Towards Understanding and Answering Comparative Questions

February 21–25, 2022

Alexander Bondarenko\textsuperscript{1}  
Yamen Ajjour\textsuperscript{1}  
Valentin Dittmar\textsuperscript{1}  
Niklas Homann\textsuperscript{1}  
Pavel Braslavski\textsuperscript{2,3}  
Matthias Hagen\textsuperscript{1}

\textsuperscript{1}Martin-Luther-Universität Halle-Wittenberg,  \textsuperscript{2}Ural Federal University,  \textsuperscript{3}HSE University

www.webis.de

Fifteenth ACM International Conference on Web Search and Data Mining (WSDM ’22); Virtual Event, Tempe, AZ, USA.
Towards Understanding and Answering Comparative Questions

Motivation

- Simple comparisons: “Did Messi or Ronaldo score more goals in 2021?”
- Life-changing and highly subjective: “Is it better to move abroad or stay?”
- For big decisions, 80% of Americans rely on online research [Turner & Rainie; 2020].
- 3% of search engine’s questions are comparative [Bondarenko et al.; WSDM'20].
- 50% of these comparative questions are non-factual [Bondarenko et al.; WSDM'20].
Towards Understanding and Answering Comparative Questions

Motivation

- Simple comparisons: “Did Messi or Ronaldo score more goals in 2021?”
- Life-changing and highly subjective: “Is it better to move abroad or stay?”
- For big decisions, 80% of Americans rely on online research [Turner & Rainie; 2020].
- 3% of search engine’s questions are comparative [Bondarenko et al.; WSDM’20].
- 50% of these comparative questions are non-factual [Bondarenko et al.; WSDM’20].

Contributions

- Dataset: comparative questions with objects, aspects, and answers’ stances.
- Classifiers for comparative and subjective comparative questions.
- Identifying objects, aspects, and predicates.
- Stance detector for answers.

https://github.com/webis-de/WSDM-22
Is a cat or a dog a better friend?

Pro obj. 1: Cats can be quite affectionate and attentive, and thus are good friends.

Pro obj. 2: Cats are less faithful than dogs.
Is a cat or a dog a better friend?

Pro obj. 1: Cats can be quite affectionate and attentive, and thus are good friends.

Pro obj. 2: Cats are less faithful than dogs.

- 31,000 questions, 3,500 comparative, 1,690 subjective from MS MARCO, Google Natural Questions, Quora, Stack Exchange.
- 950 answers (text passages) with 4 stance labels from Stack Exchange.

Source: https://pixabay.com/images/id-2606759/
Towards Understanding and Answering Comparative Questions

Comparative Question Classification

- Cascading ensemble recalls 71% of comparative questions at prec. of 1.0.
  1. 10 rules: e.g., “Is a cat or a dog a better friend?” Recall 54%.
  2. Feature-based: Logistic regression with word 4-grams Recall 62%.
  3. Neural: RoBERTa, BART, SBERT for representations + DNN Recall 69%.
  4. Averaging the classifiers’ decision probabilities Recall 71%.

- Operating points (probability thresholds) chosen for precision of 1.0.

- Remove comparative questions after each classifiers’ group: more sophisticated classifiers for more difficult cases.

- 10-fold cross-validation.
Towards Understanding and Answering Comparative Questions

Parsing Comparative Questions

Direct:  
Is a **cat** or a **dog** a **better** **friend**?

Indirect:  
What **pet** is the **best** **friend**?

Without aspect:  
Who is **better**, a **cat** or a **dog**?
Towards Understanding and Answering Comparative Questions

Parsing Comparative Questions

Direct:  
Is a cat or a dog a better friend?

Indirect:  
What pet is the best friend?

Without aspect:  
Who is better, a cat or a dog?

- 10-fold cross-validation.
- Baseline: BiLSTM with 300-dimensional GloVe embeddings [Arora et al.; CIKM’17].

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Object</th>
<th>Aspect</th>
<th>Predicate</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiLSTM</td>
<td>0.80</td>
<td>0.52</td>
<td>0.85</td>
<td>0.98</td>
</tr>
<tr>
<td>RoBERTa</td>
<td>0.93</td>
<td>0.80</td>
<td>0.98</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Towards Understanding and Answering Comparative Questions
Answer Stance Detection

Is a **cat** or a **dog** a better friend?

object 1 object 2

Pro obj. 1: *Cats can be quite affectionate and attentive, and thus are good friends.*

- 4 labels: pro object 1, pro object 2, neutral, no stance.
- RoBERTa and Longformer for representations + DNN and logistic regression.
- RoBERTa and Longformer with sentiment prompts.
- Masking comparison objects.
Towards Understanding and Answering Comparative Questions

Answer Stance Detection

Is a OBJECT 1 or a OBJECT 2 a better friend?

Pro obj. 1: OBJECT 1 can be quite affectionate and attentive, and thus are good friends.

- Most effective classifier RoBERTa.
- Identifying subjective questions: F1 0.95.
- Comparison objects are masked in questions and answers.
- Add a sentiment prompt: OBJECT 1 is better.
- Input: OBJECT 1 is better [SEP] ANSWER.
- Highest accuracy on 4 labels (pro object 1 / 2, neutral, no stance) 0.63.
Towards Understanding and Answering Comparative Questions

Conclusions

- Dataset: comparative questions with objects, aspects, and answers’ stances.
- Classifiers for comparative questions, objects, aspects, and predicates.
- Stance detector for potential answers.

11 © Alexander Bondarenko 2022
Towards Understanding and Answering Comparative Questions

Conclusions

- Dataset: comparative questions with objects, aspects, and answers’ stances.
- Classifiers for comparative questions, objects, aspects, and predicates.
- Stance detector for potential answers.

Future Work

- Matching comparison objects in questions and answers.
- Improving the stance detection of comparative answers.
Towards Understanding and Answering Comparative Questions

Conclusions

- Dataset: comparative questions with objects, aspects, and answers’ stances.
- Classifiers for comparative questions, objects, aspects, and predicates.
- Stance detector for potential answers.

Future Work

- Matching comparison objects in questions and answers.
- Improving the stance detection of comparative answers.

SIGIR Student Author Registration Award.

thank you!
References

