Journal of Applied Science and Agriculture, 9(20) Special 2014, Pages: 23-28

AENSI Journals

Journal of Applied Science and Agriculture

ISSN 1816-9112

Journal home page: www.aensiweb.com/JASA

A Subjective Question and Answer System: An Implementation of Stemming, Stop List, Hamming, Latent Semantic Indexing, Log-Likelihood Ratio Summarization and Vector Space Model Algorithms

Gan Leng Keong, Maziani Sabudin, Siti Khaotijah Mohammad

Universiti Sains Malaysia, School of Computer Sciences, Malaysia

ARTICLE INFO

Article history:
Received 30 September 2014
Received in revised form
17 November 2014
Accepted 25 November 2014
Available online 13 December 2014

Keywords:
Subjective Question and Answer System, Stemming, Stop List, Hamming, Latent Semantic Indexing, Log-likelihood Ratio, Vector Space Model

ABSTRACT

Background: There are only few systems have been developed for the SQ&A System. This is because the subjective questions did not have definite evaluation standard and examiners had different evaluation standard for different students. **Objective:** The Subjective Question and Answer System (SQ&A System) is a system that provides a platform for examiners to automatically grade their subjective examinations' questions. The algorithms used are Stemming, Stop List, Hamming, Latent Semantic Indexing (LSI), Log-likelihood Ratio and Vector Space Model (VSM). These algorithms were used to process the user's input as the answer and determine its relevancy according to the answer stored in the question and answer system database. Stemming algorithm is used to improve retrieval effectiveness and to reduce the size of indexing files. LSI is used to check the spelling mistake of the answer that consists of one word and Log-Likelihood ratio is for summarization. Vector Space Model is used to ensure that the performance of the system component is high by effective implementation of data type, data searching algorithms, sorting algorithms and mathematical equations. **Results:** Comparison using four users shows significant different between the manual evaluation and SQ&A system. **Conclusion:** Implementing the Hamming algorithm and LSI had increased the accuracy of answer evaluation as well as enhanced the SQ&A system. However, there is a significant different between the SQ&A system with manual evaluation due to limited synonyms words in the dictionary.

To Cite This Article: Gan Leng Keong, Maziani Sabudin, Siti Khaotijah Mohammad, A Subjective Question and Answer System: An Implementation of Stemming, Stop List, Hamming, Latent Semantic Indexing, Log-Likelihood Ratio Summarization and Vector Space Model Algorithms. ***J. Appl. Sci. & Agric.,*** *9(20): 23-28, 2014*

INTRODUCTION

Many efforts had been carried out to design and develop the Objective Question and Answering System (OQ&A Systems). Some developers had successfully developed the OQ&A System either in web-based platform or desktop application in order to solve the evaluating module for multiple choice questions' problem (Xuejun Liu *et al*., 2004). Unfortunately, only few systems have been developed for the SQ&A System. This is because the subjective questions did not have definite evaluation standard like objective questions. The examiners always had different evaluation standard for different student (Zhang Xiao Shu, 2012). This will cause the unfair among the students.

The subjective questions can be answered in the length of few paragraphs until few pages. Therefore, evaluating subjective answers will be time consuming because evaluating strategy need to be applied while evaluating those questions (Roy Johnson, 1992). Examiners may get frustrated as the action keeps repeating many times and the answers, which are given by the students, may be too ridiculous (Daniel Midgley, 2011). Mistakes happened while evaluating the answers of the subjective questions. Therefore, a lot of processes are needed to correct the subjective answers. Human evaluation can be better in handling subjective questions, which are subtle judgments, complex reasoning and expression of attitude. However, human evaluator used a lot of time, sensitivity and skill to evaluate the answers to get the feedbacks after a delay for scoring (William Horton, 2011). Inexperienced examiner would decrease the accuracy in evaluate the subjective answers and created a lot of redundant processes to correct the evaluation.

