# Understanding Comparative Questions and Retrieving Argumentative Answers

The oral exam of **Alexander Bondarenko** 

Friedrich-Schiller-Universität Jena

June 29, 2023



Should I put on a blue or a white shirt for my defense?



Should I put on a blue or a white shirt for my defense?

Maybe a "small" decision ...



Is buying a house better than renting?

Big, life-changing decision ...



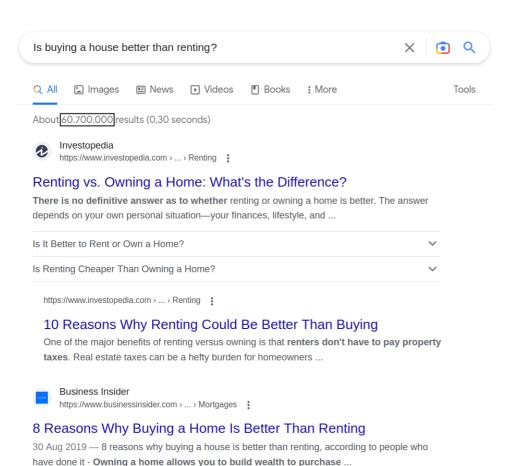
Is buying a house better than renting?

Big, life-changing decision ...

Decision making involves:

Collecting information about the alternatives.

Weighing pro and con arguments.

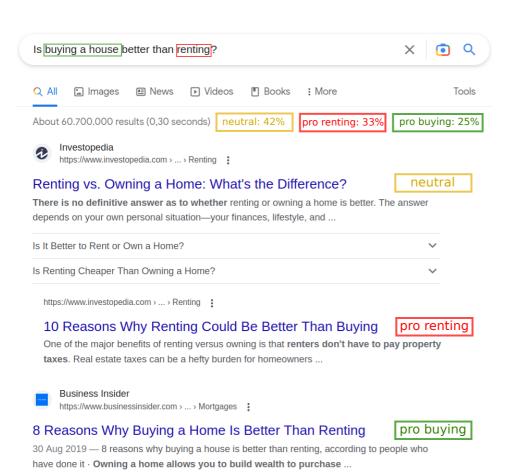


□ Decision making involves:

Collecting information about the alternatives.

Weighing pro and con arguments.

□ 80% prefer online research over asking friends [Turner and Rainie; Pew Research Center 2020].



□ Decision making involves:

Collecting information about the alternatives.

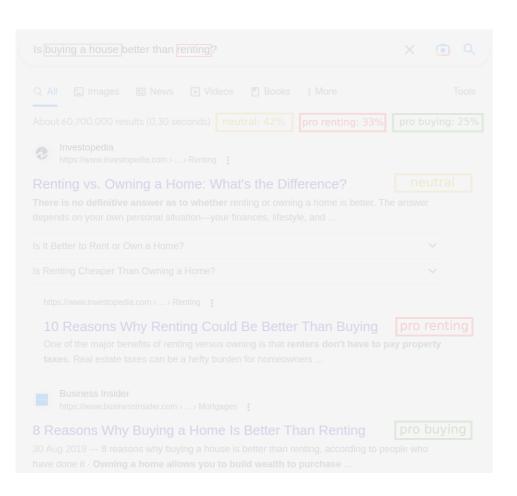
Weighing pro and con arguments.

- 80% prefer online research over asking friends [Turner and Rainie; Pew Research Center 2020].
- □ Additional support for users:

Identify comparative questions and components. Identify arguments and their pro/con stance.

racinity arguments and their preveen stance.

Inform about the number pro/con arguments.



Decision making involves:

Collecting information about the alternatives. Weighing pro and con arguments.

- 80% prefer online research over asking friends [Turner and Rainie; Pew Research Center 2020].
- □ Additional support for users:

Identify comparative questions and components.

Identify arguments and their pro/con stance.

#### **Contributions**

#### 1. Understanding Comparative Questions.

"Towards Understanding and Answering Comparative Questions" (full paper at WSDM 2022). "Comparative Web Search Questions" (full paper at WSDM 2020).

