Implementing a question classification system to increase student engagement in higher education

Drew Costello, Vladyslav Falach, Maria Weishaar, Wang Yang
Supervised by Dr. Anand

Introduction

Current trends in technology and pedagogy have pushed institutions within higher education to adopt new methods emphasizing personalised teaching and learning. Methods utilised in a traditional lecture format foster increased anxiety towards verbal participation and limit student-instructor interaction (Trees et al. 2007). In addition, given the usual passive nature of lecture, it is difficult to tell if students are learning the material (Toto, 2009). To overcome these challenges, several institutions have adopted a flipped classroom model, where pre-recorded lectures are assigned to students as homework, leaving class time open for advanced concepts and collaborative learning (Schlairet et al., 2014). This model allows students to absorb course material at their own pace and adopt active learning in the presence of their peers and instructor (Trees et al. 2007) Studies have found that lecture slides for the video presentations often do not contain sufficient quantity of information for each student (Toto, 2009, Bishop et al., 2013). In a traditional lecture, misleading or confusing slides can be mended through class discussions originating from student inquiries (Bishop et al. 2013). To facilitate both participation within a traditional lecture setting, and inquiry formulation throughout video lectures, we propose a novel application.

Objective

The goal is to create an effective system that facilitate both participation within a traditional lecture setting, and inquiry formulation throughout video lectures. Through this application, students can input questions at any point during a lecture and rank their peers questions based on importance. The collected questions are then clustered into a set of related groups, in order to decrease redundancy and sort based on student perceived importance. This allows the instructor to gather organised, readable feedback from the students. This organizational system can help mitigate the negative effects of large classroom sizes on the instructor's ability to answer a large number of questions. It also allows the students to anonymously participate in class discussion within a traditional lecture and to ask questions to an instructor, throughout a video lecture. In addition, questions that students previously asked throughout the course and questions that the student ranked as important, are stored within their personal accounts. In regards the implementation methods, one of the main factors that motivated our choices is that the program needs to be able to produce results in near-real time.

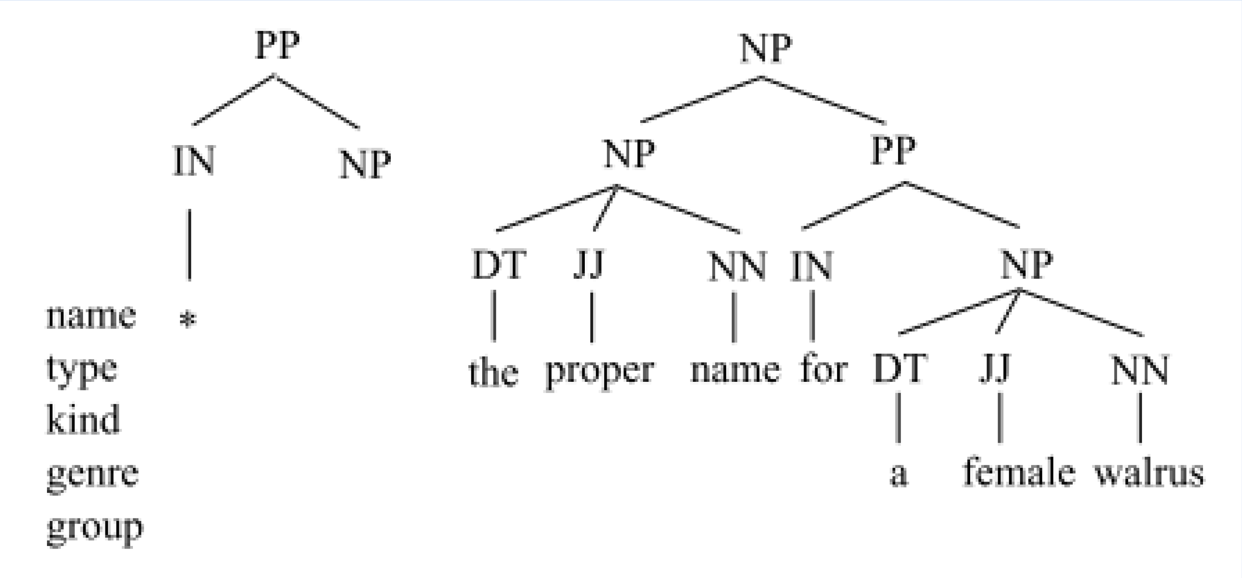

Figure 1. Example of a postfix syntax tree for a natural grammar, taken from (Thint et al., 2008).

Methods

Part of speech (POS) tagging is the process of applying tags, such as noun, verb, adjective, to individual words. One of the primary uses of tagging for our application is to minimize polysemy, which is the overlap of multiple different meanings for the same word. Tagging is performed via a perceptron neural network algorithm trained on the Penn treebank corpus. An example of this can be seen in Figure 2 (Daniel et al., 2016).

