

Learning to Identify Students' Relevant and Irrelevant Questions in a Micro-Blogging Supported Classroom

Suleyman Cetintas, Luo Si Sugato Chakravarty Hans Aagard, Kyle Bowen
Department of Computer Department of Consumer Rosen Center for Advanced
Sciences Sciences and Retailing Computing
Purdue University, West Lafayette, IN, 47907, USA
{scefinta,lsi}@cs.purdue.edu sugato@purdue.edu {hans,kbowen}@purdue.edu

Abstract. This paper proposes a novel application of text categorization for two types questions asked in a micro-blogging supported classroom, namely relevant and irrelevant questions. Empirical results and analysis show that utilizing the correlation between questions and available lecture materials in a lecture along with personalization and question text leads to significantly higher categorization accuracy than i) using personalization along with question text and ii) using question text alone.

1 Introduction

Micro-blogging, a Web 2.0 technology, is a type of blogging that lets the users post short text messages to their community in real time. Recently, micro-blogging tools have been used in classroom environments as a communication tool between students with the instructor [2]. An important issue with large micro-blogging supported classrooms is that the number of questions/comments an instructor receives from the students can be many more than what s/he can answer in a limited time.

To the best of our knowledge, there is no prior research on the categorization of relevant and irrelevant micro-blogging messages or questions in classroom environments. Prior work on teacher agents has a question ranking capability that utilizes the questions text as well as a personalized approach to differentiate between the relevant and irrelevant questions [5]; but ignore the educational materials that are available in most classrooms.

This paper proposes a text categorization approach that can automatically identify relevant and irrelevant questions asked in a lecture by utilizing multiple types of evidence including question text, personalization, correlation between questions and lecture materials. We show that i) utilizing personalization along with question text is more effective than using question text alone, ii) utilizing the correlation between question and available lecture materials improve the categorization accuracy and iii) tf-idf weighting scheme is more effective than okapi while estimating the correlation between questions and available lecture materials.

2 Data

Data collected from a micro-blogging supported personal finance class (a 300 level u.g. course) during Fall 2009 has been used in this work. The study was conducted in a large classroom with 243 students during 24 lectures (each of them 50 minutes

Table 1. Results of the SVM_TermsOnly, SVM_TermsPers and SVM_TermsPersLMSim classifiers in comparison to each other for two main configurations (while correlations between questions and available lecture materials are calculated) i) with tf-idf and ii) okapi is used as weighting schemes. The performance is evaluated with the F_1 measure [1] and “ F_1 (precision, recall)” triplets are reported for each configuration of each classifier.

Methods	Tf-Idf	Okapi
SVM_TermsOnly	0.7252 (0.8797, 0.6169)	
SVM_TermsPers	0.7830 (0.8675, 0.7135)	
SVM_TermsPersLMSim	0.8627 (0.9070, 0.8225)	0.7948 (0.8514, 0.7452)

long). Data from first 4 lectures are used for training and the remaining 20 lectures are used for testing. Each lecture has an average of 26.9 relevant questions with a standard deviation of 9.4 and 10.8 irrelevant questions with a standard deviation of 8.5 (i.e. totally 645 relevant and 260 irrelevant questions). We employ two human annotators (the first author and an expert in finance) and ask them to annotate each question as either being relevant or irrelevant. The annotators reach a Kappa of 0.868 on 162 questions (i.e. questions of first 4 lectures) and therefore the rest of the data was annotated by the first annotator only. Every lecture has 1 publicly available presentation file relevant to the lecture (which are used as the available relevant lecture materials) and the course has a syllabus file that discusses about course policy, exams, projects, etc. (which is used as the available irrelevant lecture material).

3 Techniques: Support Vector Machine and Cosine Similarity

Micro-blogging questions/messages are in textual format; therefore we use the widely used Support Vector Machines (with a linear kernel) as our text classifier [4]. The categorization threshold of each SVM classifier is learned by 2-fold cross validation in the training phase (i.e. 2 of the 4 training lectures for each fold).

