Chapter 5
A Yes/No Answer Generator Based on Sentiment–Word Scores in Biomedical Question Answering

Mourad Sarrouti

Sidi Mohammed Ben Abdellah University, Morocco

Said Ouatik El Alaoui

Sidi Mohammed Ben Abdellah University, Morocco

ABSTRACT

Background and Objective: Yes/no question answering (QA) in open-domain is a longstanding challenge widely studied over the last decades. However, it still requires further efforts in the biomedical domain. Yes/no QA aims at answering yes/no questions, which are seeking for a clear "yes" or "no" answer. In this paper, we present a novel yes/no answer generator based on sentiment-word scores in biomedical QA. Methods: In the proposed method, we first use the Stanford CoreNLP for tokenization and part-of-speech tagging all relevant passages to a given yes/no question. We then assign a sentiment score based on SentiWordNet to each word of the passages. Finally, the decision on either the answers "yes" or "no" is based on the obtained sentiment-passages score: "yes" for a positive final sentiment-passages score and "no" for a negative one. Results: Experimental evaluations performed on BioASQ collections show that the proposed method is more effective as compared with the current state-of-the-art method, and significantly outperforms it by an average of 15.68% in terms of accuracy.

INTRODUCTION

The large size of literature in the biomedical domain makes it difficult for information seekers even in their field of interest to find the information they need (McDermid, Kristjanson, & Spry, 2010). The most used for accessing to biomedical information are information retrieval (IR) systems, such as

PubMed[1] which gives access to the MEDLINE[2] biomedical bibliographic database (Hristovski, Dinevski, Kastrin, & Rindflesch, 2015). Indeed, finding sufficient and short precise answers is a challenging task for classical IR systems (Wren, 2011). In addition, in classical IR systems, the users have often to deal with the burden of studying and filtering the returned citations of their queries so as to find the precise information they were looking for. Therefore, to minimize searching and browsing time while maximizing the usefulness of that knowledge is a growing interest for biomedical question answering systems (Bauer & Berleant, 2012). Question answering (QA) regards a sophisticated form of IR characterized by information needs that are expressed as natural language statements or questions (Wren, 2011). It aims at providing inquiries with specific pieces of information as an answer, by automatically analyzing thousands of articles, ideally, in less than a few seconds. Typically, an automated QA system consists of three main processing phases, which can be studied and developed independently (Athenikos & Han, 2010; Cao et al., 2010; Neves & Leser, 2015): (1) question processing, (2) document processing, and (3) answer processing. Figure 1 illustrates the generic architecture of a biomedical QA system.

Given an input biomedical question, the question is first handed over to the question processing phase. The latter consists of the following components: (a) question analysis for extracting some useful information such as biomedical entity names, and semantic relationships; (b) question classification for identifying the answer format and the topic (Cao et al., 2010; Patrick & Li, 2012; Roberts et al., 2014; Lopes et al., 2014; Sarrouti et al., 2015); (c) query formulation for constructing IR-style query by transforming the question into a canonical form. The output of this phase is an appropriate query which is used as input to document processing, the second phase. An IR system is normally used to retrieve the relevant documents (Sarrouti & Alaoui, 2016). Then, passages are extracted which serve as answer candidates as well as an input to the last phase, answer processing; in this phase, the system uses an appropriate answer extraction algorithm to estimate the qualities of the candidate answers. Finally, the top-ranked candidate answers and the raw texts from which the answers were extracted are shown to the user (Sarrouti & Ouatik, 2017).

In spite of the importance of the process, yes/no-style QA has not been intensively studied in the biomedical domain, compared to other types of QA, such as the factoid-style QA (Weissenborn et al., 2013; Yang et al., 2015; Hristovski et al., 2015) and the definition-style QA (Lee et al., 2006; Yu et al., 2007; Cao et al., 2011). Yes/no QA aims at answering yes/no questions, which are seeking for a clear "yes" or "no" answer (Green & Carberry, 1994). For instance, a biomedical yes/no QA system would provide "yes" as an answer for the yes/no question "Does the histidine-rich Ca-binding protein (HRC) interact with triadin?". Even though there are only two possible answers, "yes" or "no," such questions can be quite hard to answer due to the complicated sentiment analysis process of the candidate answers. This challenging task has motivated the recent launch of the BioASQ[3] challenge: an EU-funded support action to set up challenges on biomedical semantic indexing and QA (Tsatsaronis et al., 2015). The BioASQ challenge consists of two main tasks, i.e., task a: large-scale online biomedical semantic indexing and task b: biomedical semantic QA. The major goal of task b is to assess the performance of participating systems in different stages of the QA process. Moreover, task b is sub-divided into two phases: phase A and phase B. Indeed, in phase B, BioASQ provided researchers with biomedical questions and their respective question types (i.e., yes/no, factoid, list, summary) as well as gold-standard information for phase A, such as manually curated relevant concepts, documents, passages and RDF triples (Balikas et al., 2013). The participants were requested to provide an exact answer: "yes" or "no" for yes/no questions, a short paragraph for summary questions, and a single or a list of short answers for factoid and list questions, respectively.

Related Content

Loan Fraud Detection Using Machine Learning as a Data Mining Approach
Nabila Hamdoun (2022). *International Journal of Data Analytics (pp. 1-10).*
www.irma-international.org/article/loan-fraud-detection-using-machine-learning-as-a-data-mining-approach/309096

Sustainable Parking Planning for Smart Growth of Metropolitan Cities
Ozge Yalciner Ercoskunand Ebru Vesile Ocalir Akunal (2018). *Intelligent Transportation and Planning: Breakthroughs in Research and Practice (pp. 117-136).*
www.irma-international.org/chapter/sustainable-parking-planning-for-smart-growth-of-metropolitan-cities/197130

Exploratory Data Analysis
 (2018). *Spatial Analysis Techniques Using MyGeoffice® (pp. 112-136).*
www.irma-international.org/chapter/exploratory-data-analysis/189720

Importance of Big Data and Hadoop in E-Servicing
Karthiga Shankarand Suganya R. (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications (pp. 1634-1644).*
www.irma-international.org/chapter/importance-of-big-data-and-hadoop-in-e-servicing/291056

Wearable Devices Data for Activity Prediction Using Machine Learning Algorithms
Lakshmi Prayaga, Krishna Devulapalliand Chandra Prayaga (2019). *International Journal of Big Data and Analytics in Healthcare (pp. 32-46).*
www.irma-international.org/article/wearable-devices-data-for-activity-prediction-using-machine-learning-algorithms/232334