Evaluation on Text Categorization for Mathematics Application Questions

Liang-Chih Yu, Hsiao-Liang Hu and Wei-Hua Lin
Department of Information Management, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li, Taiwan 32003, R.O.C.
E-mail: lcyu@saturn.yzu.edu.tw Tel: +886-3-4638800

***Abstract*— In learning environments, developing intelligent systems that can properly respond learners' emotions is a critial issue for improving learning outcome. For example, systems should consider to replace the current question with an easier one when detecting negative emotions expressed by learners. Conversely, systems can try to retrieve a more challenging question when learners have contempt emotion or feel bored. This paper proposes the use of text categorization to automatically classify mathematics application questions into different difficulty levels. Applications can then benefit from such classification results to develop retrieval systems for proposing questions based on learners' emotion states. Experimental results show that the machine learning algorithm C4.5 achieved the highest accuracy 78.53% in a binary classification task.**

I. INTRODUCTION

In learning environments, learners may express different emotions due to various questions with different difficulty levels. For example, learners may express delight when they satisfy their performance or the questions are easy. Nevertheless, if the questions are too simple, then students may express contempt. Conversely, students may express frustration when they worry about their performance or the questions are hard [1], [2]. Previous studies have shown that such emotions may affect learning outcomes [3], [4], [5]. For instance, Rodrigo et al suggested that boredom may have a negative impact on student achievement, while confusion may have both positive and negative effects [4]. Therefore, it is worth to develop e-learning systems that can properly respond learners' emotions. For example, systems should consider to replace the current question with an easier one when detecting negative emotions expressed by learners. Conversely, systems can try to retrieve a more challenging question when learners have contempt emotion or feel bored. To achieve the goal, a critical step is to automatically identify difficulty levels of questions.

In this paper, we propose the use of text categorization to automatically classify mathematics application questions into different difficulty levels. Applications can then benefit from such classification results to develop retrieval systems for proposing questions based on learners' emotion states. We use mathematics application questions as the target domain because application questions play an important role in mathematics learning. Such questions usually have two major learning goals: propose and solve equations. To improve students' abilities to propose and solve equations, intelligent e-learning systems should be more adaptive to select adequate application questions according to the error type caused by students. For example, systems should select propose-equation questions when students have trouble in proposing equations in solving application questions. Similarly, systems should select more difficult solve-equation questions when students are good at proposing equations but have trouble in solving equations. To accomplish this goal, automatic classification of application questions into propose-equation and solve-equation types is an essential step. Information retrieval (IR) techniques can then be used to realize adaptive learning by commending different types of application questions according to students' error types.

Classification of mathematics application questions can be casted to a text categorization task, where each question is a text span (a sentence or passage). Different classification methods [6], [7], [8] and features [9], [10], [11], [12], [13], [14] can then be used to classify each text span into a set of predefined categories. A Bag-of-words is the most common feature which is usually adopted as the baseline feature for classifier training. Due to the constraint of independence assumption between words imposed on the Bag-of-words approaches, *n*-grams have been investigated to capture sequential relations between words for improving classification performance [9], [10], [11], [12]. Recent studies have indicated that *n*-grams are useful features to capture local dependencies of words but may have difficulties in capturing long-distance dependencies especially for higher-order *n*-grams [13]. Therefore, association language patterns which represent meaningful combinations of words have been induced using supervised and unsupervised methods [13], [14]. Such language patterns including intra-sentential and inter-sentential patterns can capture both local and long-distance dependencies between words, and thus can improve classification performance.

II. MATHEMATICS APPLICATION QUESTIONS

The mathematics application questions were collected from a junior high school in Taiwan. These questions are selected by an expert from the chapter of One-Variable Linear Equation, and then divided into two types: propose-equation and solve-equation. Table I presents example questions for the two types of qiestions. Generally, solve-equation questions are more difficult than propose-equation questions.

TABLE I
TWO TYPES OF MATHEMATICS APPLICATION QUESTIONS

Type	Example question
Propose equation	What is the side of a square whose circumference is x cm?
Solve equation	There are 40 questions in a quiz and the full marks are 100 points. Suppose that each fill-in-the-blank question is 2 points and each choice question is 4 points. How many fill-in-the-blank and choice questions are in the quiz?