Cosine similarity is a measure of similarity between two vectors by calculating the cosine of the angle between them which is commonly used in text mining to compare text documents. In this work, the similarity scores between questions and lecture materials are calculated as a measure of the correlation by the common Cosine measure [1]. We use and compare the two common weighting schemes: tf-idf [1] that uses term frequency and inverse document frequency (i.e. favoring discriminative terms that only reside in a small number of documents) and okapi [3] that additionally considers document sizes by favoring shorter but relevant documents.

4 Performances of Several Modeling Approaches

SVM_TermsOnly & SVM_TermsPers: Using individual features (i.e. terms) of questions along with personalization to select the best questions to respond to has been shown to be a useful approach in a recent prior work [5]. In this work, we use two features for personalization: i) percentage of relevant questions asked by a student and ii) percentage of irrelevant questions asked by a student. An SVM classifier that only uses the terms of questions is used along with another SVM classifier that uses personalization along with the terms. The two baseline classifiers will be referred as SVM_TermsOnly and SVM_TermsPers respectively.

SVM_TermsPersLMSim: In a lecture, it is intuitive that most relevant questions asked in a class will be related with the lecture being covered in class. This modeling approach makes use of this fact and adds 3 new features about the correlation between a question and available lecture materials (that may be relevant or irrelevant) to the set of baseline (i.e. terms and personalization) features. Particularly the added

features are the cosine similarity score i) between a question and available relevant lecture material(s) of the current lecture, ii) between a question and all available relevant lecture materials of that course (i.e. in this work, sum of the similarity scores of top 3 most similar relevant materials are used), iii) between a question and all available non-relevant lecture materials (i.e., in this work, the similarity score with the only irrelevant material is used: if there are more irrelevant materials, the approach in (ii) can be used).

It can be seen in Table 1 that SVM_TermsPers classifier significantly (with p-value of less than 0.01, for paired t-test) outperforms the SVM_TermsOnly classifier. This shows that utilizing personalization along with terms is a better approach than using only terms of questions and this is consistent with prior research [5]. SVM_TermsPersLMSim classifier is also shown to significantly (i.e. with p-value much less than 0.01) outperform SVM_TermsPers and SVM_TermsOnly classifiers. Utilizing the correlations among questions and available lecture materials along with terms and personalization is a better approach than using only personalization and terms of questions. To assess the similarity between questions and available lecture materials, two common weighting schemes are used with the cosine similarity measure. It can be seen that tf-idf weighting scheme significantly outperforms (with p-value less than 0.01) okapi weighting scheme.

5 Conclusions

This paper proposes a novel application of text categorization to identify relevant and irrelevant micro-blogging questions asked in a classroom. Several modeling approaches and weighting configurations are studied for this application. Empirical results show that utilizing the correlation among questions and available lecture materials along with personalization and question text significantly outperforms i) using personalization and question text and ii) using question text only. Furthermore, it is found to be significantly more effective to use tf-idf weighting scheme rather than okapi while calculating the correlations among questions and available lecture materials.

Acknowledgements: This research was partially supported by the NSF grants IIS-0749462, IIS-0746830. Any opinions, findings, conclusions expressed in this paper are the authors', and do not necessarily reflect those of the sponsor.

References

1. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press Series/Addison Wesley (1999)
2. Grossec, G., & Holotescu, C.: Can we use Twitter for educational activities? In Proceedings of the 4th Int. Scientific Conf., eLearning and Software for Education (2008)
3. Robertson, S. Walker S. Beaulieu, S., Gull, A., Lau, M.: Okapi at TREC. In Proceedings of 1st TREC Conference (1992)
4. Sebastiani, F.: Machine learning in automated text categorization. In ACM Computing Surveys, 34(1):1–47 (2002)
5. Soh, L., Khandaker, N. and Jiang, H. I-MINDS: A Multiagent System for Intelligent Computer-Supported Collaborative Learning and Classroom Management. Int. J. Artif. Intell. Ed. 18, 2, 119-151 (2008)