#### 2. Stance Detection for Answers.

"Towards Understanding and Answering Comparative Questions" (full paper at WSDM 2022).

#### 3. Retrieving Argumentative Answers.

"Overview of Touché 2020–2022: Argument Retrieval" (overview papers at CLEF 2020–2022).

#### 4. Clarifying Comparative Questions.

"A User Study on Clarifying Comparative Questions" (short paper at CHIIR 2022).

**Identifying Comparative Questions** 

Is buying a house better than renting? vs. How to rent a house?

**Identifying Comparative Questions** 

Is buying a house better than renting? vs. How to rent a house?

Annotated data:

50 000 questions from Yandex (Russian; 3% comparative).

31 000 from Bing, Google, Stack Exchange (English; 10% comparative).

**Identifying Comparative Questions** 

Is buying a house better than renting? vs. How to rent a house?

#### Annotated data:

50 000 questions from Yandex (Russian; 3% comparative).

31 000 from Bing, Google, Stack Exchange (English; 10% comparative).

High-precision classifier (precision is 1.0).

**Identifying Comparative Questions** 

Is buying a house better than renting? vs. How to rent a house?

#### Annotated data:

50 000 questions from Yandex (Russian; 3% comparative). 31 000 from Bing, Google, Stack Exchange (English; 10% comparative).

High-precision classifier (precision is 1.0) that combines:

Rules (patterns)

 $[is] \land RBR \land [than] \rightarrow Is buying a house better/RBR than renting? \rightarrow Recall: 0.44.$ 

**Identifying Comparative Questions** 

Is buying a house better than renting? vs. How to rent a house?

#### Annotated data:

```
50 000 questions from Yandex (Russian; 3% comparative).
31 000 from Bing, Google, Stack Exchange (English; 10% comparative).
```

#### High-precision classifier (precision is 1.0) that combines:

```
Rules (patterns) Feature-based classifiers
```

```
 [is] \land RBR \land [than] \rightarrow Is buying a house better/RBR than renting? \rightarrow Recall: 0.44.
```

+ Logistic regression  $\rightarrow$  probability of being comparative  $\rightarrow$  combined Recall: 0.54.

**Identifying Comparative Questions** 

Is buying a house better than renting? vs. How to rent a house?

#### Annotated data:

```
50 000 questions from Yandex (Russian; 3% comparative).
```

31 000 from Bing, Google, Stack Exchange (English; 10% comparative).

#### High-precision classifier (precision is 1.0) that combines:

Rules (patterns) Feature-based classifiers Neural classifiers.

```
 \begin{tabular}{lll} [is] $\land$ RBR \land [than] $\to$ & Is buying a house better/RBR than renting? $\to$ & Recall: 0.44. \\ \end{tabular}
```

- + Logistic regression  $\rightarrow$  probability of being comparative  $\rightarrow$  combined Recall: 0.54.
- + BERT  $\rightarrow$  probability of being comparative  $\rightarrow$  combined Recall: 0.60.

**Identifying Comparative Questions** 

Is buying a house better than renting? vs. How to rent a house?

#### Annotated data:

```
50 000 questions from Yandex (Russian; 3% comparative).
31 000 from Bing, Google, Stack Exchange (English; 10% comparative).
```

Rules (patterns) Feature-based classifiers Neural classifiers.

High-precision classifier (precision is 1.0) that combines:

```
\lceil is \rceil \land RBR \land \lceil than \rceil \rightarrow Is buying a house better/RBR than renting? <math>\rightarrow Recall: 0.44.
```

- + Logistic regression  $\rightarrow$  probability of being comparative  $\rightarrow$  combined Recall: 0.54.
- + BERT  $\rightarrow$  probability of being comparative  $\rightarrow$  combined Recall: 0.60.

