Adaptive Music Video Recommendation System Framework for Foreign English Language Learning

Chin-Hwa Kuo and Chen-Chung Chi*

Department of Computer Science and Information Engineering, Tamkang University, Tamsui, Taiwan 251, R.O.C.

Abstract

For many EFL learners, reading English articles or listening to music has been a good way to improve their English proficiency. However, it’s not always easy to find the appropriate learning multimedia materials that can satisfy the learners’ interest. The approach proposed in this paper uses a music video recommendation framework which fetches videos from Youtubeto be accessed either via mobile device or web browser. Learners can practice the usage of Verb-Noun collocations by watching the dynamic display of lyrics sentence by sentence, as well as listen to a recommended music video playing at the same time. The proposed system fetches collocation words by calculating mutual information of corpus, choosing music lyrics in the range of appropriate reading difficulty by cluster music lyric based on a particular vocabulary difficulty estimation method, and then providing a practice tool for mobile language learning.

Key Words: Language Learning, Mutual Information, Document Difficulty, Document Readability, Music Recommender System

1. Introduction

EFL (English as Foreign Language) learners seldom use English in daily life, which causes the corresponding obstacles to Taiwanese students [1]. To strengthen the effect of English learning, there is a huge amount of teaching materials on the Internet, such as foreign TV programs, movies, newspapers, books, radio, music, teaching websites, etc. These retrieved learning materials will be more interesting and effective for EFL learners if the researchers combine multimedia with language learning [2].

Vygotsky’s “zone of proximal development” [3] follows the pedagogical rule ‘recommended LAs should have a level a little bit above learners’ current competence level’. Thus, RSs in TEL have to take into account competence levels in order to suggest an appropriate L.A. According to Krashen’s theory of second language acquisition [4], there are five hypotheses about second language acquisition, and one of them is the input hypothesis: (1) the input hypothesis relates to acquisition, not learning; (2) language is acquired by understanding language that contains structures a bit above the learners’ current level of competence ‘i+1’. This is done with the help of context or extra-linguistic information. (3) If a learner is at stage i, then acquisition takes place when learner is exposed to ‘Comprehensible Input’ that belongs to level ‘i+1’. This is the most important feature in our study, since the recommendation system needs to consider the learners’ reading ability or vocabulary using ability. Therefore, a pre-test procedure or a reading degree setting needs to be referenced for a precise estimation of i+1 level material [5]. This study evaluates the learners’ stage ‘i’ by using the essays database in the learners’ corpus and the system makes recommendations in which the reading difficulty of the music lyrics can fit stage ‘i+1’ to meet learners’ learning demand.

Some research related to the topic of providing multimedia for language learning are described below and can be divided into, have two categories: (1) Retrieved
readings as language learning material: Hsu et al. [6,7] designed a reading recommendation system based on the knowledge approach. Using domain knowledge from multiple experts to construct repertory grids for English article recommendation, they designed a recommendation strategy for their proposed system. Heilman et al. [8] and Miltsakaki [9] have both implemented a reading material retrieval system for the purposes of vocabulary practice and matching the readers’ reading skill. Both systems not only provide a reading display interface, the use of vocabulary and sample sentences, and provide filtering options like reading level range, text length, and topic of interest, but they also use some commonly used readability analysis formula, and collect experts’ or learners’ feedback to verify the performance of the reading retrieval-recommender systems. In contrast to their point of view, this study tries to provide song lyrics for reading material so that a proposed document difficulty estimation method can be used for analyzing the reading difficulty of lyrics. Besides, collocation quizzes that take words from the lyrics can help learners to become more familiar with vocabulary at a specified level, since some commonly misused collocation words from the learner corpus will also appear in the quiz. (2) Retrieved audio, music video or movies as language learning material: Ono and Ishihara [10] show that the podcast system is a very effective tool, which sends digital materials to their players automatically. This study suggests that a podcasting system can be used as a supplementary tool for language learning; Shea [11] indicates that a well-designed video application can motivate, save time, and help learners overcome their difficulties; Choi and Johnson [12] found that video applications can enhance learners’ motivation and satisfaction and keep their attention. Moreover, Kuo et al. and Hung [13,14] both use movie dialogue and process the video to provide a Syntax-Base teaching system; Tsao et al. [15] construct an automatic collocation correction and subtitle collocation extraction system, where the misused collocations are provided by English learning platform. All these studies use some form of natural language processing, multimedia processing and information retrieval technologies for language learning, and the proposed system in this study contains the idea of incidental language learning that was proposed by Brett [16]: if messages possess the property of salient, they can catch the learners’ attention and strengthen their comprehension. Due to the above-mentioned property, the goal of incidental language learning is achieved.

