Recommend Video Materials to Enhance Language Learning Motivation by Collaborative Filtering Method

Chih-Kai Chang, National University of Tainan, Taiwan

The Asian Conference on Language Learning 2016
Official Conference Proceedings

Abstract
Learning English on Internet has become increasingly common, where Internet movies for learning English are also booming. For instance, the VoiceTube, an Internet video platform for learning English, has pluralistic free videos with both Chinese and English captions. Meanwhile, VoiceTube can combine social media to create a learning network community. However, a good recommendation system is necessary to select proper videos from the English film resources according personal preferences. Hence, the present study used collaborative filtering method to recommend videos, which were found in the learners' lists with similar preferences. First, we used a web crawler to crawl user information on VoiceTube. Then, the Crab, which is recommender engine in Python, was used to analyze the collected data for identifying similar learners. According the preference scores, Crab can precisely recommend proper English learning films to every learner. Finally, we created a query interface for the data crawled from VoiceTube. Thus, learners can use the query function to search friends, collecting similar favorite movies, through VoiceTube social networks. As a result, learners can passively get recommended videos or actively select proper English movies that can enhance their motivation of learning English through watching videos.

Keywords: VoiceTube, social media, recommendation system, collaborative filtering, learning motivation.

Introduction

With the popularity of the Internet, it is relative convenient to access rich and free online resources for learning English. Learning English is no longer just to acquire knowledge from books. A large number of English-language teaching-related Web sites, software, videos and other resources can be used for various purposes. English-language film is a great learning tool while watching entertainment movies. Learners can listen to conversations while watching, and can try to recognize the words on caption. In addition, the situations in the film can help learners into the context of a foreign culture, which is difficult to learn in books.

There are many video sharing platform and YouTube is a popular video sharing site for users to upload, watch, and share videos freely. The VoiceTube (URL: www.voicetube.com) is an English learning website based on watching YouTube videos at Taiwan. Learners can repeatedly play a single sentence to enhance English listening comprehension and reading. In addition to the both English and Chinese captions, learners can use the instant dictionary to check words. Furthermore, VoiceTube combined with social media Facebook. Because VoiceTube is so successful, this study explored how to integrate with recommendation system on this platform.

Furthermore, it is difficult to make a selection correctly among a large number of, various types of film resources according to learners' different preferences. Moreover, quality of online videos cannot judge due to lack of appropriate assessment and management mechanism. How can learners quickly find the right information on Internet becomes a very important issue. Therefore, using recommender system to help users filter out useful information from large amounts of data is necessary. The recommender system can actively provide information to users according to user preferences at the right time.

In summary, this study uses Crab, a flexible, fast recommender engine, to recommend video clips on VoiceTube. To develop the collaborative filtering recommendation system, the web crawler should first crawl through VoiceTube to collect data to be analyzed and discussed. Then, the calculation formula should be designed to obtain the scores of the featured film in order to meet learners' interests. Finally, learners' responses, including English learning motivation and system usability, were collected as feedback to improve the recommender system.

Related Studies

Collaborative filtering

O'Donovan & Smyth (2005) pointed out that collaborative filtering recommender system, sometimes called social filtering, are built mainly on property or interest similar or user experience to provide personalized information service. Based on preference data, users can be divided into a number of group with high degree of preference similarity. Herlocker, Konstan, & Riedl (2000) also mentioned the collaborative filtering system is connecting the users by the same group of people interested to originally predict the extent of a transaction or information. To sum up, the main concept of collaborative filter is to create recommend mechanism from a

large group of users with similar preference records. Then, the recommendation system will try to count assessment scores of every item for the user. Finally, a list of items will be recommended to the user.

Web crawler

Web crawler (Web Crawler) is an automated web browsing programs, also known as Web Robot, or Web Spider (Kausar, Dhaka, & Singh, 2013). Web crawler is an orderly, automated way to visit and retrieve specific information from the web by simulating web browser. It will analyze links to other document or information in retrieved files and will continue to retrieve other files, and so repeated. Therefore, Web crawler often used as one of the basic components in search engine technology, such as: Google, AltaVista, Lycos, and Infoseek. There are some commercial crawler architectures in literatures. For instance, Bingbot is the name of Microsoft's Bing web crawler. Yahoo! Slurp was the name of the Yahoo! Search crawler before Yahoo turn to use Microsoft's Bingbot instead. Googlebot is described in some reports, but only the early version of its architecture was described. Fortunately, there are some Open-source crawlers available. In this study, Scrapy, a free and open source web crawling framework written in Python, was used to systematically collect data from learning website.

System Design

This study designed a collaborative filtering recommendation system for videos of learning English. The system was implemented on the VoiceTube website through its members on the social media Facebook. There are three main steps in the whole process, including Python web crawling, data analyzing and formulating, and Crab recommending steps. In the first step, Web crawler will retrieve users'' favorite videos as the Movie list and collect users' identification with common favorite videos as Friend list. The difference set between user's Move list and his/her friends' favorite videos is the candidate videos as New Movie list. Then, the retrieved data by Web crawler was analyzed and formulated as a Matrix for Crab recommendation system. Finally, this system is expected to recommend a proper list of videos for learning English. The following Figure 1 illustrates the processes of the system design in this study.