Word Similarity is a method of measuring semantic similarity between a pair of words. For this we will be relying on WordNet, which is a lexical database that stores words as a graph of related concepts. POS tagging is used to identify the specific instance of a word within WordNet. Similarity is measured as the distance between two words within the concept graph. Simple path distance normalized to the range [0,1] provides adequate results, given the time restrictions placed on our application. (Ferlež et al, 2004)

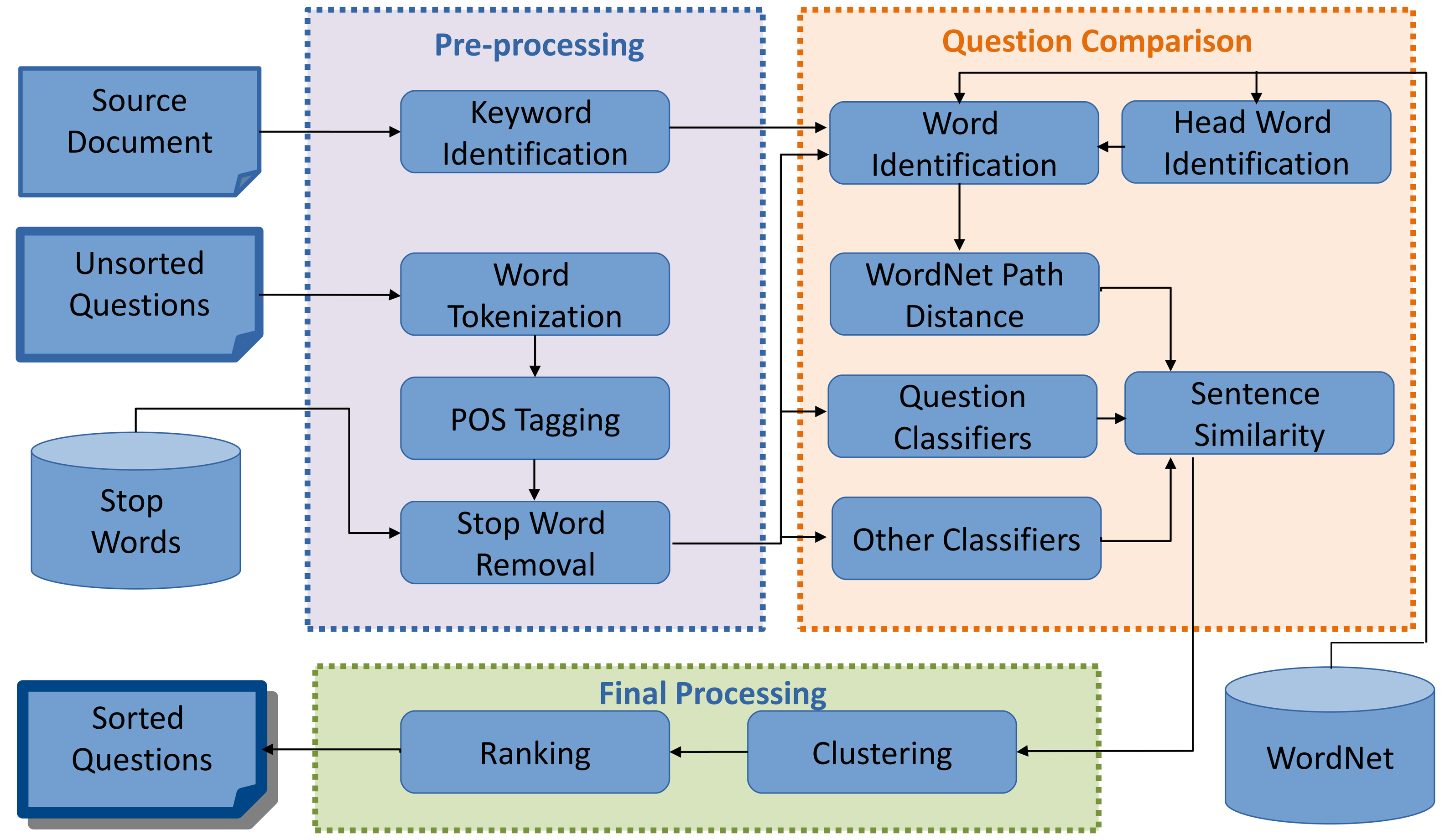

Figure 2. Simplified methodology diagram.

Keyword Extraction via Term Frequency-Inverse Document Frequency (TF-IDF) on a single document is utilized to identify words within the source text that describe a word unknown to WordNet such as an acronym. TF-IDF is the product of two factors. The first is the frequency of a word in a document, and the second is used to give a higher weight to words that only appear in a few documents (Jurafsky, 2017). For the purpose of this paper, words with a high TF-IDF value that are present in the slide containing the unknown term are used as a placeholder for the word similarity calculation (Jha et al., 2012).

