Designing a Reading Material Recommendation System for EFL Learners

Chin-Hwa Kuo and Chen-Chung Chi

CSIE, Tamkang University, Tamsui, Taiwan 251, R.O.C.

Abstract

For numerous people who are English as a foreign language (EFL) learners, reading English articles is an effective activity for improving reading comprehension. In this research, an article recommendation system that identifies articles of suitable difficulty levels for EFL learners was designed. The system design was based on the vocabulary sets of the General English Proficiency Test (GEPT). Using text mining and classifying techniques, the system compares the difficulty levels of articles found on news Web sites and in textbooks, as well as articles written by EFL high school students. In this study, language learners’ current language proficiency levels were assessed to create the learning environment introduced in Krashen’s second language acquisition theory. The document classification verification results indicated that the reading material recommendation system (which is based on GEPT vocabulary sets as the foundation of article feature extraction) can effectively classify the difficulty levels of vocabularies contained in articles of various difficulty levels. Additionally, articles that complied with learners’ language levels based on the evaluation results were used as the reading materials for learning purposes.

Key Words: Article Recommendation System, Cosine Similarity, Document Readability, Second Language Acquisition

1. Introduction

The English language has become the dominant language worldwide because of globalization. Learning English is becoming essential for students in countries where English is not natively spoken. The reading material recommendation system introduced in this study was designed to provide learners with articles written in difficulty levels appropriate for learning purposes. The design of this system was motivated by the concept of flow [1–3]. A flow state can be entered while performing any activity, although it is most likely to occur when a person is wholeheartedly performing a task or activity for intrinsic purposes. Several challenges to staying in a flow state include states of apathy, boredom, and anxiety. These states generally differ from being in a state of flow because flow occurs when challenges match the skill level of a person. In addition, Krashen proposed similar concepts regarding language learning [4]. According to his second language acquisition theory, five hypotheses concerning second language acquisition exist. One of them is input hypothesis: (1) the input hypothesis relates to acquisition, not learning; (2) language is acquired by understanding language samples that contain structures that are a level above a learner’s current level of competence (i + 1), which is accomplished by analyzing context clues or extralinguistic information; and (3) when a learner is at a stage i, then acquisition occurs when a learner is exposed to comprehensible input that belongs to the level i + 1. Therefore, when designing a reading material recommendation system, the recommended texts must correspond with learners’ abilities instead of being excessively easy or difficult. Adopting this principle prevents learners from losing learning motivations because...
of emotions such as apathy and anxiety.

To recommend reading materials written in suitable difficulty levels for language learners during their learning process for them to retain positive learning motivations (i.e., staying in the flow state), reading material recommendation systems must assess learners’ language abilities. Numerous relevant studies have proposed article readability calculation formulas to predict learners’ language levels for providing learners with reading materials at appropriate difficulty levels. Danielson and Bryan, Chall and Dale, Chall, Klare, and Williams and Lixiao [5–10] have proposed that the vocabulary difficulty level and average sentence length are the factors that affect the readability of traditional English articles. Collins-Thompson and J. Callan [11] proposed using a smoothed unigram model for analyzing the difficulty level of English articles, and the American 12th grade vocabulary was used for analyzing the difficulty level of English articles. Petersen and Ostendorf [12] proposed a machine learning method for reading level assessment. Heilman [13] and Miltsakaki [14] both have proposed English article recommendation systems in which users can select an appropriate reading level, the systems provide users with articles, and users can provide feedback after reading the articles. Martin [15] used three types of readability formulas [16–18] to analyze how to reliably measure the readability of Web documents; Pitler and Nenkova [19] proposed a unified framework for text quality prediction and collected readability ratings from college students for performance evaluations. The aforementioned studies proposed numerous assessment methods for evaluating learners’ language abilities and obtained article features that are typical of articles appropriate for learners at various proficiency levels. These data served as the design bases for some recommendation systems. However, the teaching materials used for learners of different nationalities during their learning processes vary. Hence, these formulas may not be suitable for English learners whose first language is not English. When inappropriately applied, learners may experience anxiety because of extremely difficult teaching materials. This study addressed the problem of recommending reading materials with appropriate difficulty levels to EFL learners for reading and learning purposes.