Based on Krashen’s theory of second language acquisition, this study designs a music video retrieval and recommendation system, which provides some language learning elements, e.g. listening to the song, reading they dynamic display of lyrics, as well as V-N collocation quiz testing, such that learners can learn the use of vocabularies and collocations, and how to sing the song/speak the lyrics in a song. This study makes use of information technology on multimedia learning tools to help language learning. As the music is being played, the lyrics will show up simultaneously. When verb-noun collocations appear in the lyrics at a particular moment, the system will show a quiz for learners. Learners can study the correct usage of verb-noun collocations via the process of listening to music, answering quizzes, and getting feedback from the system.

In the second section, system architecture, a recommendation system framework is introduced. This study uses Facebook for keeping personal public information, a web-based music video recommending and playing interface, and various databases and corpora that were applied in this framework. In the third section ‘Method’, both mutual information calculation methods and lyric difficulty score evaluation equations are introduced to create a verb-noun collocations database and estimate the vocabulary difficulty from music lyrics. In the fourth section ‘Implementation’, a music video recommendation system interface, quiz in lyrics, and various system performance evaluation methods are explained. Finally, ‘Conclusion’ section is placed in the final section.


The proposed system is divided into several parts, which are described below:

2.1 Login Process for Keeping Personal Information

The system connects with Facebook Graph API [17] at the top of the system using flow. Users can use their familiar login process, e-mail and password, and then the system will keep their Facebook ID and collect public social information as the basis of personal recommendations, since most people do not want to disclose too much personal information in our experiment. In the
user’s first time, the system will recommend music videos according to the difficulty of lyrics to meet the learner’s vocabulary ability, or they may select a vocabulary difficult degree (from 1 to 10) of lyrics in the music video themselves. Then the system will adjust the music video recommendation list according to the users’ feedback – include collocation word testing score and lyric difficulty feedback from user.

2.2 Web-Based Music Video Recommendation List and Playing Interface

The system makes a music video list to meet users’ demand. When users click on an item from a list of music video, the system will play the music video and display lyrics (sentence by sentence, each sentence has a unique timestamp for the reference of music playing program). The music playing process will be paused when a collocation word-pair appears in a lyric sentence; the user will need to choose one the answers and then the system will give feedback to the user before the music resumes.

2.3 Databases

This section describes some databases that were applied in the proposed system, such as a collocation (right collocation and mis-collocation) database, a lyrics database and the implementation of music video player. The system calls Youtube API [18] at any time to meet learners’ demand and the collocation database was constructed for providing the right and wrong use of collocation for language learning.

2.4 Lyrics’ Vocabulary Difficulty Evaluation

The system collects the learners’ feedback when they click on a song from the recommended list, providing some useful feedback information for analyze the users’ English vocabulary ability such as (1) Collocation word testing score; (2) Lyric difficulty feedback in their selected song. Therefore, the system can use both to evaluate the lyric vocabulary difficulty score, and then try to match the users’ vocabulary ability. Figure 1 describes the system architecture in the proposed music video recommendation.