In order to save the time for evaluating the subjective answers, SQ&A System in an effective way to evaluate the subjective answers. Furthermore, the system helps the examiners in increasing the accuracy of

Corresponding Author: Maziani Sabudin, Universiti Sains Malaysia, School of Computer Sciences, 11800, USM, Pulau Pinang, Malaysia.
Ph: 6012 4958894 E-mail: maziani@cs.usm.my

Journal of Applied Science and Agriculture, 9(20) Special 2014, Pages: 23-28

evaluating the subjective questions by performing the algorithms. Besides that, the system also could help the lecturers to keep track of the students' records and this would indirectly help the lecturers understand more about the academic status of the students.

2. Literature Review:

Many research papers are carried out on a subjective question and answer system (Xuejun Liu *et al*., 2004), (Fanjin Mai *et al*., 2010), (Xiaoli Liu *et al*., 2007), (Chunli Mao *et al*., 2009), (Li Liu *et al*., 2010), (Alessandro Moschitti, 2003). According to (Fanjin Mai *et al*., 2010), research paper on Chinese subjective question scoring algorithm based on Natural Language Processing, a prototype that had been developed to solve Chinese subjective question and answer system.

There are several algorithms suggested by researchers, such as Forward Maximum Matching (FMM), semantic similarity computation and contrary degree calculation (Chunli Mao *et al*., 2009), in order to solve the problem faced by a SQ&A system.

FMM is a method that looks for the longest string in the text files and matches the first one until four characters of the string with a dictionary that contains thousands of items, which consist of single words with four characters. Once the match is found, the words will be taken out from the string as segment and the algorithm continue matching the next one until four characters of the string with the dictionary until the end of the string (Yan Niu *et al*., 2009). Strength of FMM is the accuracy of the word segmentation of this algorithm can go up to 95% and the algorithm is easy to achieve (Yan Niu *et al*., 2009).

However, the algorithm might lead to the longer words cannot be segmented correctly because the initial value of the maximum word-length is a constant. The longer words might also be matched repeatedly and this causes the waste of time (Wang Ruilei *et al*., 2011). The algorithm is more suitable to use to match Chinese words (Yan Niu *et al*., 2009), (Wang Ruilei *et al*., 2011).

Semantic Similarity Algorithm compares the sentences by using the union of set and semantic similarity of vectors (Yuhua Li *et al*., 2004), (Xu Liang *et al*., 2011). It involves a lot of formulae in the process of calculating the semantic similarity (Yuhua Li *et al*., 2004). Three classes of measures are used to identify the similarity between sentences. They are Word Overlap Measures, Term Frequency-Inverse Document Frequency (TF-IDF) Measures and Linguistic Measures (Palakorn Achananuparp *et al*., 2009). Linguistic measures are the best measures for identifying paraphrases in low-complexity data set. They are also performing well in high-complexity data set with certain condition (Palakorn Achananuparp *et al*., 2009). The order of the words is considered when evaluating the sentences (Yuhua Li *et al*., 2004).

Somehow, the performance between the classes of measures disappeared when the characteristic of the data had changed such as contained high degree of word overlap (Palakorn Achananuparp *et al*., 2009). A lot of improvements need to be done to make the algorithm can be applied in question answering system (Xu Liang *et al*., 2011). Words with similar meaning may appear more time after union of sets (Yuhua Li *et al*., 2004). Obviously, the formulae using are too hard to understand (Yuhua Li *et al*., 2004), (Xu Liang *et al*., 2011), (Palakorn Achananuparp *et al*., 2009).

The FMM algorithm is easy to achieve but it would be time consuming if the answer, which is going to be evaluated, is lengthy. Furthermore, the algorithm is not flexible, as the initial value of the maximum word-length had been set as constant which is immovable. This makes the algorithm not much reliable and usable as the complexity of the data increased.

The Semantic Similarity algorithm can process well in the low and complexity data. The characteristics of the data such as high degree of word overlap will destroy the performance gap of the between the measures (Palakorn Achananuparp *et al*., 2009). In addition, sentence pairs of data in high complexity because it did not serve like other measures tend to reject the dissimilar sentences pairs. The union of sets will cause the repetition of words which have similar meaning because the problem of singular and plural words (Yuhua Li *et al*., 2004).