TABLE II
CLASSIFICATION RESULTS

	NB	C4.5	SVM
Propose equation	0.777	0.830	0.851
Solve equation	0.663	0.735	0.651
Average	**0.732**	**0.785**	**0.757**

III. EXPERIMENTAL RESULTS

This section evaluates different methods for classification of mathematics application questions. The classification is binary, i.e., to classify each test question into propose-equation or solve-equation question. A total of 177 mathematics application questions were collected where 94 questions were propose-equation questions and the remaining 83 were solve-equation questions. The questions were first segmented into word sequences using the CKIP segmentation system [15], and then used to build three classifiers: NB, C4.5, and SVM using Weka Package [16]. In performance evaluation, we used a 10-fold cross validation, and the evaluation metric was accuracy defined by the number of correctly classified questions divided by the total number of questions. Experimental results (Table II) show that the respective accuracies of NB, C4.5, and SVM were 72.3%, 78.5% and 75.7%. This finding indicates that the decision tree method C4.5 yielded highest classification performance.

IV. CONCLUSIONS

This paper has presented use of text categorization to automatically classify mathematics application questions into different difficulty levels. Experimental results show that the machine learning algorithm C4.5 achieved the highest accuracy 78.53% in a binary classification task. Future work will focus on investigating equation-specific features and developing more effective machine learning algorithms to boost classification performance.

ACKNOWLEDGMENT

This work was partially supported by the National Science Council, Taiwan, R.O.C., under Grant No. NSC 102-2221-E-155-029-MY3.

REFERENCES

[1] L. C. Yu, S. F. Liang, W. H. Lin, K. R. Lai, and B. J. Liu, "Analysis of Students' Emotion from a Text Corpus," In *Work-In-Progress Proceedings of the 19th International Conference on Computers in Education (ICCE-11)*, pp. 1-3, 2011.

[2] L. C. Yu, S. F. Liang, W. H. Lin, "Analysis of Students' Emotion from a Text Corpus," In *Proceedings of the 20th International Conference on Computers in Education (ICCE-12)*, pp. 51-55, 2012.

[3] S. D'Mello, A. Graesser, and R. W, Picard, "Toward an Affect-Sensitive AutoTutor," *IEEE Intelligent Systems*, vol. 22, no. 4, pp. 53-61, 2007.

[4] M. M. T. Rodrigo, R. S. J. D. Baker, and J. Q. Nabos, "The Relationships between Sequences of Affective States and Learner Achievement," In *Proceedings of the 18th International Conference on Computers in Education (ICCE-10)*, pp. 56-60, 2010.

[5] M. H. M. Yusoff and B. du Boulay, "A Tutoring System Using an Emotion-Focused Strategy to Support Learners," In *Proceedings of the 18th International Conference on Computers in Education (ICCE-10)*, pp. 89-91, 2010.

[6] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, vol. 34, no. 1, 2002, pp. 1-47.

[7] A. Khoo, Y. Marom and D. Albrecht, "Experiments with Sentence Classification," In *Proceedings of Australasian Language Technology Workshop*, 2006, pp. 18-25.

[8] M. Naughton, N. Stokes, and J. Carthy, "Investigating Statistical Techniques for Sentence-Level Event Classification," In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-08)*, Manchester, United Kingdom, 2008, pp. 617-624.

[9] R. Chitturi and J.H.L. Hansen, "Dialect Classification for online podcasts fusing Acoustic and Language based Structural and Semantic Information," In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies (ACL-08)*, Columbus, OH, 2008, pp. 21-24.

[10] S. Li and C. Zong, "Multi-domain Sentiment Classification," In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies (ACL-08)*, Columbus, OH, 2008, pp. 257-260.

[11] L.C. Yu, C.H. Wu, R.Y. Chang, C.H. Liu and E.H. Hovy, "Annotation and Verification of Sense Pools in OntoNotes," *Information Processing & Management*, vol. 46, no. 4, 2010, pp. 436-447.

[12] L.C. Yu and W.N. Chien, "Independent Component Analysis for Near-synonym Choice," *Decision Support Systems*, vol. 55, no. 1, pp. 146-155, 2013.

[13] L.C. Yu, C.L. Chan, C.C. Lin and I.C. Lin, "Mining Association Language Patterns Using a Distributional Semantic Model for Negative Life Event Classification," *Journal of Biomedical Informatics*, vol. 44, no. 4, 2011, pp. 509-518.

[14] J.L. Wu, L.C. Yu and P.C. Chang, "Detecting Causality from Online Psychiatric Texts Using Inter-Sentential Language Patterns," *BMC Medical Informatics and Decision Making*, vol. 12, 2012, article 72.

[15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10-18, 2009.

[16] W.Y. Ma and K.J. Chen, "A Bottom-up Merging Algorithm for Chinese Unknown Word Extraction," In *Proceedings of the ACL workshop on Chinese Language Processing*, 2003.