International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Question retrieval via a diverse social influence network for community-based question answering

By:
Uday kumar. BK
Mohana vamsy. A
Anil kumar. VS

A B S T R A C T

community-based question-answering platforms have attracted a large number of users who want to share their knowledge and learn from one another. As the number of community-based question answering (CQA) platforms grows, so does the number of overlapping questions, making it difficult for users to choose the right resource. It is critical that we implement good automated algorithms to reuse historical queries and answers. We focus on the subject of question retrieval in this study, which tries to match relevant or semantically equivalent historical questions to immediately resolve one's inquiry. The lexical gaps between questions for the word ambiguity and word mismatch problem pose a barrier in this job. Furthermore, the restricted number of words in the query sentences results in a lack of word characteristics. To address all of these issues, we suggest a HSIN is a new framework that enhances question embedding performance by encoding not only the question contents but also the asker's social interactions. To match the similarities between the asker's inquiry and previous inquiries supplied by other users, we employ a random walk based learning method using a recurrent neural network. Extensive tests on a large-scale dataset from the real-world CQA site Quora show that using heterogeneous social network information beats other state-of-the-art solutions in this task.

KEYWORDS CQA

Question retrieval Deep learning Social network

Introduction:

Users can submit their puzzles and share their knowledge through community-based question answering (CQA) services. CQA sites such as Yahoo! Answers, Baidu Knows, Wiki Answers, Zhihu, and Quora have amassed significant question-answer pairs throughout the years. However, a huge number of recommended questions are excessively overlapping and redundant, reducing the query efficiency of users. For many years, scholars have worked in the fields of question retrieval, question answering, expert finding, and natural language processing to efficiently automate the selection of appropriate references from large-scale pre-queried questions with corresponding replies.The domain of question retrieval is the subject of this paper.

The most difficult difficulty in question retrieval is assisting users in retrieving historical questions that are semantically related or relevant to their current questions. Before deciding whether or not to ask a new inquiry, users can look at the good matches. The feature provides users with a great deal of ease while also lowering the repCQA platforms have a high etition rate. As a result, it is critical for CQA services to provide timely and accurate results. This task has been the subject of numerous investigations. However, issues persist due to lexical gaps between questions created by word ambiguity and the problem of word mismatch. For instance, there are two questions on the Quora site: "What are some decent basic textbooks on machine learning?" and "How can I begin learning machine learning?" These two inquiries are semantically related and express the same meaning in the eyes of our human readers. While the major stream models used in question retrieval share a few common words, these two questions may generate a mismatching difficulty. Even for the same term, it's possible.produce ambiguity, for example, when we say the word 'apple,' we can't know if we're talking about the apple firm or the apple fruit unless we apply context information to categorise it. The feature sparsity is- sue is another problem in question retrieval. Because question titles are typically short and contain a variety of irregular noise, it is difficult to obtain a precise modelling topic from the whole information of questions.The majority of extant research treats the question retrieval problem as a supervised learning approach, in which an evaluating model is trained using both the question textual content and its belonging category. The language model is primarily used by researchers in the field of question retrieval to learn the semantic representation of questions.the contents of the question Although existing question retrieval methods have achieved good results, they do not entirely address the word sparsity bottleneck and do not properly utilise questions side information such as the asker's history, which is crucial for question interpretation. Furthermore, because askers have their own social networks and their interests may overlap with those of their friends, it is reasonable to predict one scenario: individuals may pose questions that are similar to those of their friends. It's a typical occurrence among classmates and coworkers. As a result, how to make use of the available social data is crucial for the question retrieval task.

The textual con- tents of questions are required for question retrieval tasks, in addition to the valuable social information. Recent research in the field of question retrievalTo learn semantic representations, CQA data uses a variety of retrieval models, including the language model, the translation model, and the learning-to-rank model. These prior studies have repeatedly demonstrated the viability of retrieval performance. Traditional hand-crafted features such as bag-of-words, on the other hand, are bound to have difficulty with properly embedding the word sequence of queries. Various embedding approaches are presented for learning the semantics of comparable words and encoding the word sequence into low-dimensional continuous embedding space, inspired by the flourishing of deep learning application in natural language processing. Recurrent neural networks are an excellent choice for learning the semantic representation because the question contents are constantly consecutive data of varying duration.We propose a new framework called HSIN in this paper (Heterogeneous Social Influential Network). In particular, we use a random walk method to uncover useful side information from heterogeneous social network and question category data. In addition, we use a recurrent neural network to simulate the textual content of the inquiry. To express the query and rate similarities with historical questions, we concatenate the question textual content with user embedding. The questions textual content, their linked categories information, and the askers social information are all simultaneously learned in our proposed HSIN framework, allowing us to take use of the extensive interactions between CQA data and users data. When a user asks a new question, HSIN can rank the previously proposed related questions so that users can refer to them.Without having to wait for his own inquiry to be addressed, he was given suggestions for queries and responses.

It's worth mentioning a few of our work's contributions here:

- We provide HSIN, a novel framework for integrating question textual content with information from the asker's social network. To learn the semantic representation of queries and users at the same time, we use a random deep walk method with a recurrent neural network.

- Unlike earlier studies, our suggested methodology in the field of question retrieval leverages the semantic representation of questions and the rich heterogeneous social network information. Because it is scalable for heterogeneous network learning, the framework can be applied to various fields of information retrieval.

- Our suggested framework outperforms state-of-the-art methods that rely just on question textual data.The performance of our concept of incorporating rich social network side information improved greatly in question retrieval ranking, demonstrating the potential of our concept.

Existing question retrieval methods can be divided into four categories: categories- model-based approaches, translation-model-based approaches, topic-modeling-based approaches, and neural network-based approaches.

The first technique is the one that most people think of when they're trying to figure out how to solve a problem like question retrieval. It takes into account the metadata of questions by considering question categories and labels. Can et al. developed three language models that use question categories smoothing to estimate question similarity within the same category. Several methods for utilising category side-information were proposed by Zhou et al. They used user-selected categories to filter irrelevant queries under leaf categories in. They developed group non-negative matrix factorization in by learning both category-specific and shared themes for each category.Throughout all categories Zhou et al. used a fisher kernel to combine variable-length word embedding vectors into fixed-length word embedding vectors.

As a result, a continuous word em- bedding model was learned.