Both of the algorithms were more used to evaluate the answers, which contained Chinese words. It will be time consuming because the complexity of the algorithm is high. Moreover, the algorithm did not do any modification on the answers before processing and evaluating the answers. Therefore, the effectiveness of both algorithms is low and some improvements need to be done to increase the effectiveness of the algorithms.

3. SQ&A System:

FMM algorithm evaluate every words in the answer and this will cause the time consuming because there will be 60% of words are high frequency words and sight words which are not necessary to be evaluated. Therefore, the answer should be gone through Stemming and Stop List algorithms to remove the high frequency words. The Semantic Similarity algorithm's performance will be destroyed if the answer given is complex because it tries to reject the words, which are not similar to the answer in the database. Therefore, comparison algorithm like Log-likelihood Ratio Summarization is used to compare the answer instead of rejecting the answer word by word. VSM is used to calculate the relevancy value for both user input and system answer. The value is then used to calculate the results of the answer.

Journal of Applied Science and Agriculture, 9(20) Special 2014, Pages: 23-28

To further enhance the accuracy and flexibility of the SQ&A System, two algorithms such as Hamming and LSI had been introduced. Hamming algorithm will allow the answer, which consists of one word and got one character different to pass the evaluation as long as the length of the answer is same as the modal answer. LSI helps to check the synonyms of the words before comparing the answers.

The architecture of the system is shown as the flow chart in Figure 1. Those algorithms are suitable to marking the subjective question in English. In addition, those algorithms will take less time to process the answer because they reduce the size of the answer by removing the high frequency words and sights word.

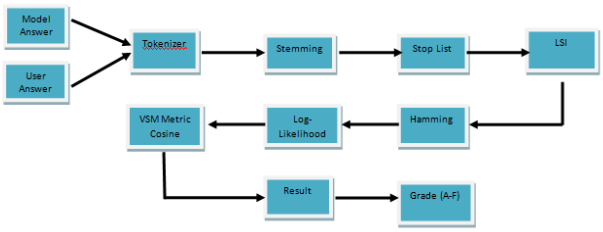


Fig. 1: Architecture Of SQ&A System.

The architecture of the system is shown as the flow chart in Figure 1. Those algorithms are suitable to marking the subjective question in English. In addition, those algorithms will take less time to process the answer because they reduce the size of the answer by removing the high frequency words and sights word.

A. Stemming and Stop List:

Stemming is the process of normalizing tokens by giving a standard form for tokens by looking for prefixes or suffixes and removing or occasionally rewriting them. In other words, stemming strips a string to its root word, which is called stem. This happens over a number of phases; the stemmer looks for and deals with one set of affixes, and when it's dealt with them, it looks for another set. At the end of this process, the part of the token that remains is called the stem. Stemming is one technique to provide ways of finding morphological variants of search terms. It is used to improve retrieval effectiveness and to reduce the size of indexing files. The usefulness of stemming depends on the morphological complexity that is, the number and frequency of prefixes and suffixes for a particular language. Because stemming algorithms generally run very fast, stemming is seen as an almost free way to boost search recall. Empirically, the effect of stemming on search is small, but it ultimately improves search by a measurable amount, especially on a smaller scale where recall becomes more important.

Stop List is use to filter the sentences where some of the words can be ignore. A Stop List is often used to omit high frequency words such as 'the', 'a', 'and' and so on. Ignoring the high frequency words can reduce the size of a finished concordance since they are appearing so often. Thus, after stemming and Stop List, some of the words, which are less useful, were ignored.

B. Latent Semantic Indexing (LSI):

LSI is an algorithm that finds the synonym of the words in a sentence. If the sentence of the answer did not contain any synonyms then this algorithm will not be executed. The algorithm also is applied in the Hamming algorithm which will be discussed at next section.

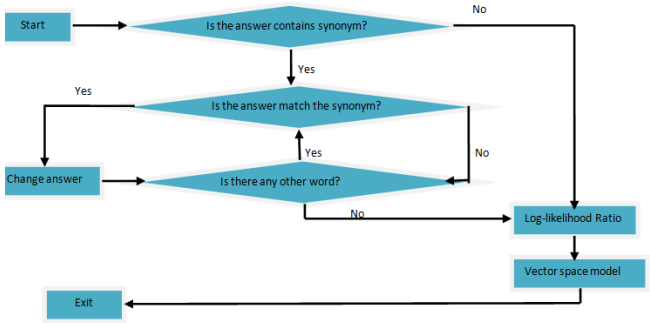


Fig. 2: Flow Chart Of LSI.