This study is innovative and valuable because of the reading material recommendation system design. The General English Proficiency Test (GEPT) glossary was used as the feature value acquisition basis for English article difficulty classification. The writing assignment data that high-school-level learners uploaded to learning platforms and high school English texts were employed for evaluating the language levels of this specific learner population, thus creating the learning environment introduced in Krashen’s second language acquisition theory and identifying online English articles that are appropriate for the learners’ language levels. These articles were subsequently used as the reading materials that can be recommended for certain learner populations. The reading material recommendation system proposed in this study contains natural language process (NLP) preprocessing, information retrieval models, and document similarity calculation methods. Article vocabulary difficulty feature value sets are generated before they are used for performing feature comparisons. The corresponding results served as the basis for article filtering and recommendation by using naïve Bayes and k-Nearest Neighbors (KNN) document classifiers. Subsequently, document classification accuracy and 2D data point expression methods were used to verify the recommendation results of this study. Online news articles that exhibited appropriate vocabulary difficulty levels were selected and recommended to certain learner groups for learning purposes. By providing learners with articles that contained vocabulary words that were appropriately difficult, learners were aware of the difficulties of reading the articles, but they were able to cope using their current language abilities and skills. Accordingly, learners’ study motivations remained and positive learning environments were created.

The remainder of this paper is structured as follows: section 2 introduces vocabulary databases, article databases, and research tools. The three types of article database sources used in this study were senior high school English textbooks (SHSETs) used in Taiwan, Intelligent Web-based Interactive Language Learning (IWiLL), and online English news articles. The research tools include Orange and the Natural Language Toolkit (NLTK). Section 3 presents a system overview by using system framework flow diagrams to explain the filtering system and the procedure of the EFL learner reading material recom-
mendation system designed in this study. Section 4 verifies system effectiveness by employing a series of approaches to verify the corresponding results, thus indicating the effectiveness of the article recommendation filtering mechanism designed in this study. Lastly, section 5 offers a conclusion.

2. Text Databases and Research Tools

To design the recommendation system, document features by which English articles can be classified into their respective levels of difficulty were extracted. The document features were derived from vocabulary sets and text databases. The vocabulary set used was from the GEPT. The three text databases used were 1) the Sanmin version of SHSETs [20], 2) IWiLL [21], and 3) news articles from online sources.

2.1 Vocabulary Sets

The vocabulary set from which document features were computed was taken from the GEPT [22]. Based in Taiwan, the GEPT was established by the Language Training and Testing Center [23] for examining the English proficiency levels of Taiwanese EFL learners. Specifically, this vocabulary set was adapted from bands 2 to 5 of the Collins Cobuild English Dictionary and is the primary English vocabulary set for high school students in Taiwan. It was revised by the College Entrance Examination Center in the second half of 2004 to include 6 levels of difficulty. The entire database consists of 6,604 English words. A higher band number indicates a greater level of difficulty.

2.2 Text Databases

(1) SHSETs are the English textbooks used in Taiwanese high schools. For this research, SHSETs published by Sanmin were used. In total, 106 SHSET articles selected from 10th, 11th, and 12th grade English textbooks were included in the database.

(2) Established in 1999, IWiLL is an interactive online English learning Web site for Taiwanese high school students. IWiLL has been widely adopted as a digital learning platform across the country. As of December 2009, there are 174 high schools, 1,482 English teachers, and 111,415 students registered on IWiLL. To date, more than 5,000 English articles have been published by registered students. In this research, the English vocabulary proficiency of high school students was inferred from the articles published by registered students on IWiLL. For this research, 4,486 articles written by 2,532 students and teachers were randomly selected. IWiLL was used in this study because it is a unique and regular source for student-authored articles; otherwise, formulating learner profiles would have been challenging. By analyzing these articles, students’ language learning progress can be observed and articles suitable for their current level of proficiency can be subsequently suggested.

(3) News articles were collected from English news Web sites including CNN [24], The China Post [25], and the BBC [26]. The recommendation system identifies online news articles appropriate for high school EFL learners in Taiwan. More than 10,000 articles were collected from these news Web sites for the database.

2.3 Research Tools

This research used (1) NLTK toolkit application programming interface and (2) Orange software for preprocessing and conducting data mining analysis:

(1) NLTK [27] is a platform used for building Python programs that function with human language data. It provides easy-to-use interfaces for over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning.

(2) Orange [28] is an open source data visualization and analysis software. It supports visual programming, visualization, interaction and data analytics, a large toolbox, and a scripting interface. In this research, Orange was used to classify English articles and evaluate classification accuracy.

