A Computational Linguistics-based Framework for Question Generation and Recommendation in E-Learning Systems

Maryam Moshrefizadeh[a]*

[a]Ms.c of Artificial Intelligence, Department of Computer Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, Email: moshrefizadeh@aut.ac.ir

Article History: Received: 14 July 2020; Accepted: 2 January 2021; Published online: 5 February 2021

Abstract: Technology has provided new ways for languages, cultures and the world to be represented, expressed and understood. Nowadays, e-learning is being used more intensively. New approaches are involved with innovative pedagogical tools. Various fields of research can help us provide a new framework for educational tasks. In this paper we use two main fields of research, namely computational linguists and recommender systems to propose a new framework for e-learning. The implemented framework can be used for language learners to improve their language skills in an interactive scenario which automatically generates new questions and recommends the best items to each uses.

Keywords: Natural Language Processing, Recommendation systems, Question Generation, Online Language Learning, E-Learning, Self-study.

1. Introduction

Developing technology-oriented systems is one of the active research topics in recent years. These systems have been widely used in various areas and applications. Educational systems are an example of such systems which received researchers' attention. E-learning systems have been successfully used for language learners to improve their language skills.

E-learning is one of the best technology achievements that allows people learn anywhere. Using technical devices for training makes teaching more exciting and facilitating; e.g., developing online classes gives employees the possibility to be trained and work together in the same period of time (Osipov et al., 2015).

The main advantages of e-learning are as follows (Mohammadyari & Singh, 2014):

a) Using e-learning significantly affects individual performance[1].

b) In e-learning systems, learners can learn without place and time limitation which forms a worldwide training.

c) E-learning provides a more efficient communication among learners and teachers which makes it the basic part of learning and teaching process in the near future.

d) Digital literacy strengthens access, search, evaluate, modify and distribute digital media, and develops skills through the use of new technologies.

e) E-learning systems provide access to a wide network of peers, more up-to-date learning resources, and lowers training costs.

In this paper, we introduce a new system developed for language learners to improve their language skills by generating and recommending various kinds of tests on different topics. Generating tests is fully automatic and based on computational linguistics and natural language processing techniques. Moreover, we benefit from recommendation systems' techniques to personalize the learning process by automatically suggesting the best test items to each user based on their history and weaknesses in educational materials.

The structure of this paper is as follows: in Section 2, we introduce two main research concepts used in our research, namely computational linguistics and recommendation systems. Section 3 provides a background on computer-based testing. In Section 4, we describe our techniques for technology-oriented language learning. Section 5 presents the experimental results and finally, Section 6 summarizes the paper.

1. Background

[1] According to a research model adapted from UTAUT.

1.1. Computational Linguistics

Computational linguistics is a way for computers to analyze, understand, and derive meaning from human languages in a smart and useful way. It benefits from two different research areas, namely linguistics and computer science. This field belongs to cognitive sciences and overlaps with artificial intelligence which aims at computationally modeling of human cognition. It can be studied in both applied and theoretical approaches.

According to Uszkoreit (2005), this field of research is defined as follows: "Computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistics phenomena".

By utilizing computational linguistics techniques, developers can organize and structure knowledge to perform tasks such as parts-of-speech tagging, named entity recognition, syntactic parsing, relationship extraction, stemming, etc (Jurafsky & Martin, 2009).

Computational linguistics is used to analyze text in order to make machines understand how humans speak. This human-computer interaction enables real-world applications like automatic text summarization (Bui et al., 2016), translation (Costa-jussa and Fonollos, 2016), sentiment analysis (Sun et al., 2016), automated question answering, topic extraction, and speech recognition.

Moreover, computational linguistics techniques have been used in educational systems. They are mainly used in educational domain to evaluate exams (Chen & Sheehan, 2015). The other interesting application of computational linguistics in education which has not been studied widely is automatic generation of questions. In this paper we use several applied computational linguistics techniques including parts-of-speech tagging, named entity recognition, and stemming to generate questions for language learners.

1.2. Recommendation systems

One of the main differences between the physical and on-line world is the number of choices that their customers have. While a physical bookstore offers hundreds of books to their customers, an on-line bookshop offers millions of books. While a physical newspaper can only print limited articles per day, on-line news services offer thousands of articles per day (Leskovec et al., 2014).

