformulated query that involves a particular choice of words and phrases is explicitly modeled, this framework captures dependencies between those query components. On the other hand, this framework naturally combines query segmentation, query substitution and other possible reformulation operations, where all these operations are considered as methods for generating reformulated queries. In other words, a reformulated query is the output of applying single or multiple reformulation operations.

Mariana Damova, Ivan Koychev [7] presents a survey of recent extractive query-based summarization techniques. We explore approaches for single document and multi-document summarization. Knowledge-based and machine learning methods for choosing the most relevant sentences from documents with respect to a given query are considered. Further, expose tailored summarization techniques for particular domains like medical texts. The most recent developments in the field are presented with opinion summarization of blog entries. This survey is motivated by the idea of making e-books more intelligent, in particular enabling them to “answer” users’ queries. To find the needed information in books users usually do not want to spend a long time searching, browsing or skimming them. They will be happy to have a “guru” nearby that can provide them with the right answer almost

simultaneously. For this purpose to had a close look at the area of automated text summarization. Recently, with the increasing of information available online, those approaches have been developed very extensively. In the realm of automatic summarization different kinds of summarization have been attempted. Along with the study distinguish between the following types of summaries according to specific criteria.

METHODOLOGY

The purposes of this research are 1) to explore the categories of teachers’ reflective thinking and realize the automatic classification of the online discussion data by integrating the inductive content analysis and educational data mining techniques; and 2) to analyze teachers’ reflective thinking in the online teacher professional development program, including teachers’ reflection levels and evolution. Teachers’ online discussion data has been collected and analyzed for understanding their reflective thinking mainly because:

- The peer coaching model has long been used in professional development programs to enhance teachers’ teaching practices and students’ learning out-comes. In reciprocal peer coaching, teachers
- Share the coaching role, discuss with each other, and elicit the reflection [10]. Teachers’ online discussion data embodies their reflective thinking.
- Based on the understanding of teachers’ reflective thinking, teacher training managers and educational researchers can design more appropriate online learning activities and provide proper services and interventions to support teachers’ online reflection.
- Content analysis is commonly used to analyze transcript of online discussion for educational purposes. In this study, the coding frameworks, which focused on the main content and level of teachers’ reflection, had been obtained from inductive content analysis. In addition, we had developed a Data collector, the Hawk, to collect posts on OPDP. Therefore, we chose to analyze teachers’ posts for understanding their reflective thinking

Stem Word Document

In this module, enter the given word and stem word using text box control and click save button stem word saved into the table. The details are saved in 'Stem word' file. The stem word details view on view controls.

Stop Word Document

In this module, enter the stop word using text box control and click save button stop word saved into the table. The details are saved in 'Stop word' file. The stop word details view on view controls.

Add Synonym Word Document

In this module, enter the given word and synonym word using text box control and click save button synonym word saved into the table. The details are saved in 'synonym word' file. The synonym word details view on view controls.

Add Test Dataset

The test dataset in this context are different which technical terms are, and hence they are fixed-terms. In other words, the keywords are natural language like (i.e., not specific technical terms), so they are diverse and have various synonyms. For example, in the sentence "Trust influences communication between organizations". To identify the Trust and communication is an entity of success factor and influences as an influencing relationship.

The relevant and irrelevant sentences are annotated into positive class and negative class respectively. The annotation is manually performed by a domain expert based on the given related keywords and those keywords are used as a guideline for the annotation. In this module, user can submit the certain keywords which are stored into the Key Words file.

Subset of Six levels preparation

✓ Phase 1: A data collection tool to collect teacher-generated online discussion data and obtained 21,388 posts by using the data collection tool over a period of 6 months.

- ✓ Phase 2: A random sample (2,000 posts) were drawn from the online discussion data set and used for inductive content analysis.
- ✓ Phase 3: Three experts who had experience with qualitative research and familiarity with pre-existing coding schemes collaborated on the inductive content analysis process. Three experts conducted an inductive content analysis on the random sample of online discussion data set.
- ✓ After phase 3, we obtained the categories used for text classification.
- ✓ Phase 4: A single-label Naïve Bayes Classification algorithm had been implemented to classify the labeled data. Then, we evaluated the performance of the single-label Naïve Bayes classification algorithm by comparing it with other commonly used text classification algorithm.
- ✓ Phase 5: Implemented a large-scale text classification based on the trained classification model. All the online discussion data has been automatically coded.
- ✓ Phase 6: Visually represented teachers' reflection levels after we obtained the results of large-scale text classification. Additionally, we created a cumulative adjacency matrix to characterize the evolution of teachers' reflective thinking. Then, the results of text classification and visualization were sent back to the data storage.