Sentence Similarity between two sentences is calculated based on a word similarity matrix, which is constructed from all the pairwise combinations of words between the two sentences. This can be seen as a function which takes in a pair of sentences, and returns a single similarity value. An important feature of such a function in order to produce a similarity matrix suitable for clustering is that on comparing a sentence to itself, similarity is always 1, and that the order in which we input the pair does not matter. (Jha et al., 2012)

Question classification is the process of assigning a set of questions to a subset of semantic categories based on the expected answer. The question category criteria used is a two-layered taxonomy, containing 6 coarse classes and 50 fine classes (Sangodiah et al., 2015). The question is assigned a coarse and fine class based on the interrogative and head words that are present in the sentence, which are identified through a set of parser rules. The shortest path within the WordNet hierarchy connecting the head word to a fine class word assigns a fine class to each questions (Thint et al., 2008).

Clustering is done based on the sentence similarity matrix that was calculated in the previous steps. The similarity matrix for n questions can be seen as an n-dimensional vector space, where each dimension is similarity to a certain question. Because of this, we can use k-means clustering. This method is relatively fast and gives good results given the input of a similarity matrix. The number of clusters can either be given, or calculated automatically (Jurafsky, 2017).

Results

A simplified implementation of the intended application was produced. Questions were collected in a number of different class lectures, with subjects ranging from computer science to biology. Collection was done through student written questions, as well as questions recorded from verbal participation during lecture. These questions form the sample dataset on which the model was tested. To validate the performance of algorithm, an online survey was conducted. The survey asked participants to evaluate their perceived similarity of pairs of questions. The ranking was based on the scheme proposed at Semeval 2014 (Albertross et al. 2016), which can be seen in Figure 3 . This was done in order to ensure the test ranking values are unbiased and clear to participants. The similarity matrix produced by human participants can then be compared with the similarity matrix produced by the program. Due to time constraints, the study remains ongoing.

0 - The two items do not mean the same thing and are not on the same topic.
1 - The two items describe dissimilar concepts, ideas and actions, but might be likely found together in a longer document on the same topic.
2 - The two items have dissimilar meaning, the share concepts, ideas, and actions in the smaller text are related (but not similar) to those of the larger text.
3 - The two items share many of the same important ideas, concepts, or actions, but those expressed in the smaller text are similar but not identical to the most important in the larger text.
4 - The two items have very similar meanings and the most important ideas, concept, or actions in the larger text are represented in the smaller texta

Figure 3. Ranking scheme proposed at Semeval 2014 (Albertross et al., 2016)

Conclusions

There are a number of potential improvements that can be studied, which may increase the accuracy of the model. The primary one being research into implementing additional classifiers for the model, such as classifiers to measure the essential terms required to convey the meaning of a question (Khashabi, 2017). There is additional semantic and statistical data that we can obtain from the questions, which is not currently taken into account since the training dataset for this purpose is not yet available. In terms of pedagogical applications and future research, it is possible for the program to be integrated with video lecture systems as an additional feature. In addition, it is also essential that the survey validation experiment is conducted with an audience that has a high level of familiarity with the topics involved to reduce the noise in the data.

References

Albertsson, S. et al., 2016. Similarity-Based Alignment of Monolingual Corpora for Text Simplification Purposes. , pp.154–163.

Daniel, James, H., & Martin. (2016). Part-of-Speech Tagging. Speech, Language Processing, (Chapter 12). Retrieved from https://web.stanford.edu/~jurafsky/slp3/10.pdf

Ferlež, J., & Gams, M. (2004). Shortest-path semantic distance measure in WordNet v2.0. Informatica (Ljubljana), 28(4), 381–386.

Jha, S., Schwartz, H., & Ungar, L. (2012). Penn: using word similarities to better estimate sentence similarity. The First Joint Conference on Lexical Anf Computational Semantic (*SEM), 679–683. Retrieved from http://dl.acm.org/citation.cfm?id=2387751

Khashabi, D. & Roth, D., 2017. Learning What is Essential in Questions. Proceedings of the 21st Conference on Computational Natural Language Learning, pp.80–89.

Trees, A.R. & Jackson, M.H., 2007. The learning environment in clicker classrooms: Student processes of learning and involvement in large university-level courses using student response systems. *Learning, Media and Technology*, 32(1), pp.21–40.

Schlairet, M.C., Green, R. & Benton, M.J., 2014. The Flipped Classroom. *Nurse Educator*, 39(6), pp.321–325. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00006223-201411000-00017.

Toto, R., 2009. Flipping the Work Design in an industrial engineering course. *2009 39th IEEE Frontiers in Education Conference*, pp.1–4. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5350529.

Lowell, J. et al., 2013. The Flipped Classroom : A Survey of the Research The Flipped Classrom : A Survey of the Research. *Proccedings of the Annual Conference of the American Society for Engineering Education*, p.6219.

Jurafsky, D. & Martin, J.H., 2017. Speech and Language Processing - An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Available at: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf.

Sangodiah, A., Muniandy, M. & Heng, L.E., 2015. Question classification using statistical approach: A complete review. *Journal of Theoretical and Applied Information Technology*, 71(3), pp.386–395.

Thint, M. & Qin, Z., 2008. Question Classification using Head Words and their Hypernyms. , (October), pp.927–936.

Khashabi, D. & Roth, D., 2017. Learning What is Essential in Questions. *Proceedings of the 21st Conference on Computational Natural Language Learning*, pp.80–89.