2.5 Corpora Source

This approach use the British National Corpus to collect Verb-Noun collocations and uses GEPT level Six as a document difficulty estimation method to fit the learners’ demand, especially for senior high school students in Taiwan. It also uses the IWiLL language learning platform, SHSETs and Web News corpus to estimate classifier accuracy based on the proposed vocabulary difficulty estimation method in this study, which is described below:

2.5.1 British National Corpus (BNC)

This approach chooses various corpuses for constructing the verb-noun collocation database. For example, the proposed system uses the BNC corpus which contains 100,106,008 words and the size of the database is about 1.5 GB. One-third of the BNC corpus has been used in this approach. Therefore, some natural language processing such as POS tagging, lemmatizing, tokenizing process has been used in order to find all the verb-noun collocations in corpus. The goal of tokenizing to N-gram is to get the subsequence of n items from a given sequence; this study only keeps unigram, bigram and word frequency data for MI calculating.

2.5.2 GEPT Level Six

The document feature retrieval process is based on the vocabulary database from the General English Proficiency Test. GEPT was founded by the Language Training and Testing Center in Taiwan for the purpose of examining the English proficiency of Taiwanese English learners. The vocabulary database comprises of the Collins-Cobuild English Dictionary (Bands 2–5), the English vocabulary for senior high school education revised by the College Entrance Examination Center (CEEC) in the second half of 2004 (a total of six levels), and the level four/six English vocabulary from China. The database divides the English vocabulary into six levels with a total of 6,604 English words/phrases. A higher level indicates a higher degree of difficulty.

2.5.3 Learner Essays in Intelligent Web-Based Interactive Language Learning (IWiLL)

Founded in 1999, IWiLL [19] is an interactive online English learning website that targets senior high school students. There were a total of 174 senior high schools, 1,482 English teachers, as well as 111,415 students registered at IWiLL. There were over 5,000 English articles published by registered students. This study infers the proficiency in English vocabulary of senior high school students from the articles published by the regis-
tered students at IWILL. There were 4,486 articles that have been written by 2,532 students and a few teachers, which is valuable since collecting effective learning profiles (e.g., articles written by students) from students in Taiwan is otherwise difficult in the experiment of this study. Through the analysis of the corpus data, students’ language learning progress and their vocabulary using ability can be figured out both by syntactic and semantic information retrieval, and some appropriate articles can be recommended for their language learning and reading demand.

2.5.4 Readings in Senior High School English Textbooks (SHSETs)

SHSETs are the English textbooks that have been used in senior high schools in Taiwan. The textbooks published by Sanmin are adopted in the proposed system.

2.5.5 Web News Collected on the Internet

Web news articles have been collected from some English news websites, e.g., CNN, The China Post, and BBC. This study tries to identify Web news articles that are appropriate for English learners, especially for senior high school students in Taiwan.

2.6 Collocation and Mis-Collocation

This approach also focuses on choosing verb-noun collocations for language learning. For EFL learners in Taiwan, they usually rely on the Chinese equivalent without hesitation when they use collocations in essay...
writing. For example, “acquire knowledge” is the correct usage. But many learners in Taiwan will use “learn knowledge” or “study knowledge,” as evident in the IWiLL learning platform. The main reason is the interference of the mother tongue. Learners usually make mistakes on verb usage. So this approach applies the mutual information calculating method, and sets a threshold for filtering the right or wrong collocations. If the similarity value is higher than the threshold, then the verb-noun collocation is the correct usage. Therefore each word-pair distance relationship in documents will be observed. If the former one is a verb and the latter one is a noun, and the distance between two words is larger then threshold value, then the proposed system treats the word pair as a collocation. This process collects 2,291,050 collocations from various corpuses that have described in previous section, but they are not necessarily correct. A process to count frequency for each collocation is necessary in order to find out the correct usage of the verb-noun collocations by mutual information calculating. Figure 2 is the data collection process flow for constructing the collocation database.