Having such characteristics in on-line world leads to the long tail phenomenon; i.e., users find huge amount of information in front of them when they want to look for specific issues. A recommendation system can be known as a program which recommends the most appropriate items to specific users. A recommendation system recommends items according to user's behavior and interaction between items and users. Reducing information overload by retrieving the most relevant information and services from a huge amount of data for providing personalized services is the aim of developing recommender systems. The ability to guess a user's priority and interests is the most important feature of recommenders.

Recommender systems use personalization techniques in e-services and have gained much attention in recent years. Generally, there are two types of recommendation systems:

a) Content-based systems which focus on properties of items.

b) Collaborative-filtering systems which focus on the relationship between users and items.

These systems involve the areas of e-commerce, e-learning, e-library, e-government, and e-business service, and have been widely used in various applications, including recommending products, movies, music, television programs, books, scientific literature, new articles, websites, conferences, hotels, tourist attractions, and learning materials.

Using a recommender system in education such as e-learning services is the main focus of our work. Educational institutions have become interested in e-learning recommender systems since the early 2000s. This type of recommender system usually helps learners choose their favorite courses, subjects, learning materials, and their learning activities. We aim at using it to recommend questions/tests to users and help them by improving their language learning process.

2. History of Computer based Testing

Computer-based testing is one of the examples of e-learning systems for language learners. Computer-based testing is administered at computer terminals, or on personal computers. Adapting to the test taker's language level is one of the main characteristics of this system. Computerized tests were first introduced by The Educational Testing Service (ETS) in October 1992. Converting non-computerized tests to their computerized versions became an important goal of ETS in their next step (Yunxiang et al., 2010).

Although, while some studies indicate that reading is more difficult on a computer screen than on paper, others imply that there is no difference between the two. As reported by Meyer and Poon (1997), reading speed and text comprehension in young adults are better than in older people. The authors mentioned that this difference is due to generational differences; e.g., older people are less familiar with computers.

According to Mayes, Sims, and Koonce (2001) who studied the differences between paper and computer in terms of reading speed, the main issue is the way information is displayed. Another comparison between paper and computer-based learning, shows that students' speed of reading from computer screens is slower than reading from paper (Porion et al., 2015).

In a study from Ball and Hourcade (2011), a reading task from Wikipedia is taken into the consideration in which comprehension and reading speed of participants were compared. The results of their study showed that the performance of older participants was better than that of younger ones. These results suggest a decrease in intergenerational differences. We conclude that reading from a computer would elicit better performances in comprehension test than reading from paper (Porion et al., 2015).

One of technology-based fundamentals of foreign language learning in a short period of time is using pre-defined educational materials. Such systems which are based on live interaction between teachers and the students called "eye to eye". Most of these systems are automated and are divided into two categories: autonomous and social. Our system also has the ability to perform as a social system and interact with users.

"Learning a foreign language is also associated with acquiring the knowledge about other cultures, which is impossible without speech communication and knowledge about linguistic and cultural features" (Osipov et al., 2015). To achieve this goal, we propose a system that can be effective in listening. In eye to eye relation between teacher and student emotional conflicts may occur and cause confusion in decision making, whereas this confusion do not occur when performing tests in virtual environment.

3. The Proposed Model

We propose a system that consists of two main parts, the first part examines users based on two types of questions: namely vocabulary questions that focus on nouns, verbs and conjunctions, and grammatical questions. The question generation is performed automatically using computational linguistics techniques such as parts-of-speech tagging and named entity recognition. One of the important features of this system is the ability to extract questions from a sentence with a minimum number of words. The second part is directly in interaction with user. The questions suggested to each user are selected based on user's history in the system. Using recommendation techniques, the system selects the next appropriate question for user. The following two sections describe our method in more details.

3.1. Test Generation

This system is able to produce questions from any context. In the first step, the system receives an input text containing one or more sentences. In the second step, it determines the number of questions that can be generated according to the number of words in each context. In the third step, the target word for each question is selected based on statistical computational linguistics techniques. Finally, the system generates questions and stores the questions with their answers in the database.

The main part of this scenario is Step 3 which aims at detecting target word using the following techniques:

a) Parts-of-speech tagging to find questions with special parts-of-speeches, namely noun, verb, and conjunction.

b) Named entity recognition to filter out nouns that are named entity.

c) TF-IDF to rank words based on their importance.