3. System Overview

The reading recommendation system design includes NLP preprocessing, followed by feature extraction mechanism (FEM) and classification mechanism (Figure 1). The articles selected from high school textbooks and
essays from IWiLL were used to establish the templates for reading difficulty level $i + 1$ and level $i$, respectively. Subsequently, the features were used to identify the difficulty levels of the articles crawled from online news sources; these articles were classified as either difficulty level $i + 1$ or $i + 2$. Thus, the system can select and recommend articles of reading difficulty level $i + 1$ to learners.

The mechanisms in Figure 1 are described in detail as follows:

3.1 Natural Language Process Preprocessing

During the preprocessing stage, common NLP occurred. First, unnecessary document elements such as control characters and HTML tags were removed. Second, the first 400 tokens from each article taken from the three text databases were chosen to avoid distortions in document similarity caused by document size (i.e., token count). The selected tokens consisted of four parts of speech only, verbs, nouns, adjectives, and adverbs. Third, the WordNetLemmatizer function was used to perform POS-tagging and lemmatizing.

3.2 Feature Extraction Mechanism

This step defined document features by calculating the similarity value between GEPT base 6 vocabulary words and selected articles. These document features would then represent the difficulty level of each article. Whereas most search engines use the cosine similarity calculation method to find the similarity between document-document or document-query terms, in this research, searching for document-vocabulary level relationships was achieved. This rendered constructing a vocabulary difficulty measurement mechanism based on a set of vocabulary words that EFL learners are familiar with possible.

The two information retrieval models used were the Boolean model and the vector space model, both of which are information retrieval models introduced by Liu [29]. In the Boolean model, a term is considered either present or absent in a document. The weight $w_{ij}$ is represented as $w_{ij} = \{0, 1\}$ of terms $t_i$ in document $d_j$; therefore, $w_{ij}$ is 1 if $t_i$ appears in document $d_j$, and 0 otherwise. In the vector space model, a document is represented as a weight vector in which each component weight is computed based on some variation of the term frequency (TF) or term frequency-inverse document frequency (IDF) scheme.

In the feature extraction process, document similar-

Figure 1. Document preprocessing, feature extraction and classifier accuracy estimation procedure.
ity was determined using the Euclidean dot product formul\textsuperscript{a} [30], [31] to obtain the dot product and the cosine value between two vectors. As shown in (1), the similarity value between vector \(v\) (vocabulary sets from different levels) and \(w\) (an article) was calculated using the cosine similarity method; \(v_i\) and \(w_i\) were the term weights associated with each keyword \(t_i\) of each vocabulary set and article. When the cosine similarity value was 1, the vector elements between vectors \(v\) and \(w\) were completely identical. This approach involved using two information retrieval models to evaluate two document feature sets by using similarity calculation.

\[
\text{Similarity}_{\text{vvec}}(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|} = \frac{\sum_{i=1}^{N} v_i w_i}{\sqrt{\sum_{i=1}^{N} v_i^2} \sqrt{\sum_{i=1}^{N} w_i^2}} \quad (1)
\]

To illustrate, Table 1 shows two sentences from two documents. The sentence from Document 1 is “English is one of the most popular languages in the modern world.” The sentence from Document 2 is “The server has a small manufacturer logo.” After implementing the feature extract mechanism, the system assigned a difficulty level for each keyword, as shown in Table 1.

Within the Boolean information retrieval (IR) model, the vector element \(w_{ij}\) in vectors \(v\) and \(w\) was set as 1 to denote that a certain keyword \(t_i\) of a certain level appears in article \(d_j\). Hence, the similarity score between vectors \(v\) (vocabulary set from different levels) and \(w\) (the article) can be computed as shown in Table 1.

Table 3 shows the resulting document feature data. The Boolean IR model produces a similarity score between the article and vocabulary set. In Document 1, no keywords from levels 3–6 were identified; therefore, the values for those levels are 0, as shown in Table 3.