A quiz making process has been included in the proposed recommendation system. In order to simulate quizzes that will teach learners to avoid popular mistakes, mis-collocation data has also been collected from two sources: (1) IWiLL English learning platform: IWiLL is an English learning platform that contains “IWiLL Community” and “IWiLL Campus”. The corrections of the exams and essays are stored into database. Then these data is used to build candidate lists for the quiz. (2) WordNet: find the synonyms of the verb in collocation. WordNet provides the semantics for each word in different parts of speech and groups words together based on their meanings. The group, called “SynSets”, is used to build candidate lists for the quiz. The candidate lists forms the test quiz base: The proposed system will first select the correct collocation and get all its synonyms. Then it combines the synonyms with the noun in its correct collocation; Figure 3 describes recordings of the collocation/mis-collocation words database.

3. Collocation Database Creation and Document Difficulty Estimation Method

3.1 Mutual Information

Church and Hanks [20] argued for the usefulness of mutual information in identifying monolingual collocations in lexicography. This study applies a mutual information equation to test the similarity value between two instances and compare it with a threshold value. If the similarity value is higher than the threshold value, then the proposed system treats the verb-noun collocation as the correct usage. The mutual information calculating process is shown below: (1) Select the collocation and get frequency from database. (2) Split the collocation into each word and get their frequency. (3) Substitute the result as variables into formula (1). The proposed system sets a threshold of 4. If the similarity value is higher than the threshold, then the verb-noun collocation is the correct usage.

\[
MI(\alpha, \beta) = \log_2 \frac{P(\alpha\beta)}{P(\alpha) \times P(\beta)} = \log_2 \frac{\frac{N}{\text{count}(\alpha)} \times \frac{\text{count}(\beta)}{N}}{\text{count}(\alpha) \times \text{count}(\beta)}
\]

Figure 2. Data collect flowchart in proposed collocation database.

Figure 3. Data layout of the collocation database.
where the variable alpha and beta indicate two instances, respectively; and \( N \) represents the size of the BNC corpus.

The collocation/mis-collocation words database has 37,644 records. This database was created by calculating MI values between 6-pairs of verb words and a related noun word, so that each record has a noun, a right verb and other wrong verbs – at most 5 words. In general the system picks a verb word with the highest MI value for the right answer, and randomly picks other verb words as wrong answers with lower MI values (usually less than 4) in a database record. The average MI value in the collocation/mis-collocation words database is shown in Figure 4. This figure can be observed that the right collocation candidate has a higher MI value than the other wrong collocation candidates, which have an average MI value less than 4.

### 3.2 Lyric Difficulty Score Evaluation

*Document Similarity* utilizes the Euclidean dot product formula \([21,22]\) to obtain the dot product and the cosine value between two vectors, as shown below:

\[
\text{similarity}_{\text{cosine}}(\bar{v}, \bar{w}) = \frac{\bar{v} \cdot \bar{w}}{\|\bar{v}\| \|\bar{w}\|} = \frac{\sum_{i=1}^{N} v_i \times w_i}{\sqrt{\sum_{i=1}^{N} v_i^2} \sqrt{\sum_{i=1}^{N} w_i^2}} \tag{2}
\]

where vector \( v \) represents vocabularies in each level of GEPT vocabulary database; there are 6 levels of GEPT vocabulary, each level has 1,024 words, the lower the vocabulary level is, the more frequently the words in this level were used in documents. Vector \( w \) represents a bag of words from each lyric; and Cosine Similarity means the cosine value between vector \( v \) and \( w \) is taken as the feature value for the document. Vector \( v \) and \( w \) are similar if the similarity cosine value is close to 1.

The equation (2) is commonly used for calculating similarity between query words and a document in a search engine. However, this approach applies GEPT Level Six for similarity calculating, where the word set in each GEPT level is treated as a query word set. Therefore, a relationship between GEPT levels and a document can be discovered. A vocabulary difficulty calculating equation is proposed, which uses a weighted difficulty value to represent a document, as in (3).

\[
difficulty(\text{doc}) = \frac{\sum_{i=1}^{6} (i \times \text{similarity}(\bar{v}_i, \bar{w}))}{\sum_{i=1}^{6} i} \tag{3}
\]