Online Collaborative Learning Approach

The approach of online collaborative learning provides Internet-based professional development opportunities, including individual reflection, sharing resource, work- shops, online interactions with colleagues, and mentors

- In the first stage, every teacher reflected the discussion topic individually. After the chief teacher posted the activity plan and discussion topic onto the OPDP, every teacher downloaded and read the plan and topic



Fig 3.1 Teacher Communication

- In the second stage, all teachers discussed the topic collectively. Every teacher expressed their own views, and exchanged views with others. This is a process of externalization. Teachers expressed their ideas or thoughts about a specific topic via posts.
- In the final stage, every teacher submitted a document, which recorded his/her learning experience in the online collaborative learning activity. This is a process of internalization. Teachers recorded the knowledge learned from communication with mentors and colleagues.

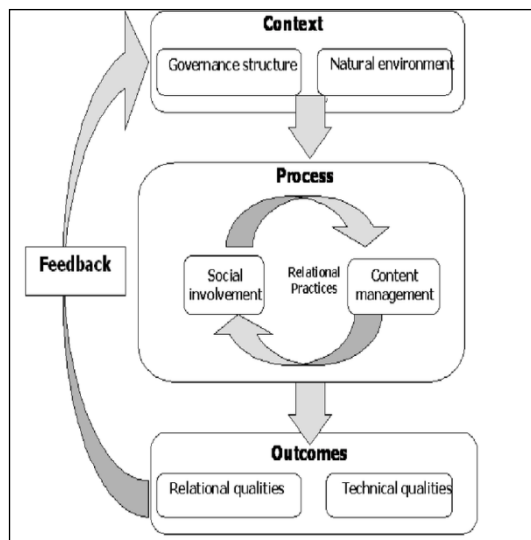


Fig 3.2 Feedback Communication

- Each three-stage online collaborative learning activity usually lasted for one month. After three stages were completed, the chief teacher and the individual teachers summarized the results.
- In the three-stage online collaborative learning activity, teachers' reflection thinking was expressed in two main ways: the online discussion posts and the reflection documents. In this study, we collected teachers' online discussion posts to analyze their reflective thinking.

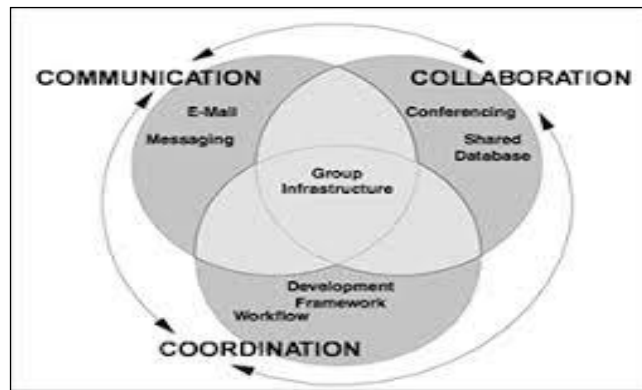


Fig 3.3 Collaborative Detection

Term Frequency – Inverse Document Frequency (TF-IDF)

The TF measures how frequently a particular term occurs in a document. It is calculated by the number of times a word appears in a document divided by the total number of words in that document. It is computed as $TF(\text{the}) = (\text{Number of times term the 'the' appears in a document}) / (\text{Total number of terms in the document})$. The IDF measures the importance of a term. It is calculated by the number of documents in the text database divided by the number of documents where a specific term appears. While computing TF, all the terms are considered equally important. That

means, TF counts the term frequency for normal words like “is”, “a”, “what”, etc. Thus we need to know the frequent terms while scaling up the rare ones, by computing the following: $IDF(\text{the}) = \log_e(\text{Total number of documents} / \text{Number of documents with term 'the' in it})$.