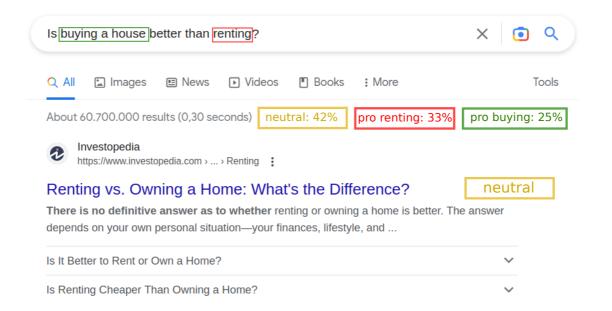
Recall: 0.60 (Russian), 0.71 (English); precision is 1.0 (of being comparative).

Taxonomy of Comparative Questions (Yandex)

Factual 35% Who is richer Jeff Bezos or Elon Musk?

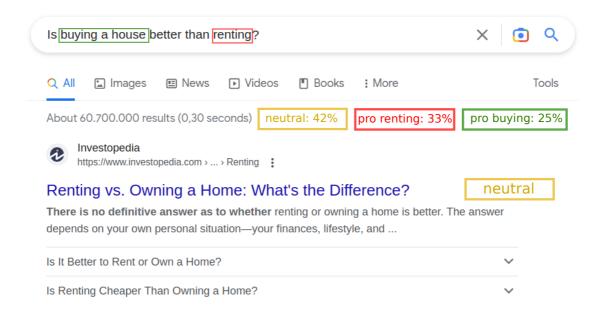
Taxonomy of Comparative Questions (Yandex)

Subjective 65% Is buying a house better than renting?



Taxonomy of Comparative Questions (Yandex)

Direct 43% Is buying a house better than renting?



Taxonomy of Comparative Questions (Yandex)

Indirect 57% What is the best housing option?

Taxonomy of Comparative Questions (Yandex)

Indirect

57% What is the best housing option?

Housing option (clarify)



■ Which options do you want to compare?

Buy a house vs. rent a house

Buy a flat vs. rent a flat

A house vs. a flat

I don't care

Taxonomy of Comparative Questions (Yandex)

With aspect 22% Is buying a house more <u>risky</u> than renting?

Taxonomy of Comparative Questions (Yandex)

Without aspect 78% Is buying a house <u>better</u> than renting?

Taxonomy of Comparative Questions (Yandex)

Without aspect 78% Is buying a house <u>better</u> than renting?

<u>Better</u> (clarify)



• Over which aspect do you want to compare?

Risks

Flexibility

Monthly payments

I don't care

Question Parsing for Answering Comparative Questions

Is buying a house more risky than renting?

Question Parsing for Answering Comparative Questions

What should be compared?

Buying a house involves more substantial financial risks than renting a house.

Question Parsing for Answering Comparative Questions

Is buying{a house} more risky than renting?
object 1 aspect object 2

- What should be compared?
- Should a particular shared property be emphasized?

Buying a house involves more substantial financial risks than renting a house.

Question Parsing for Answering Comparative Questions

Is buying{a house} more risky than renting?
object 1 predicate aspect object 2

- What should be compared?
- Should a particular shared property be emphasized?
- □ Guide the direction of the answer composition (e.g., a "more" or a "less" option).

Buying a house involves more substantial financial risks than renting a house.

Question Parsing for Answering Comparative Questions

- □ 3500 comparative questions labeled with objects, aspects, and predicates.
- □ Baseline: BiLSTM with 300-dimensional GloVe embeddings [Arora et al.; CIKM 2017].
- □ Our: Fine-tuned RoBERTa [Liu et al.; arXiv 2019].

| Classifier | F1 scores |      |        |        |
|------------|-----------|------|--------|--------|
|            | Predicate | None | Object | Aspect |
| BiLSTM     | 0.85      | 0.98 | 0.80   | 0.52   |
| RoBERTa    | 0.98      | 0.94 | 0.93   | 0.80   |

Question Parsing for Answering Comparative Questions

- □ 3500 comparative questions labeled with objects, aspects, and predicates.
- □ Baseline: BiLSTM with 300-dimensional GloVe embeddings [Arora et al.; CIKM 2017].
- □ Our: Fine-tuned RoBERTa [Liu et al.; arXiv 2019].