The proposed system used two types of IR models to represent key word term weight \(w_{ij}\) in vectors \(v\) and \(w\), and six similarity values were applied to represent the document features in an article. The FEM calculation process based on the vector space model was also incorporated in (1), except that the keyword term weight \(w_{ij}\) was calculated using the TF-IDF instead of \([0, 1]\). In the vector space model, both document and vocabulary sets were represented as sets of terms, and the weight \(w_{ij}\) of each term \(t_i\) in documents were represented as (2), (3), and (4), respectively.

\[
tf_{ij} = \frac{n_{ij}}{\sum_{i=1}^{N} n_{i,j}} \quad (2)
\]

The term frequency \(tf_{ij}\) represents the importance of a certain keyword \(t_i\) found in document \(d_j\), whereas \(n_{ij}\) represents the number of times keyword \(t_i\) appears in document \(d_j\). The denominator is the sum of the number of times keywords appear in document \(d_j\).

\[
\text{idf}_t = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (3)
\]

To obtain the IDF of a vocabulary word, the total number of documents was divided by the number of documents that contain this word. The logarithm of the product acquired was then obtained. The symbol \(|D|\)

Table 2. Similarity scores between documents and vocabulary sets (obtained by applying the Boolean IR model)

<table>
<thead>
<tr>
<th>Similarity</th>
<th>(GEPT_voc_level_1, document_1)</th>
<th>0.57735</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(GEPT_voc_level_2, document_1)</td>
<td>0.2582</td>
</tr>
<tr>
<td></td>
<td>(GEPT_voc_level_1, document_1)</td>
<td>0.36514</td>
</tr>
<tr>
<td></td>
<td>(GEPT_voc_level_4, document_2)</td>
<td>0.1836</td>
</tr>
<tr>
<td></td>
<td>(GEPT_voc_level_5, document_2)</td>
<td>0.25822</td>
</tr>
</tbody>
</table>

Table 3. Document features (obtained by applying the Boolean IR model)

<table>
<thead>
<tr>
<th>Doc. no.</th>
<th>F11</th>
<th>F12</th>
<th>F13</th>
<th>F14</th>
<th>F15</th>
<th>F16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.57735</td>
<td>0.2582</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.36514</td>
<td>0</td>
<td>0.1836</td>
<td>0.2582</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Two example documents

<table>
<thead>
<tr>
<th>Document 1</th>
<th>English is one of the most popular languages in the modern world</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary level</td>
<td>(1) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)</td>
</tr>
<tr>
<td>Document 2</td>
<td>The server has a small manufacturer logo</td>
</tr>
<tr>
<td>Vocabulary level</td>
<td>(1) (5) (1)(1) (1) (4) (5)</td>
</tr>
</tbody>
</table>
represents the total number of documents in the article database, and \( \{j: t_i \in d_j\} \) represents the number of documents that contain vocabulary word \( t_i \).

\[
\text{w}_{ij} = \text{tf}_{ij} \times \text{idf}_{ij}
\]

(4)

The variable \( w_{ij} \) represents the TF-IDF weight of vocabulary word \( t_i \) in the entire document set \( d \).

As shown in Figure 2, cosine similarities have been computed for each English article and each GEPT vocabulary word from levels 1 to 6 (corresponding to series 1 to 6 in Figure 2). The X-axis (Document No.) represents the serial number of the English textbook. In this figure, each article has a feature set based on the article and six levels of words in the GEPT vocabulary set. Vocabulary words in GEPT levels 1 and 2 (labeled as Series 1 and 2) are the words most frequently used in these English articles.

3.3 Classifier Accuracy Estimation and Document Recommendation

After the FEM process, two sets of document features based on the Boolean and vector space models were obtained. Two sets of document features can be obtained as input data in this step, and each set had six features to represent the vocabulary weights in GEPT levels 1–6. This study employed these numerical data as the document feature values that document classifiers used for determining readability levels. Consequently, the classification accuracies of the various types of feature value sets were compared to determine the feature value sets that can be used as the foundation of the reading material recommendation system. A classifier accuracy estimation flowchart, as shown in Figure 3.

(1) Raw text file input

The two document feature sets represented two IR models. Each type of IR model can generate 12 groups of similarity feature values between the vocabulary words and articles to represent the vocabulary difficulty level of...
each article. The corresponding data are represented as L11–L16 and L21–L26 in the data table. The other seven document readability measurement methods that were described in section 1 were applied for method comparisons in the next section.

(2) Data selection and data sampling

Dozens of articles were randomly selected from the three types of text databases in the program. Class labels were assigned to the articles, which were marked as Class 0: text obtained from the IWILL learner corpus, Class 1: text obtained from the SHSETs, or Class 2: text obtained from English news Web sites.