For example, consider a document containing 1000 words, wherein the word give appears 50 times. The TF for give is then $(50 / 1000) = 0.05$. Now, assume that, 10 million documents and the word give appears in 1000 of these. Then, the IDF is calculated as $\log_e(10,000,000 / 1,000) = 4$. The TF-IDF weight is the product.

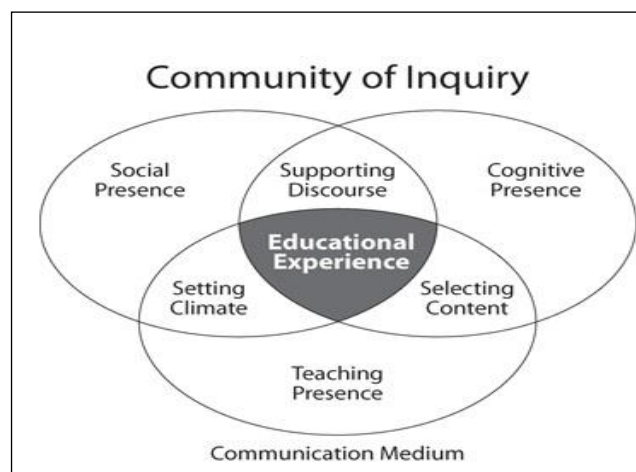


Fig 3.4 Community of Inquiry

CONCLUSION

An inductive content analysis on samples taken from 17,624 posts was implemented and the categories of teachers' reflective thinking were obtained. Based on the results of inductive content analysis, we implemented a single-label text classification algorithm to classify the sample data. Then, we applied the trained classification model on a large-scale and unexplored online discussion text data set and two types of visualizations of the results were provided. By using the categories gained from inductive content analysis to create a radar map, teachers' reflection level was represented.

In addition, a cumulative adjacency matrix was created to characterize the evolution of teachers' reflective thinking. This study could partly explain how teachers reflected in online professional learning environments and brought awareness to educational policy makers, teacher training managers, and education researchers. The

considerably better results are found for the planned technique compared to existing strategies, regardless of the classifiers used. All the results according in this paper demonstrate the practicability and effectiveness of the planned technique. It's capable of distinctive co-regulated clusters of genes whose average expression is strongly related to the sample classes. The known gene clusters could contribute to revealing underlying category structures, providing a useful gizmo for the explorative analysis of biological information.

In future, the here fast algorithm enforced in information set solely, additional planned to implement in numerical information set. During this application, enforced cancer dataset solely, in future planned to real time dataset like diabetics, pressure so on. Additional a lot of, planned to compare with algorithm like naïve bayes, K-Nearest neighbor so on.

REFERENCES

- [1]. W. Kong and J. Allan, "Extending faceted search to the general web," in Proc. ACM Int. Conf. Inf. Knowl. Manage., 2014, 839–848.
- [2]. K. Balog, E. Meij, and M. de Rijke, "Entity search: Building bridges between two worlds," in Proc. 3rd Int. Semantic Search Workshop, 9, 2010, 1–9:5.
- [3]. C. Li, N. Yan, S. B. Roy, L. Lisham, and G. Das, "Facetedpedia: Dynamic generation of query-dependent faceted interfaces for wikipedia," in Proc. 19th Int. Conf. World Wide Web, 2010, 651–660.
- [4]. W. Dakka and P. G. Ipeirotis, "Automatic extraction of useful facet hierarchies from text databases," in Proc. IEEE 24th Int. Conf. Data Eng., 2008, 466–475.
- [5]. A. Herdagdelen, M. Ciaramita, D. Mahler, M. Holmqvist, K. Hall, S. Riezler, and E. Alfonseca, "Generalized syntactic and semantic models of query reformulation," in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. retrieval, 2010, 283–290.
- [6]. X. Xue and W. B. Croft, "Modeling reformulation using query distributions," ACM Trans. Inf. Syst., 31(2), 2003, 1–6, 34.
- [7]. M. Damova and I. Koychev, "Query-based summarization: A survey," in Proc. S3T, 2010, 142–146.