

A Novel Approach for Assisting Teachers in Assessment of Student Reading Ability in web-based Learning Environment

Jun Ming CHEN^{a*}, Meng Chang CHEN^b & Yeali S. SUN^a

^a *Department of Information Management, National Taiwan University, Taiwan*

^b *Institute of Information Science, Academia Sinica, Taiwan*

*d96725001@ntu.edu.tw

Abstract: Reading exercises are critical for developing strong reading comprehension. However, due to resource constraints and a lack of accurate evaluation methods, English instructors can hardly assess student reading ability effectively. In past decades, learning by reading is known to be challenging for both teachers and students involved, especially for students learning English as a Foreign Language (EFL). To cope with this problem, in this paper, we proposed a Tag-based assessment approach to both elicit reading behaviors from EFL learners and assist the teachers in tracing and evaluating the student reading ability effectively. The experimental results showed that the novel approach can not only find out the relations between learners' tags and their comprehension, but also help teachers to evaluate students' reading ability.

Keywords: Social tagging, Collaborative learning, Evaluation methodologies, Information retrieval

Introduction

For learners of English as a Foreign Language (EFL), accurate assessments of student literacy are critical to the success of language instruction, but unfortunately several studies have demonstrated that teachers, due to either a lack of administrative support or time constraints, are unable to assess student literacy effectively [5]. English language teachers rarely have enough time to properly teach reading comprehension, forcing students to rely predominantly on their own intuitions and perceptions when students attempt to understand the structure and concepts of instructional reading material. In some cases, this problem can lead to significant learning obstacles for students in the future.

Consequently, the challenges teachers face when assessing student reading ability levels, are not only limited to a lack of resources, training, or time, but also involve concerns with how assessment tests are conducted. To cope with these problems, we develop an on-line Tag-based Collaborative reading learning (TAC) system designed to aid teachers in accurately evaluating English reading ability.

The remainder of this paper is organized as follows. First, the paper outlines how this paper's social tagging system works to evaluate and score reading comprehension. It then describes how users interact with the TAC system website. Next, it presents our experimental design, reading comprehension results, and the survey feedback data from teacher and student participants. Finally, section five concludes, discusses potential problems with the interpretation of the paper's results, and proposes ideas for future research.

1. A Tag-based Reading Assessment Approach

The assessment approach proposed within this paper gives serious consideration to improving the ability of teachers to evaluate student reading comprehension skills. This section further explains how student tags are used to judge a student's progress in reading comprehension.

1.1 Data Preprocessing

Figure 1 illustrates how this preprocessing phase works, with particular emphasis on several preprocessing techniques for information retrieval, including Porter stemming, irregular verb return to base form, and stop word. Given the different types of input data contained with articles and article summaries, our preprocessing method generates a number of different types of input data for scoring function on different data sparseness processes and different text summarization algorithms.

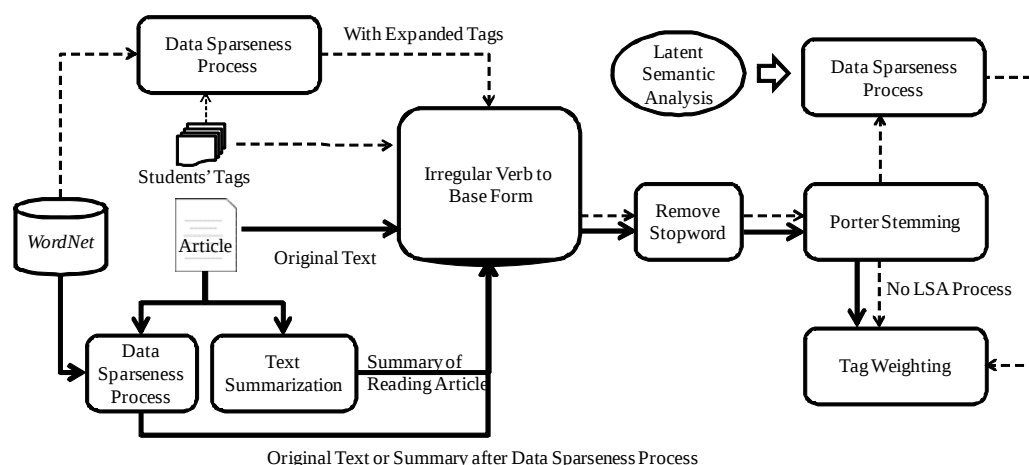


Figure 1: The Process of Student Tag and Article Preprocessing

First, the data sparseness problem is an important concern when implementing text-processing techniques, especially with data retrieved from the web. In order to diminish the impact of data sparseness, this study uses both the WordNet and Latent Semantic Analysis (LSA) methods. WordNet and Latent Semantic Analysis (LSA) also work to ensure the true meaning of the student's tag is successfully captured. Moreover, we apply two summarization algorithms to extract important terms and paragraphs from articles: Context Sensitive Frequency-based text summarization [4] and Latent Semantic Analysis-based text summarization [2]. These two methods of text summarization are sentence-based. Each extracts one sentence for each step in the algorithm process.