(3) Attribute selection (selecting document features for classifier input)

Document features were selected as attributes for testing classifier accuracy. Regarding this procedure, this study applied 12 groups of feature values obtained using two IR models and seven types of article readability calculation formulas as the references for comparing the classification accuracies of various document classifiers.

(4) Document classification

This study used two classifiers (naïve Bayes and KNN) in the evaluation step. These two classifiers accepted the different feature value data sets obtained from the various types of articles shown in Table 4. This study used classifiers to compare whether online English news articles conform to the article features of the texts used in senior high schools. When articles from English news Web sites conformed to the corresponding features, they were classified as document types that high school texts belong with and were thus recommended to learners as reading materials.

(5) Classifier evaluation

Data training and testing programs were conducted using the widget “Test Learners.” The “cross-validation” [32] of this widget was adopted in this study. The number of folds was set as 10. The “Cross-validation” settings in the widget split the data into the training and testing sets by the given number of folds by 10. The algorithm was tested by holding out the examples from one fold at a time; the model was induced from the other folds and the examples from the held out fold are classified. Once each step was performed as described, widgets called “Test Learners” and “Confusion Matrix” were used to show the classifier performance estimation results.

4. System Performance Evaluation

M. Konchady described document classification as “Documents are assigned to one or more categories based on the degree of similarity description. A classifier uses a similarity measure to evaluate documents against categories to find the closest category” [33]. This study used article similarity calculation methods and vocabulary difficulty evaluation formulas as the article classification bases. Additionally, based on learners’ writing samples and high school text information, the articles in the databases were classified into three categories: 1) Stage i articles that comply with learners’ language levels, 2) stage i + 1 articles that require slightly more advanced language abilities than learners’ current lan-