Natural language processing tries to understand and parse languages. For this reason, various intermediate tasks are introduced to recognize some of the inherent structures in a language without requiring complete understanding of the language. Part-of-speech tagging and named entity recognition are examples of such tasks. In this application we need to understand the tag of each word to focus on some particular parts-of-speeches such as verbs, nouns, or conjunctions. Moreover, we need to find named entities to avoid selecting them as target question terms.

Considering that the TF-IDF is a factor for identifying the most important words in a text data collection, we also use this factor to find important terms in a text. The calculation of TF-IDF is based on two factors: (1) Term Frequency (TF), and (2) Inverse Document Frequency (IDF).

Term frequency and document frequency are the main important measures that have been used for term weighting. Term frequency can be derived from a single document, but to estimate document frequency, a collection of text data is required.

Term frequency shows how salient a word is within a given document. The higher the term frequency (the more often the word occurs in a document), the more likely it is that the word is a good description of the contents of a document. Considering that different documents have different lengths, there is a possibility to have more occurrences of a term in long documents than short documents. Therefore, term frequency should become normalized by the length of the document (the total number of terms in the document):

$$TF(t) = \frac{Number\ of \times term\ tap\ pears \in a\ document}{Total\ number\ of\ terms \in the\ document}$$

Document frequency, on the other hand, shows how much the word is used within various documents. Of course, words that occur in a large number of documents cannot carry important information. Stop words are a good example in this issue. Therefore, the higher the document frequency (the more often the word occurs in different documents), the more likely it is that the word is a general term that cannot describe the content well. This form of term weighting is often called inverse document frequency or IDF weighting (Manning & Schütze, 1999).

IDF is scaled logarithmically using the following formula:

$$IDF(t) = log\left(\frac{Total\ number\ of\ documents}{Number\ of\ documents\ with\ term\ t \in it}\right)$$

Using the above formula, each word that occurs in only 1 document receives full weight (log N/df_t = 1og N/1 = log N), while a word that is occurred in all documents receives zero weight (log N/df_t = 1og N/N = log 1 = 0).

Since some of the written words are misspelled in the corpus and the number of mistakes are usually less than other words, misspelled words get higher score with this criterion compared to other words. We use a threshold for TF factor to solve this problem. After selecting candidate target word using the above procedure, two different question styles are generated: fill in the blanks questions and multiple-choice questions.

The fill in the blanks tests are generated by removing all selected terms from the text and then are shown to the user to find the correct place of each word. The multiple-choice tests require one additional step to generate other new words to be presented to the user together with the target word as possible choices. To this aim, additional computational linguistics techniques such as stemming and synonymous extraction are used and that results generating both vocabulary and grammatical questions.

3.2. Self-examiner

In general, the developed framework can deal with text from different topics which should be identified when loading a text in the system. For example, the following topics have been used in our experiments: family, education, business, environment, and hobby. In addition, the questions could be in two styles, namely academic and general. Besides, users can deal with questions in five levels of difficulty: very hard, hard, medium, easy, and too easy.

Each user can select his/her favorite topic and create a new exam for themselves to test their language skills and try to improve them. For each user, after defining tests, each question is chosen from the list of questions saved in the database during the question generation step. When a user enters for the first time, the first question is selected according to the chosen category and randomly selected from questions with different difficulty levels.

When user takes at least one test in the system, the system starts using its content-based recommender to recommend a new question to the user, based on their previous mistakes (whether it was a vocabulary or a grammatical point) while considering the topic and style of the wrongly-answered test questions. If user correctly answers all the questions from a particular level of difficulty, next question will be selected from higher level questions.

4. Evaluation

4.1. Experimental Setups

To perform our evaluation, we selected 55 English text documents, each contains 148 words on average. The documents are in five categories: family, education, business, environment, and hobby. For each document, our proposed method extracted both fill in the blanks questions and multiple-choice questions. For both type of questions our method selects the complex words according to TF-IDF factor using the *frequency word lists2012*[2]. For pos-tagging and NER, we used the NLTK toolkit (Bird et al., 2009).

As mentioned, to generate the other options of multiple-choice questions, word stemming and word meaning should be taken into consideration. For word stemming, we used the *morph corpus*[3] and the stemming of NLTK. For word meaning, we used WordNet from NLTK.

4.2. Results

To evaluate the accuracy of our question generation method, we asked two annotators to assess our questions manually. The annotators were English teachers with more than 2 years of experience.