Weight calculator for the Article Terms and Student Tags

When weighting terms for articles, we use local weighting $L(i)$ for each term i within the article. This weight is based on the term's frequency, although two possible alternatives are also applied. The first of these alternatives is no weight, where $L(i) = tf(i)$, and the second is a logarithm weight, where $L(i) = \log(1 + tf(i))$. In the case of more than one article, a global weight by factoring the frequency of each term and document length, $G(i)$ is applied to each term i , which is defined as follows:

$$G(i) = \log(N / n(i)) \cdot \frac{(k_1 + 1)tf(i)}{k_1((1-b) + b(L_d / L_{ave})) + tf(i)}$$

Where N is the total number of articles (documents), and $n(i)$ is the number of articles (documents) that contain term i ; L_d is the article length; L_{ave} is the average article length for the whole collection; k_1 and b are represented as the term frequency scaling and the effect of article length normalization, respectively, these parameters are positive tuning parameters. In the absence of such optimization, experiments have shown reasonable values are to set k_1 and b to a value between 2 and 0.75 [3]. After local and global weights are determined, the weight of each article term is finalized as $L(i) \times G(i)$. Initial student tags are weighted as $S(i) = 1$ for each student tag i . When tag i contains an expanded tag from WordNet, the synonym tag is $S(i) \times 0.5$, and the hypernym tag is $S(i) \times 0.3$.

In order to extract semantic relationships from terms and students' tags, the follow procedure was followed. First, a bi-graph was constructed to perform a spreading activation [1] to find all patterns of tags related to a set of article terms. Then, such patterns facilitate to construct semantic relationships, which would be used to draw semantic inferences from generated vector. After final tag activation vectors are generated, the tag weighting ratio was measured by the percentage of tag weight on specific tag, which is defined as

$$S(i) \times w_i / \sum_{j=1}^m w(j)$$

Where $W(j) = \{w_{i1}, w_{i2}, \dots, w_{im} \mid w_i \in W(j)\}$ denotes the final activation vector of tag i , and w_{im} denotes the activation energy (weight) between tag i and tag m .

1.2 Scoring Function for Reading Comprehension

In this paper, the overall *Scoring Function* uses a conventional information retrieval method for weighting articles (documents) by their term frequency, and then computes the cosine similarity between the reading article and student tag vector.

This score can then serve as an important reference for evaluating student comprehension. Since the *Scoring Function* is based on the techniques of a vector space model and vector similarity, the result of the *Scoring Function* is a numerical value. Here, large values signify a strong reading comprehension score. It is important to note however, that in some cases, this value may not be an interval between 0 and 1. As stated previously, the *Scoring Function* is *Scores of Students* = $f(S_j, C_k)$, where $S_j = \{s_{j1}, s_{j2}, \dots, s_{jn}\}$ is a set of student tags, and $C_k = \{c_{k1}, c_{k2}, \dots, c_{kn}\}$ is the different type of input elements related to the reading article.

As mentioned within the preprocessing section earlier, each element in S_j and C_k is a vector by its weight (the vector may or may not be a unit vector), where vector $s_1 = \{s_{11}, s_{12}, \dots, s_{1n}\}$ implies that student s_1 has n tags, and vector $c_1 = \{c_{11}, c_{12}, \dots, c_{1n}\}$ implies that the input content c_1 has n terms. The next step is to then perform the dot product calculation between these two vectors in order to calculate each student's individual score:

$$\text{Score of one Student's Tags} = s_j \times c_k = \sum_{i=1}^n s_{ji} c_{ki}$$

2. A Tag-based Collaborative Reading Learning (TAC) System

Given the scoring methodology of our system outlined above, this section also covers the teacher interface designed to aid teachers in student assessment.

Figure 2 shows the teacher interface for browsing the portfolios of individual students, and provides an assessment of the student's learning status, recently tagged discussions, TACA score and tags related to discussions. The 'Learning Status' information reflects the reading status and tagging behavior of individual students, which helps teachers observe student thought processes, track changes in ideas to item tags over time, and create channels for social interaction.