Table 4. Raw data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1600001</td>
<td>.0615</td>
<td>.0339</td>
<td>.0336</td>
<td>.0137</td>
<td>.0094</td>
<td>0</td>
<td>.1704</td>
<td>.0302</td>
<td>.0218</td>
<td>.0109</td>
<td>.0082</td>
<td>0</td>
<td>6.82</td>
<td>8.66</td>
<td>6.73</td>
<td>72.04</td>
<td>9.3</td>
<td>9.8</td>
<td>9.39</td>
<td></td>
</tr>
<tr>
<td>1600002</td>
<td>.1139</td>
<td>.0182</td>
<td>.0039</td>
<td>.0144</td>
<td>.0041</td>
<td>0</td>
<td>.2168</td>
<td>.0207</td>
<td>.0051</td>
<td>.0051</td>
<td>.0026</td>
<td>0</td>
<td>3.99</td>
<td>6.71</td>
<td>4.74</td>
<td>81.78</td>
<td>7.2</td>
<td>8.49</td>
<td>8.14</td>
<td></td>
</tr>
<tr>
<td>1600003</td>
<td>.0923</td>
<td>.0186</td>
<td>.0143</td>
<td>.0106</td>
<td>.0021</td>
<td>0</td>
<td>.1995</td>
<td>.0311</td>
<td>.0206</td>
<td>.0129</td>
<td>.0026</td>
<td>0</td>
<td>7.46</td>
<td>9.05</td>
<td>7.05</td>
<td>70.96</td>
<td>9.2</td>
<td>9.96</td>
<td>9.54</td>
<td></td>
</tr>
<tr>
<td>1600004</td>
<td>.0749</td>
<td>.0167</td>
<td>.01</td>
<td>.0045</td>
<td>.0052</td>
<td>0</td>
<td>.1869</td>
<td>.0192</td>
<td>.0109</td>
<td>.0055</td>
<td>.0055</td>
<td>0</td>
<td>6.03</td>
<td>7.41</td>
<td>6.75</td>
<td>75.5</td>
<td>9.2</td>
<td>9.39</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>1600005</td>
<td>.083</td>
<td>.0357</td>
<td>.0108</td>
<td>.0059</td>
<td>0</td>
<td>0</td>
<td>.2032</td>
<td>.046</td>
<td>.012</td>
<td>.0072</td>
<td>0</td>
<td>0</td>
<td>7.66</td>
<td>9.45</td>
<td>7.68</td>
<td>65.99</td>
<td>9.2</td>
<td>8.89</td>
<td>9.48</td>
<td></td>
</tr>
<tr>
<td>1600006</td>
<td>.0933</td>
<td>.0275</td>
<td>.0151</td>
<td>.0062</td>
<td>.0023</td>
<td>0</td>
<td>.1877</td>
<td>.0276</td>
<td>.0109</td>
<td>.0055</td>
<td>.0028</td>
<td>0</td>
<td>6.94</td>
<td>10.6</td>
<td>6.45</td>
<td>69.07</td>
<td>7.6</td>
<td>8.84</td>
<td>8.47</td>
<td></td>
</tr>
<tr>
<td>1600007</td>
<td>.0769</td>
<td>.0304</td>
<td>.0055</td>
<td>.0105</td>
<td>0</td>
<td>.029</td>
<td>.2235</td>
<td>.0323</td>
<td>.025</td>
<td>.0148</td>
<td>0</td>
<td>0.003</td>
<td>3.87</td>
<td>5.96</td>
<td>4.71</td>
<td>83.76</td>
<td>8</td>
<td>8.97</td>
<td>8.59</td>
<td></td>
</tr>
<tr>
<td>1600008</td>
<td>.1218</td>
<td>.0555</td>
<td>.0082</td>
<td>.0091</td>
<td>0</td>
<td>0</td>
<td>.2198</td>
<td>.0527</td>
<td>.0091</td>
<td>.0068</td>
<td>0</td>
<td>0</td>
<td>7.83</td>
<td>9.42</td>
<td>7.36</td>
<td>69.01</td>
<td>10</td>
<td>10.4</td>
<td>9.99</td>
<td></td>
</tr>
<tr>
<td>1600009</td>
<td>.0607</td>
<td>.0108</td>
<td>.0081</td>
<td>.0063</td>
<td>.0028</td>
<td>0</td>
<td>.1751</td>
<td>.0201</td>
<td>.0114</td>
<td>.0086</td>
<td>.0029</td>
<td>0</td>
<td>6.58</td>
<td>8.69</td>
<td>7.58</td>
<td>67.96</td>
<td>10</td>
<td>10.6</td>
<td>10.2</td>
<td></td>
</tr>
<tr>
<td>1600010</td>
<td>.0129</td>
<td>.0287</td>
<td>.0094</td>
<td>.0049</td>
<td>0</td>
<td>0</td>
<td>.2418</td>
<td>.0299</td>
<td>.0092</td>
<td>.0069</td>
<td>0</td>
<td>0</td>
<td>6.78</td>
<td>8.38</td>
<td>7.5</td>
<td>70.07</td>
<td>9.6</td>
<td>9.86</td>
<td>9.46</td>
<td></td>
</tr>
</tbody>
</table>
guage levels, and 3) English news articles collected online. Because the SHSETs were compiled and meticulously designed by Taiwanese scholars to provide high school students with a collection of reading materials, online English news articles with vocabulary difficulty levels that were classified into this category can be provided to students as learning and reading materials.

This section presents how the following methods were used to verify the effectiveness of the reading material recommendation system: 1) exploring the ability of the system to identify article difficulty levels that are appropriate for learners’ learning, 2) investigating the article source differentiation accuracies of this system compared with those of other methods, 3) using 2D data point graph distribution to verify the rationality of the online English news articles recommended for certain learner populations by the system, and 4) comparing superior student-written articles selected by teachers and the nonselected articles to verify the ability of the proposed article feature extraction methods to identify well-written articles.

4.1 The System’s Ability to Identify Stage i + 1 Level Articles Appropriate for Learners

This section explores the mechanism proposed in this study regarding its ability to differentiate stage i articles, which correspond to learners’ language levels, and stage i + 1 articles, which can be recommended for learners for enhancing their language skills. A confusion matrix was utilized to reveal the evaluation results for the accuracy of the classifiers. In addition, 40–49 articles were randomly selected from each type of article source, and the document feature values were obtained. After being classified using Bayes classifiers, these articles were compared with the documents that were originally the same type as them to verify classification accuracy (Table 5).

The following findings can be obtained based on Table 5: 1) 93.3% (i.e., 127/136) articles can be accurately classified in the article source that they belonged to, indicating that the article recommendation system filtering mechanism designed in this study can precisely differentiate stage i + 1 language learning textbook articles designed by Taiwanese experts and stage i articles that reflect learners’ current language levels and 2) the analysis results showed in Table 5 indicated that two online news articles were classified as containing vocabulary words with difficulty levels similar to those used in SHSETs. Thus, these two articles can be recommended as reading materials for language learners.