### 4. Recommendation System Implementation and Performance Evaluation

The purpose of this study is to design a mobile English learning application for language learning. Learners can learn the correct usage of collocations anytime, anywhere via listening to music on a web browser or mobile device. The function applied for this study: lyrics have been parsed in the system and quizzes which asks the user to choose one right collocation word from alternative options – will appear in the music playing process. At first, learners need to register for accounts, login to the system, and select a music recommended level. Then learners can choose one song item in the recommended “song list” to listen to music, answer quizzes and leave feedback when the music stops playing. The learner can then select one of the 10 songs, get the dynamic lyric sentences playing in sync with the music, and answer quizzes occasionally, as shown in Figures 5 and 6, respectively.

After a video music is played, the system will ask the learner to leave feedback for both the difficulty of collocation word and the difficulty of the lyric of this music video.

Drachsler et al. [23] identify the goal, user model and conditions of recommender systems for formal and informal learning. They summarize 3 criteria of measurement methods for Recommendation Systems (RS) in Technology Enhanced Learning (TEL), which are technical, educational and social network measures, and they suggest mixing technical evaluation criteria with educational research measures for the analysis of the suitability of RS in TEL. In TEL there are neither stan-
standardized data sets nor standardized evaluation procedures available to evaluate pedagogy driven RSs for formal or informal learning. But focusing only on technical measures for PRSs in TEL without considering the actual needs and characteristics of the learners is questionable.

Drachesler mentioned that in the educational criteria, common measures are Effectiveness, Efficiency, Satisfaction, and the Drop-out rate; these methods are used to analyse the suitability of RS in TEL.

Drachesler mentioned that in the educational criteria, common measures are Effectiveness, Efficiency, Satisfaction, and the Drop-out rate; these methods are used to analyse the suitability of RS in TEL, as shown in Table 1.

4.1 Quiz in Lyrics

There are 39,366 lyrics stored in the system databases. On average, the system provides 3 quizzes during each song. Figure 7 is an example of candidate lists of options. The candidate lists of options are similar to the meaning of the verbs of the correct collocations, but they are wrong collocations. These wrong options in quizzes are similar to common mistakes in the learners’ corpus. Learners will learn the usage of verb-noun collocations via the quiz and feedback from the proposed system.

4.2 Document Classification Accuracy

The system uses the document difficulty evaluation equation (3) to evaluate the vocabulary difficulty of lyrics that have been stored in lyric DB. All music videos that have been fetched from Youtube, will be recommended with a related lyric difficulty level. The proposed approach has evaluated the classifier precision of the proposed vocabulary difficulty formula (2) and (3) by using the Confusion Matrix and other commonly used document difficulty or readability measure methods: Automated Readability Index (ARI) [24], Coleman-Liau Index [25], Flesch-Kincaid readability test [26], Flesch Reading Ease, Gunning Fog INDEX [27], SMOG and its more easily used version: Smog Index [28]. This study chose 117 documents from 3 kinds of document source like the learner essays corpus, readings in senior high school text books, and English articles fetched from news website; each document source chose 39 articles randomly. This study use 3 kinds of classifiers and 8 kinds of document readability estimation formulas to calculate the document readability values to represent a document, and these readability values are used as various kind of document features for training and testing classifiers such as KNN, SVM and Naïve Bayes. The classifier accuracy of these combinations is shown in Figure 8.

The experiment result shows that the proposed document difficulty estimation method “L2T” has a better precision than most other methods (67.42 percent, using Naïve Bayesian classifier). This experiment result indi-
cates that the proposed difficulty estimation method can identify different difficulty levels from various kinds of document sources more precisely, so that this approach can be used to cluster all lyrics for learners. Although some particular learners have a better vocabulary using skill in essay writing, such that the system may have treated their essays as a document of another corpus and affected the accuracy of the classifier result, this experiment result represents the usage of different vocabulary difficulty as document features between each corpus.