C. Hamming Algorithm:

Hamming algorithm is suitable in evaluating the one word answer. The Hamming algorithm is implemented as flow chart below. It begin with checking the number of word, if not it will go to another checking process

known as LSI which had been described in the previous part of this section, otherwise it will follow the steps as been described in the Figure 3.

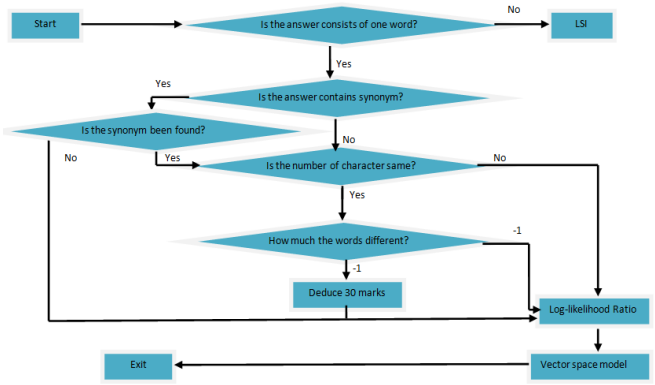


Fig. 3: Flow Chart of Hamming Algorithm.

D. Log-likelihood Ratio Summarization:

The Log-Likelihood ratio is the ratio of likelihood function varying the parameters over two different sets of the sentence. While its test is a statistical test for making a decision between two hypotheses based on the value of the ratio. Therefore, this algorithm is a good method for retrieving informative words. A word in this algorithm, generally, is called λ(w), that is the ratio between the probability of observing the word w, both in the input and in the background corpus assuming equal probabilities in both corpora, and the probability of observing w in both assuming different probabilities for w in the input and background corpus (Jurafsky D. *et al*., 2008). The summarization algorithm computes the weight for every sentence input by users and then ranks it accordingly to its score. Only top ranked sentences will retrieve and further process on it.

E. Vector Space Model (VSM):

VSM for information retrieval (William Horton, 2011) is presented as various vectors representing words in a collection of terms. The use of such representation is in order to plot vectors, which are solely based on the content of the document or query. With the use of such vector's levels of similarities between queries and documents could be then conjured. In SQ&A system, the user input will be translated into a vector denotes a point in that space. The system answer that are located close with the user's input can then be judged as being more relevant than sentences that are farther away. VSM will then output a relevancy value for both user input and system answer. The relevancy value will be taken and be converted to a score for the particular question. Lastly, student can keep track with their performance record and admin can also keep track with all the student performance.

4. System Analysis:

The results of the system had been compared with the results of the manual evaluation by using the following question and answer.

Question: How does testing fix into software quality assurance?

Model Answer: Quality assurance is a set of activities designed to evaluate the process by which software are developed or manufactured. Testing is one part of the software quality assurance. No defects will increase the confidence level and quality will be high.

The following table and graph show the result of comparing.

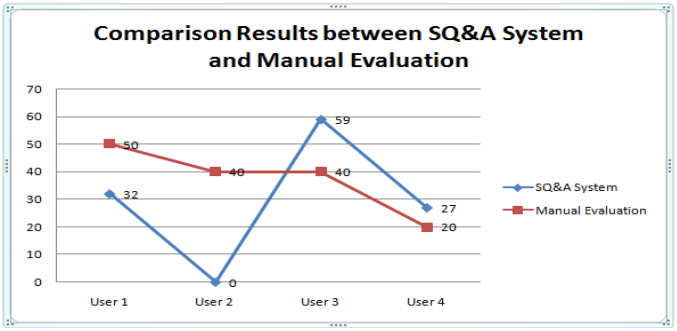
Table 1: Comparison Results Between SQ&A System And Manual Evaluation.