4.2 The Article Source Differentiation Accuracy of the System

This section explores the special article feature extraction method proposed in this study and compares it with other difficulty evaluation methods to examine the accuracy of these methods in classifying texts of different difficulty levels. For this process, 39 articles were randomly selected from three article sources, respectively. The document feature values were then acquired, and the Bayes and KNN classifiers were employed for classification purposes. Subsequently, classification accuracies were obtained. The following seven readability index formulas were adopted for comparing the objects: Automated Readability Index (ARI) [34], Coleman-Liau Index (CLI) [35], Flesch-Kincaid readability test, Flesch Reading Ease [36], Gunning Fog Index, Simple Measure of Gobbledygook (SMOG), and its relatively easier-to-use version, SMOG Index. Figure 4 shows the accuracy comparison results. F1–F7 represent the document difficulty feature values calculated using formulas such as ARI and CLI.

Experiment results indicated that using the vector element values in a vector space model to express the document feature values of the six groups of L21–L26 yielded classification accuracies that superior to those
obtained using the Boolean model (L11–L16) and other article difficulty index formulas (i.e., F1–F7). This is because the other article difficulty index formulas each had only one group of index values as classification features. By contrast, this study proposes a text difficulty evaluation mechanism based on the similarity calculation formula between the vocabulary levels in the vector space model and the texts. Thus, more accurate application features of vocabulary words on different levels can be provided for designing recommendation mechanisms. These results also verified that the article recommendation and filtering mechanisms designed in this study yielded higher accuracies than other methods regarding the ability to differentiate various article difficulty levels.

4.3 The Rationality of System-Recommended Online English News Articles for Certain Learner Populations

In this section, 2D data point graphic distribution was employed to verify the rationalities of this system in classifying online English news articles as stage i + 1 articles and recommending them to learners. A total of 40 articles were randomly selected from each of the three article sources, and (5) was used to calculate the difficulty of each article.

$$\text{difficulty(doc)} = \frac{\sum_{i=1}^{6} (i \times \text{similarity}(\vec{v}, \vec{w}))}{\sum_{i=1}^{6} i}$$

(5)

In (5), where vector $\vec{v}$ represents a bag of words in each level of the GEPT vocabulary sets, vector $\vec{w}$ represents a bag of words from the English articles (from IWiLL, SHSETs, and online news), and similarity value means the cosine value between vector $\vec{v}$ and $\vec{w}$ taken as the feature value for the document. The similarity value calculated using (1) and represented as L11–L16 or L21–L26 are shown as sample data in Table 4. It therefore using (4) to represent a difficulty score, which summarize similarity weights in each GEPT vocabulary levels by multiplying by $i$ (i indicates each GEPT vocabulary level, from 1 to 6). To evaluate document difficulty, this study randomly fetched 120 articles from three text databases, and calculated the document difficulty by using (5). The estimation results are shown in Figure 5.

In Figure 5, each point represents an article, and a higher value indicates a higher vocabulary level in an article. L1 uses a Boolean expression to represent word elements in a document vector whereas L2 uses word frequency. Variable T1 indicates the use of (5) to evaluate document difficulty. The distribution of article difficulty in different corpus can be observed: articles in class 0 (from learner essay corpus, shown as round data points) represent a lower difficulty score, articles in class 1 (from SHSETs, shown as under triangle points) in a degree of moderate difficulty between the two other text databases, and articles in class 2 (from English news Web sites, shown as X points) have on average the highest difficulty scores.

As shown in Figure 5, under triangle points typically had higher difficulty scores than round points did; however, 20 X points were located in the stage i rectangle.
area, which represents an area with the same difficulty as those of the under triangle points. Therefore, these X points with low difficulty scores can be recommended to learners. The results indicated that 1) the vocabulary difficulty levels of several English news articles were similar to those of high school texts and were slightly more advanced than learners’ current language levels; therefore, these articles are more appropriate to be recommended to learners as reading materials compared with the English news articles that contained vocabulary words that were excessively difficult. 2) If the system recommended articles exhibiting difficulty levels higher than those of stage i + 1 articles, a state of anxiety may occur when challenges are so difficult that they exceed ESL learners’ perceived skill level, causing learners great distress and uneasiness. Hence, the mechanism designed in this study can be used to select articles that contain vocabulary words of difficulty levels that are similar to those of stage i + 1 articles. These articles can create suitable learning environments that are sufficiently challenging (reading difficulties), conform to learners’ language skills, and enable learners to stay in the flow state. If the system can collect sufficient learners’ writing samples, it can continuously determine the language abilities of individual learners. This evaluation served as a feedback for the recommendation system. Accordingly, the system can adjust the difficulties of the recommended articles for individual learners for reading and learning purposes.