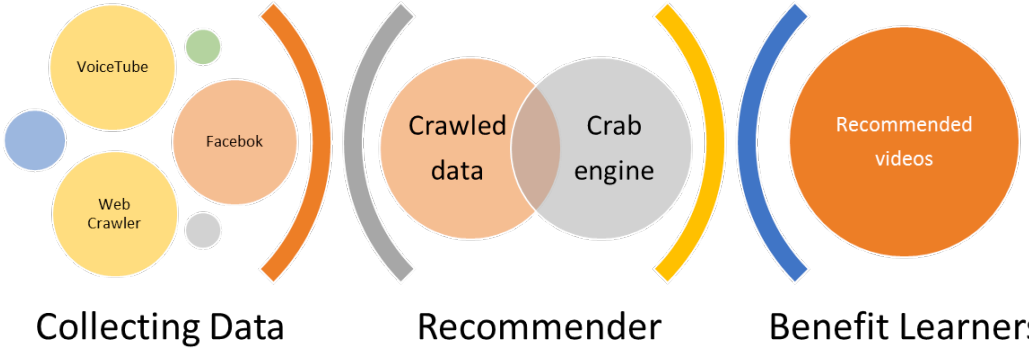

Figure 1. System Design Flow.

Clustering Analysis of Crawled Data

Based on the collected data from VoiceTube by Python web crawler, the study used cluster analysis to observe the distribution of users according to characteristics in their learning profiles. Through analyzing VoiceTube members' learning profiles, we can discover the majority of users viewing characteristics on learning English movie websites. The results of clustering analysis can help us to make more precise adjustments to a pre-established threshold and avoid the prevalent cold start problem (i.e. the system cannot draw any inferences for recommended items about which it has not yet gathered sufficient information in early stage) of a collaborative filtering system.

Since the data is too large, 2000 profiles were randomly selected for clustering analysis. There are users account id, total time to watch videos, number of collected videos, number of collected vocabularies, and other usage information of VoiceTube so on. The results of cluster analysis divide users into five similar groups according to aforementioned data. The following Figure 2 shows the results of clustering analysis, where x-axis for number of watched videos for learning English and y-axis for the total time of watching those videos.

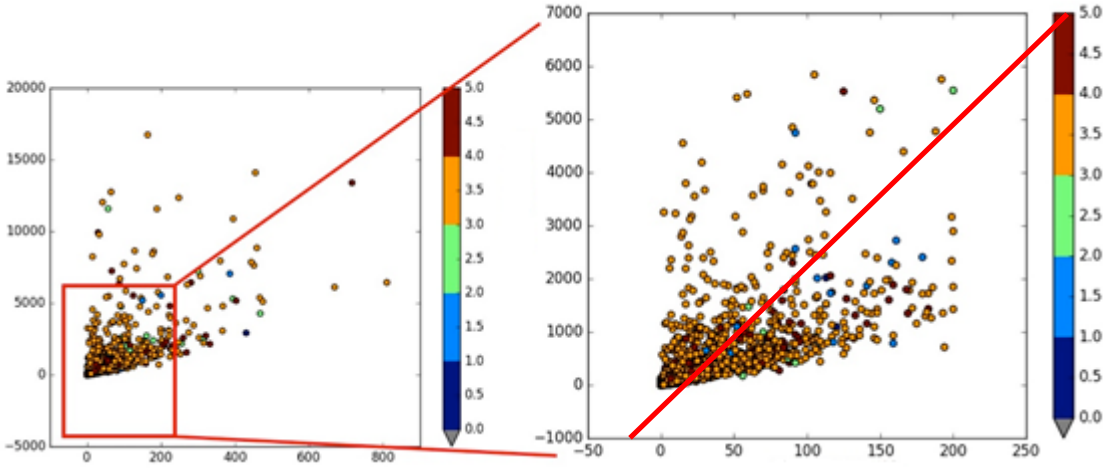

Figure 2. Result of Clustering Analysis

Based on the results shown in Figure, yellow blocks can be found mostly in the figure. The yellow blocks actually occupied 89%, so it can be used as a major group. Furthermore, yellow blocks are widely distributed, mostly gathered in the lower left corner. However, there are still some scattered at another end (i.e. upper right corner). Moreover, the number of yellow blocks at the lower right part is more than the number of yellow blocks at the upper left part. This phenomenon shows that users' learning strategies can be divided into two groups in spite of their similar features (such as collections of videos and vocabularies). The group at the lower right part, may called Extensive Learning group, watched as many different kinds of videos as they can, chiefly for pleasure, and only needing a general understanding of the videos. Another group watched videos with concentration and great care in order to understand exactly the meaning in the videos. It can be seen that most users on VoiceTube will learn English in a Extensive Learning style. In the future, we will

report details about characteristics for that group and how we used those feature to achieve more accurate recommendations.

Conclusion and Future Work

This study uses collaborative filtering method through social media scoring mechanism and Crab recommender engine to recommend English movies according to learners' preferences. Thus, the system can stimulate interest in watching videos to enhance English learning motivation. Consequently, learning English through watching videos are not just interesting; they are engaging and compelling. In addition, we found that members of VoiceTube mostly prefer Extensive Learning style. In other words, learning website, such as VoiceTube, may not suitable for learners who prefer looking up the words and grammatical structures and translating every word. Through this analysis, we not only can understand VoiceTube user features, but also can use the results to revise the computing equation of the recommender system in the future.

References

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000, December). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (pp. 241-250). ACM.

Kausar, M. A., Dhaka, V. S., & Singh, S. K. (2013). Web crawler: a review.*International Journal of Computer Applications*, *63*(2).

O'Donovan, J., & Smyth, B. (2005, January). Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent user interfaces* (pp. 167-174). ACM.

Contact email: chihkai@mail.nutn.edu.tw