	SQ&A SYSTEM	Manual Evaluation
User 1	32	50
User 2	0	40
User 3	59	40
User 4	27	20

5. Discussion:

From Table 1 and Graph 1, it shows significant different between the manual evaluation and SQ&A system. This is due to the synonyms words in the dictionary which is saved in the database is not so much since the database is developed by the developer. Therefore, it is more suitable to be used as guideline evaluating the

subjective answer, which is entered by the users. The reliability of the system will be enhanced if the dictionary of database, which is used to detect the synonyms of words, is increased.



Graph 1: Comparison Results Between SQ&A System and Manual Evaluation.

6. Conclusion and Future Work:

Using the Hamming Algorithm and LSI had enhanced the SQ&A system. The flexibility of the answer evaluation had been increased once the two algorithms had been implemented. Somehow, there is a significant different between the SQ&A system with manual evaluation. Therefore, there is still got a lot of improvement space could be done in the future. The improvements are listed as below:

- Create the system in Android platform so that could be used for the commercial purpose.
- Enable the system connect to a device to scan the user input to evaluate the subjective question.
- Automate the LSI algorithm to find the synonyms without involvement of the human.

This paper is hoped to be a reference for future use on SQ&A system.

ACKNOWLEDGMENT

This research was supported by School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia.

REFERENCES

Alessandro Moschitti, 2003. Answer filtering via Text Categorization in Question Answering Systems, IEEE International Conference on Tools with Artificial Intelligence.

Chunli Mao, *et al*., 2009. Research on Answer Extraction Method for Domain Question Answering System (QA), International Conference on Computational Intelligence and Security.

Daniel Midgley, 2011. How to mark exam question. [Accessed 12 September 2012]. Available from: http://goodreasonblog.blogspot.com/2011/11/how=to=mark=exam=question.html.

Fanjin Mai, Xiaoguang Yue, Ziqiang Zhao, 2010. Research on Chinese Subjective Questions Scoring Algorithm Based on Natural Language Processing, Advanced Computer Control (ICACC).

Jurafsky, D., J.H. Martin, 2008. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice Hall.

Li Liu, Quan Qi, Fangfang Li, 2010. Ontology-based Interactive Question and Answering System, IEEE.

Palakorn Achananuparp, Xiaohua Hu, and Shen Xiajiong., 2009. The Evaluation of Sentences Similarity Measures. IEEE.

Roy Johnson, 1992. Marking Essays: Guidance Notes for Tutors. [Accessed 12 September 2012]. Available from: http://books.google.com.my/books?id=Zfo2t7iERbAC&pg=PT10&dg=marking+question&hl=zh-CN.

Wang Ruilei, Luan Jing, Pan Xiaohua, Lu Xiupei, 2011. An Improved Forward Maximum Matching Algorithm for Chinese Word Segmentation. CNKI.

William Horton, 2011. E-Learning by Design. [Accessed 13 September 2012]. Available from: http://books.google.com.my/books?id=YeBTJrW95KYC&pg=PA219&dg=subjective+question+and+answer+system&hl=zh-CN.

Xiaoli Liu, *et al*., 2007. Software architecture for a pattern based Question Answering system, IEEE.

Xuejun Liu, Yoshinao Aoki, Jianling Xu, 2004. Development and Evaluation of the Question and Answer System for Core Courses, IEEE.

Xu Liang, Dongjiao Wang and Ming Huang, 2011. Improve Similarity Algorithm based on VSM nd Its Application in Question Answering System. IEEE.

Yan Niu, Lala Li, 2009. An Improved Chinese Segmentation Algorithm Based on Segmentation Dictionary. IEEE.

Yuhua Li, Zuhair Bandar, David McLean and James O'Shea, 2004. A Method for Measuring Sentence Similarity and its Application to Conversational Agents. Intelligent Systems Group, Department of Computing and Mathematics.

Zhang Xiao Shu, Objective Test and Subjective Test [Online]. [Accessed 23 November 2012]. Available from: http://www.yanji.edu.sh.cn/teacher/wy/yy/zhangxiaoshu/newpage2.4.htm.