For fill in the blanks questions, they were asked to rate each text based on the target candidate words that were selected to make the question. The results of manual assessment of fill in the blanks questions are presented in Table 1.

Table 1: Evaluations results for fill in the blanks questions

	Annotator 1	Annotator 2	Average
score of word selection	76.6%	76.89%	76.7%

For multiple-choice questions, in addition to the above factor, we also asked the annotators to rate the questions based on the quality of the other words suggested as options. The results of manual assessments of multiple-choice questions are presented in Table 2.

Table 2: Evaluations results for multiple-choice questions

	Annotator 1	Annotator 2	Average
score of word selection	50.2%	82.5%	66.35%
score of word suggestion for other options	67.35%	80.8%	74.1%

As can be seen in the tabulated results, the evaluators rated word selection as 76.7% for fill in the blanks questions and 66.35% for multiple-choice questions. Moreover, word suggestion for other options in multiple-choice questions was rated 74.1% by evaluators. According to the results, evaluators approved fill in the blanks

2 https://invokeit.wordpress.com/frequency-word-lists

3 http://www.issco.unige.ch/downloads/multext/mmorph-2.3.4_2.tar.gz

questions more than multiple-choice questions. As mentioned, to the best knowledge of the authors, this is the first system developed for such task and there is no baseline system to compare our method.

For estimating the recommender system, we extracted 10 various exams for two students, and they didn't see any repetitious questions. Also, contexts in their exams and questions were designed from various contexts.

5. Conclusion

In this paper, we presented a new framework for a language learning system, which works based on computational linguistics techniques as well as recommender system tools. Our proposed framework provides a technology-oriented system for e-learning.

Although the proposed framework is developed for language learners, it can also be used by teachers to generate new questions for their exams. Our system is very effective for improving learners' language level and considered a self-study language tool.

References

V. Osipov, A. Volinsky, E. Nikulchev, Y. Prasikova (2015). Study of Gamification Effectiveness in Online eLearning Systems. International Journal of Advanced Computer Science and Applications(IJACSA).

S. Mohammadyari, & H. Singh (2014). Understanding the effect of e-learning on individual performance: The role of digital literacy. Computers & Education.

H. Uszkoreit (2005). What is Computational Linguistics? http://www.coli.uni-saarland.de/~hansu/what_is_cl.html (accessed at 2017/3/20).

D. Jurafsky, & J. H. Martin (2009). *Speech and Language Processing (2nd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

D. D. A. Bui, G. Del Fiol, J. F. Hurdle, S. Jonnalagadda (2016). Extractive text summarization system to aid data extraction from full text in systematic review development. Journal of Biomedical Informatics.

M. R. Costa-jussa, & J. A. R. Fonollosa (2016). Character-based Neural Machine Translation. In proceedings of ACL.

Sh. Sun, Ch. Luo, J. Chen (2016). A review of natural language processing techniques for opinion mining systems. Information Fusion.

J. Chen, & K.M. Sheehan (2015). Analyzing and Comparing Reading Stimulus Materials Across the TOEFL Family of Assessments. Technical Report (RR-15-08, TOEFLiBT-26), ETS Research.

J. Leskovec, A. Rajaraman, J. D. Ullman (2014). *Mining of Massive Datasets* (2nd ed.). Cambridge University Press, New York, NY, USA.

Li Yunxiang, Gao Ruixue, Ren Lili, Wangjie, Qiao Qinshui and Hefei (2010). Advantages and disadvantages of Computer-based Testing: A case study of service learning. *The 2nd International Conference on Information Science and Engineering*.

B. J. F. Meyer, & L. W. Poon (1997). Age differences in efficiency of reading comprehension from printed versus computer-displayed text. Educational Gerontology.

J. Noyes, & K. Garland (2003). VDT versus paper-based text: reply to Mayes, Sims and Koonce. International Journal of Industrial Ergonomics.

B. Ball, & J. P. Hourcade (2011). Rethinking reading for age from paper and computers. Journal of Human-Computer Interaction.

A. Porion, X. Aparicio, O. Megalakaki, A. Robert, Th. Baccino (2015). The impact of paper-based versus computerized presentation on text comprehension and memorization. Computers in Human Behavior.

Ch. D. Manning, & H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA.

Bird, Steven, E. Loper, & E. Klein (2009). Natural Language Processing with Python. O'Reilly Media Inc.