Answering Comparative Questions with Web-based Arguments

Alexander Bondarenko
Martin-Luther-Universität
Halle-Wittenberg

Matthias Hagen
Martin-Luther-Universität
Halle-Wittenberg

Meriem Beloucif
Universität Hamburg

Chris Biemann
Universität Hamburg

Alexander Panchenko
Skolkovo Institute
of Science and Technology

Abstract

Question answering platforms such as Yahoo! Answers or Quora contain many comparative questions. Daily questions about comparisons range from choosing what to wear or eat, to more important decisions like where to best study or what library to use for software projects. Surprisingly, search engines or other automatic systems are not yet good at supporting answers to such questions with arguments and explanations. We propose CAM (comparative argumentative machine), an open-domain information retrieval system to argumentatively compare options using information extracted from the Common Crawl. In our empirical study, the CAM users were 15% more accurate and 20% faster at answering comparative questions of the type "How does X compare to Y with respect to Z?" than when using standard keyword-based search. In public platforms, a good (men-made) answer needs to be based on objective expert level argumentation / explanation why to favour one of the choice candidates. However, there currently is no "automatic" retrieval-based model that would be able to satisfy comparative information needs in a general domain with sufficient coverage and explanations. Web search engines, while being able to directly answer many factoid questions, do not treat comparative questions any special but simply return "ten blue links". In this extended abstract of our CHIIR 2019 publication (Schildwächter et al. 2019), we present CAM[1] a system aimed at general domain comparisons with argumentative explanations.

To ensure a wide coverage, our CAM system compares two objects based on argumentative structures extracted from a web-scale Common Crawl index. The CAM system retrieves sentences stating that one of the compared objects is superior to the other, that they are equal, or that they are not comparable. A comparison of two objects o and o' in the CAM sense is formally defined as "o $[> \mid < \mid = \mid \neq]$ o' w.r.t. $a_i, \ldots, a_j \in A$", where $A = \{a_1, \ldots, a_k\}$ is the set of comparison aspects of o and o'; for example, "Python (o) is better than ($>$) Matlab (o') for web development (a_i)." The five CAM system components are: sentence (answer) retrieval, sentence classification, sentence ranking, aspect extraction, and the user interface. For the **sentence retrieval**, CAM uses an Elasticsearch index of the Common Crawl-based DepCC containing 14.3 billion linguistically pre-processed English sentences. On a user input, sentences matching the input objects and comparison aspect(s) are retrieved. In subsequent **sentence classification**, a classifier distinguishes between four classes of support sentences: the first object from the user input is better / equal / worse than the second one ($>, =, <$) w.r.t. a comparison aspect, or no comparison is found ($\neq$). The classifier exploits the text between the objects to identify the comparison polarity. The **sentence ranking** scores sentences by combining their classifier's confidence and the Elasticsearch retrieval score; the CAM output is selected by summing up the scores of all sentences supporting and attacking the statement. In **aspect identification**, CAM also generates up to ten supplementary aspects, even when no comparison aspect is provided by the user. We use three different methods for aspect mining: (1) searching for comparative adjectives and adverbs; (2) searching for phrases with comparative adjectives, adverbs and prepositions; (3) searching for specific hand-crafted patterns to match expressions like "because of higher speed" or "reason for this is the price". Finally, **the user interface** consists of a question input form and an answer presentation component. The input form allows to enter two compared objects and their comparison aspects. The answer presentation displays the sentences retrieved from the Common Crawl for the user input, summarizes the sentences in an overall decision support score, and gives aspect-specific score bars to show the distribution for the individual user-specified aspects.

We conducted two user studies, where we compared our new CAM system to a keyword-based search. The 23 participants were prompted to answer 34 different comparative questions extracted from Quora using CAM or the keyword-based search (random shuffle of which user was prompted to use CAM on what topic; no user did one topic with both systems). The results show that the participants' CAM-based answers were 15% more accurate and 20% faster.

[1]Demo, API & code: `http://ltdemos.informatik.uni-hamburg.de/cam/`

References

Schildwächter, M.; Bondarenko, A.; Zenker, J.; Hagen, M.; Biemann, C.; and Panchenko, A. 2019. Answering comparative questions: Better than ten-blue-links? In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, CHIIR 2019, Glasgow, Scotland, UK*, 361–365.