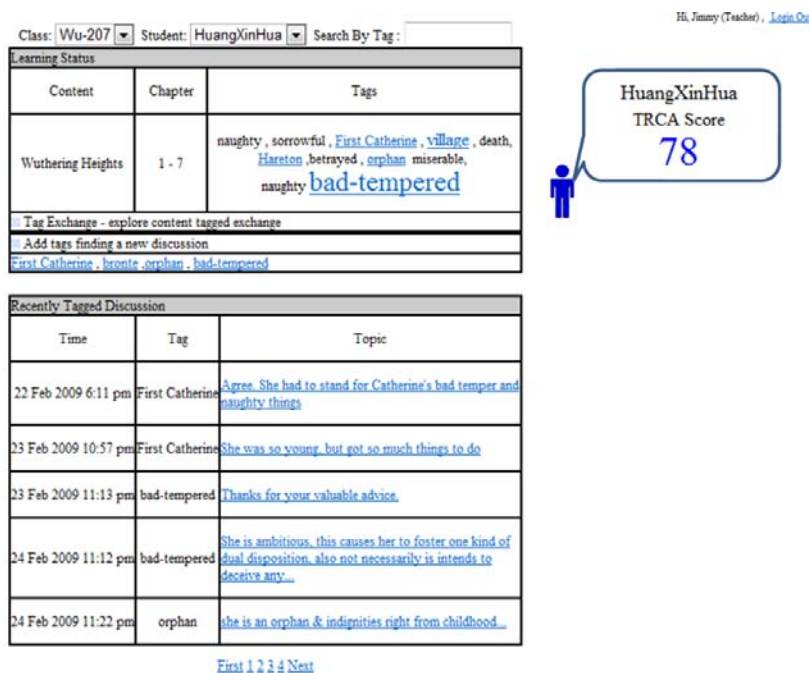


Figure 2: Interface of Tag-based Reading Comprehension Assessment

3. Experiment and Analysis

In order to thoroughly verify the relationship between reading comprehension and student data tagging, and in order to help teachers' judge student comprehension, this paper created an experimental test. First, test scores were obtained from the reading tests, and expert rating tag sets were obtained by evaluating tags that students used. This set of data represents the available variables for judging student comprehension. In our evaluation of this information, we calculated the Spearman Rank Correlation between the expert scores and the TACA scoring results of the student tags. For compiling the expert opinions of teachers, we employed a questionnaire based off of the Delphi method. In this method, researchers use multiple rounds of questionnaires to collect data until expert consensus of opinions emerges.

Table 1 depicts the scores from these experts, and their correlation with the results of the Scoring Function process. From the experimental results, it is clear that the preprocessing of tag sparseness significantly enhances the assessment of student performance. For example, the different preprocesses of student tags significantly increased the correlation value. This was verified using a 2-tailed t-test on performance between different methods that have a statistically significant difference. As the results show, "WordNet" is better than "LSA," but not significantly. However, if "WordNet" and "LSA" are combined, the results of this combination are significantly better than "WordNet" alone.

Table 1: The Spearman Rank Correlations between the Scoring Function and Expert Scoring

Spearman Rank Correlation (Delphi – Five Experts)		Different types of data sparseness preprocessing for student tags			
		Original Tags	Original Tags + WordNet	LSA Process	LSA + WordNet
Input text for scoring function	<i>Sum_{Human}</i>	0.791**	0.829**	0.790**	0.848**
	<i>Sum_{LSA}</i>	0.660**	0.715**	0.730**	0.770**
	<i>Sum_{CF}</i>	0.712**	0.740**	0.745**	0.787**
	Students' Tag Set	0.610**	0.643**	0.670**	0.700**
	Expert's Tag Set	0.688**	0.714**	0.716**	0.755**
	All Reading Text	0.702**	0.730**	0.721**	0.775**

** p < 0.01

4. Conclusions

As stated throughout this paper, the primary goals of our TAC system are to help teachers gauge student progress and literacy. The experimental results described above strongly imply that our social tagging-based method accomplishes these goals. Currently, we are planning to extend TAC to contain more functions and options, such as additional avenues for brainstorming and in-depth discussions, which might be useful in analyzing the usage of those digital materials and the reading behaviors of students.

Acknowledgements

This work was supported by the National Science Council of Taiwan, R.O.C., under contract numbers NSC- 97-2631-S-008-004.

References

- [1] Chi, E. H. & Naim, R. (2010). Information seeking with social signals: anatomy of a social tag-based exploratory search browser. *Intelligent User Interfaces Conference Workshop on Social Recommender Systems*. (2)
- [2] Gong, Y., & Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. Paper presented at the *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, New Orleans, Louisiana, United States. (4)
- [3] Jones, K. S., Walker, S., & Robertson, S. E. (2000). A Probabilistic Model of Information Retrieval: Development and Comparative Experiments (parts 1 and 2). *Information Processing and Management*, 36(6):779-840. (6)
- [4] Nenkova, A., Vanderwende, L., & McKeown, K. (2006). A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization. Paper presented at the *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, Seattle, Washington, USA. (7)
- [5] Wang, T. H., Wang, K. H. & Huang, S. C. (2008). Designing a Web-based assessment environment for improving pre-service teacher assessment literacy. *Computers and Education*, 51(1), 448-462. (8)