

Text Classification for Exam Questions

Introduction:

Text classification is a widely studied problem in natural language processing (NLP) and is applied in a variety of domains such as spam detection, sentiment analysis, and content recommendation systems. In this proposal, we aim to investigate text classification for exam questions from books. The goal is to automatically classify exam questions into different categories based on the topics and subtopics covered in the textbook. This project is intended for an Information Retrieval course at the college level.

Problem Statement :

The problem we address in this project is the difficulty students face in preparing for exams. Typically, textbooks have a vast amount of information, and it can be challenging for students to identify the most important topics and subtopics that are likely to appear on exams. Therefore, our goal is to build a system that can automatically classify exam questions from a textbook based on their content.

Importance of the Problem :

Automatically classifying exam questions can save time and effort for students who need to prepare for exams. With an automated system, students can quickly identify the most important topics and subtopics covered in a textbook, and focus their studying efforts on those areas. This can lead to more efficient studying and better exam performance. Furthermore, the proposed project can serve as a practical application of text classification techniques for students in an Information Retrieval course.

Related Work :

Several studies have investigated the classification of text data, including exam questions. For instance, in a recent study by Xie et al. (2021), they proposed a method to classify exam questions based on the learning objectives and Bloom's Taxonomy. In another study by Wang et al. (2019), they used an ensemble learning approach to classify exam questions into categories based on the difficulty level.

Our Idea :

Our proposed approach differs from previous work by focusing on the classification of exam questions based on the topics and subtopics covered in the textbook. We plan to use topic modeling techniques, such as Latent Dirichlet Allocation (LDA), to identify the topics and subtopics in the textbook. We will then train a classifier, such as a Support Vector Machine (SVM), to classify exam questions based on the topics and subtopics identified in the textbook.

Techniques/Algorithms :

We plan to use topic modeling techniques, such as Latent Dirichlet Allocation (LDA), to identify the topics and subtopics covered in the textbook. We will use a pre-trained

word embedding model such as GloVe to represent the exam questions as vectors. Then, we will train a classifier such as a Support Vector Machine (SVM) or a Random Forest (RF) on the vector representations to classify the exam questions.

Evaluation :

We plan to evaluate our approach using a dataset of exam questions from textbooks. We will use precision, recall, and F1-score to measure the performance of our classifier. We will compare our results with the state-of-the-art methods for exam question classification to demonstrate the effectiveness of our approach.

Potential Contributions :

Our proposed project can serve as a practical application of text classification techniques for students in an Information Retrieval course. Furthermore, our approach can help students save time and effort in preparing for exams by automatically identifying the most important topics and subtopics covered in a textbook. The results of our evaluation can also provide insights into the effectiveness of different text classification techniques for the classification of exam questions.

Conclusion :

In this proposal, we have presented a project idea to investigate text classification for exam questions from books. We have identified the problem of students facing difficulties in preparing for exams and the importance of building a system that can automatically classify exam questions. We have reviewed related work and presented our proposed approach using topic modeling techniques and classifiers. Finally, we have discussed our evaluation plan and potential contributions of this work.

References :

1. Xie, T., Zhang, L., Chen, J., & Chen, L. (2021). Exam Question Classification Based on Learning Objectives and Bloom's Taxonomy. *IEEE Access*, 9, 4780-4790.
2. Wang, D., Yang, Z., Gao, F., Liu, Y., & Zhang, Z. (2019). Exam Question Classification Based on Ensemble Learning. *Proceedings of the 2019 International Conference on Artificial Intelligence and Education (ICAIE)*.