4.4 System Extension Application in Identifying Articles Written with Superior Abilities

This section compares superior articles written by EFL learners that were selected by teachers with the nonselected articles to verify the document feature value acquisition methods used in the system and to assess learners’ abilities to produce well-written articles. The steps are listed below:

1) English teachers were instructed to select 100 well-written articles. 2) Equation (1) was used to calculate the similarities between the vectors of various levels of article vocabulary words and article difficulty. 3) Index weighted values were multiplied by vector similarities, which enlarged the similarity weights in high GEPT vocabulary levels (e.g., level-2 vocabulary vector similarity values were multiplied by 3; level-3 vocabulary vector similarity values were multiplied by 9; and the rest can be obtained using the same method). 4) Six vocabulary-article vector similarity means of the nonselected articles were compared with those of the selected articles to identify the features of the high-level vocabulary words used in these two types of articles.

This study used (4) to calculate the similarity values between the vocabulary vectors of various articles and the vectors of GEPT level 6 vocabularies. Figures 6 and 7 present the calculation results regarding vector element similarity, which are expressed using word frequencies and the Boolean model, respectively. Regarding the 100 well-written articles (b100), levels-5 and -6 vocabulary words exhibited vector similarity values that were higher than the average values of the reading reflection articles that were not selected (N5281). These results indicated that when using the article feature acquisition method proposed in this study as an evaluation instrument, superior student-written articles selected by teachers yielded higher high-level GEPT vocabulary usage frequencies compared with those of the nonselected articles. This feature can be used to identify article vocabulary difficulty levels and can serve as a reference for determining the quality of student-written articles.

Figure 6. Average cosine similarity comparison between the 100 well-written articles and the nonselected articles (vector space model).

Figure 7. Average cosine similarity comparison between the best 100 well-written articles and the nonselected articles (Boolean model).
5. Conclusions

Reading material recommendation systems should be designed based on learners’ reading or vocabulary usage abilities. The recommendation system developed in this study adopted the GEPT glossary, which Taiwanese learners are familiar with, as the basis for extracting feature values from English articles before classifying the articles into various difficulty levels, and for designing a reading material recommendation system. Additionally, English articles selected from high school-level textbooks and the writing data that high school-level learners uploaded to learning platforms were adopted to create suitable learning environments that are challenging (regarding reading difficulties), conform to learners’ language skills, and allow learners to stay in the flow state. The research results indicated that the method proposed in this study can accurately evaluate the general language levels (i.e., stage i) of Taiwanese high-school EFL learners and the SHSETs (i.e., stage i + 1) that were meticulously designed by Taiwanese experts. Accordingly, several online English news articles that exhibited similar difficulty levels to SHSETs were identified and recommended to Taiwanese high school students as learning tools. The recommendation system proposed in this study can facilitate learners to maintain positive learning motivations and continue learning by providing articles that are constantly at an appropriate difficulty level.

Regarding future studies involving this approach, three directions are recommended: 1) Collect learners’ feedback on reading the recommended articles and use this feedback as a collaborative-filtering mechanism to double verify the recommendation system proposed in this study. Superior results can be achieved by adjusting weights, thus improving the mechanism. 2) Collect the data of learners from an increasing number of populations to more precisely evaluate the language abilities of learners from various populations and recommend highly individualized learning tools. 3) Apply the recommendation system to learners’ article writing skill evaluation mechanisms as a reference for evaluating learners’ writing abilities.

Acknowledgment

The work described in this paper was supported by a grant from the National Science Council, Taiwan (Project No. NSC 101-2511-S-032-003-MY3).

References


[34] Senter, R. J. and Smith, E. A., Automated Readability Index, Wright Patterson Air Force Base, P. iii, AMRL-TR-6620 (1967).


Manuscript Received: May 14, 2014
Accepted: Oct. 24, 2014