4.3 Lyrics’ Clustering Based on Vocabulary Difficulty Estimation

This approach uses a K-Means Clustering algorithm for clustering lyrics in the lyric database according to their vocabulary difficulty score; the clustering result is shown in Figure 9. This experiment shows that the average of lyric difficulty increases by each cluster id, and the average difficulty score increases with the cluster id. Therefore, the system can make recommendations according to each music video by its lyric difficulty and cluster id. On the other hand, two difficulty scores were used to present average vocabulary difficulty in a cluster: M1 indicates the difficulty score is weighted by all 6 levels of GEPT vocabulary vector; M2 indicates the difficulty score is weighted by 5 levels of GEPT vocabulary vector, where the lowest vocabulary level is ignored due

| Table 1. Measurement methods in educational criteria for RS in TEL. |
|------------------------|-----------------------------|-----------------------------|
| Method | Suggested by Drachssler | Methods used in this study |
| Effectiveness | Total amount of completed, visited, or studied learning activities during a learning phase. | Recommend suitable music video and lyrics with appropriate vocabulary difficulty for language learners, which is based on the learners’ corpus and Krashen’s theory of second language acquisition. |
| Efficiency | Indicates the time that learners needed to reach their learning goal. It is related to the effectiveness variable by counting the actual study time. | In a few minutes, during a music video playing, the proposed system can attract learners to listen to the music, read music lyrics, and achieve incidental language learning. |
| Satisfaction | Reflects the individual satisfaction of the learners with the given recommendations. Satisfaction is closely related to the motivation of a learner and therefore a rather important measure for learning. | Activate learners’ learning motivation by music video listening. |
| Drop-out rate | Mirrors the numbers of learners that drop out during the learning phase. In educational research the Drop-out rate is a very important measure because one aims to graduate as many learners as possible during a learning phase. | In the evaluation step, analyze the feedbacks (MV lyrics vocabulary difficulty) and scores (system quiz for testing learners) of learners in the proposed system, to make sure that the learner can benefit during the learning process and the proposed learning activities. |

Figure 7. An example of candidate lists of word-pair options.

Figure 8. Estimation result: document classification accuracy.
to some vocabulary words that appear too frequently in each document. With the M2 difficulty curve, the experiment shows that it has more diversification than the M1 difficulty curve.

4.4 Correlation Relationship Estimation between Document Difficulty and Learners’ Feedback

The correlation relationship seems higher between the learners’ feedback for the difficulty of the displayed quiz, and the ratio of correct answers for the quiz. In Figure 10, the system chooses 28 songs for learners to listen, answer quizzes during these songs, and leave lyric word difficulty feedback at the end of the music video playing. The exam result shows that there is a negative correlation between the learners’ feedback and the ratio of correct answers or the quizzes in songs (N = 28, r = -0.69). This experiment shows that when the difficulty of quiz is harder, the right ratio will be smaller. The system still needs more learners to use this system, so that this feature can be applied to evaluate the difficulty of quiz as a voting scheme by collaborative filtering method in a recommended system.

5. Conclusion

This study proposed an English learning system with a web-based interface and mobile devices. Learners can study the use of Verb-Noun collocation word-pairs in lyrics anytime and anywhere and combine interests of music listening with learning to promote study and provide materials for English learning – both for listening and reading. The features of this paper are as follows: (1) Retrieve verb-noun collocations and automatically make quiz; (2) Provide both web-based and mobile learning environment by music video, audio, dynamic lyrics, as well as correlation word self-testing. (3) Recommend suitable music video and lyrics with the appropriate vocabulary difficulty (learning degree i+1) for language learners, which is based on learners’ corpus (learning degree i) and Krashen’s theory of second language acquisition. Personalization and collaborative filtering can be both used to provide heterogeneous system integrations for further study, such as articles in web pages, music video or any other kinds of materials. These methods still need to be developed for grouping learners’ common interests, collecting some useful public user data for analysis, and providing more multimedia materials appropriate for language learning and retrieved from Web 2.0 network services, thereby giving more accurate recommendations to individuals or their friends.

References

[7] Hsu, C. C., Chen, H. C., Huang, K. K. and Huang, Y.