| Classifier | F1 scores |      |        |        |
|------------|-----------|------|--------|--------|
|            | Predicate | None | Object | Aspect |
| BiLSTM     | 0.85      | 0.98 | 0.80   | 0.52   |
| RoBERTa    | 0.98      | 0.94 | 0.93   | 0.80   |

Question Parsing for Answering Comparative Questions

- □ 3500 comparative questions labeled with objects, aspects, and predicates.
- □ Baseline: BiLSTM with 300-dimensional GloVe embeddings [Arora et al.; CIKM 2017].
- □ Our: Fine-tuned RoBERTa [Liu et al.; arXiv 2019].
- "Pre-classifying" questions with an aspect.

| Classifier | F1 scores |      |        |        |
|------------|-----------|------|--------|--------|
|            | Predicate | None | Object | Aspect |
| BiLSTM     | 0.85      | 0.98 | 0.80   | 0.52   |
| RoBERTa    | 0.98      | 0.94 | 0.93   | 0.80   |

Question Parsing for Answering Comparative Questions

- □ 3500 comparative questions labeled with objects, aspects, and predicates.
- □ Baseline: BiLSTM with 300-dimensional GloVe embeddings [Arora et al.; CIKM 2017].
- □ Our: Fine-tuned RoBERTa [Liu et al.; arXiv 2019].
- $exttt{ iny "Pre-classifying" questions with an aspect (F1: 0.80 <math> o$  0.90).

| Classifier | F1 scores |      |        |        |
|------------|-----------|------|--------|--------|
|            | Predicate | None | Object | Aspect |
| BiLSTM     | 0.85      | 0.98 | 0.80   | 0.52   |
| RoBERTa    | 0.98      | 0.94 | 0.93   | 0.90   |

Stance Detection for Answers

#### **Stance Detection for Answers**

Is buying a house better than renting?
object 1 object 2

#### **Stance Detection for Answers**

Is buying a house better than renting?

object 1 object 2

Pro object 1: Owning your own house is better than renting, because ...

Pro object 2: Although not ideal, renting does have its advantages like ...

Neutral: Both options have their advantages and disadvantages ...

No stance: Rent (definition): an agreement where a payment is made for ...

Is buying a house better than renting?

object 1 object 2

□ 950 answers from Stack Exchange labeled with 4 stance classes.

Is buying a house better than renting?

object 1 object 2

- □ 950 answers from Stack Exchange labeled with 4 stance classes.
- Most effective stance detector: RoBERTa with a sentiment prompting and masked objects.

38 © Alexander Bondarenko 2023

- 950 answers from Stack Exchange labeled with 4 stance classes.
- Most effective stance detector: RoBERTa with a sentiment prompting and masked objects.
  - 1. Comparison objects masked in questions and answers.
  - 2. Add a sentiment prompt: [OBJECT\_1] is better.
  - 3. Input for RoBERTa: [OBJECT\_1] is better [SEP] <answer>.

39 © Alexander Bondarenko 2023

- 950 answers from Stack Exchange labeled with 4 stance classes.
- Most effective stance detector: RoBERTa with a sentiment prompting and masked objects.
  - 1. Comparison objects masked in questions and answers.
  - 2. Add a sentiment prompt: [OBJECT\_1] is better.
  - 3. Input for RoBERTa: [OBJECT\_1] is better [SEP] <answer>.
- RoBERTa with a sentiment prompting: 0.63 (accuracy).

- 950 answers from Stack Exchange labeled with 4 stance classes.
- □ Most effective stance detector: RoBERTa with a sentiment prompting and masked objects.
  - 1. Comparison objects masked in questions and answers.
  - 2. Add a sentiment prompt: [OBJECT\_1] is better.
  - 3. Input for RoBERTa: [OBJECT\_1] is better [SEP] <answer>.
- □ RoBERTa with a sentiment prompting: 0.63 (accuracy).
- □ GPT-3 with a few-shot prompting: 0.65 (accuracy).

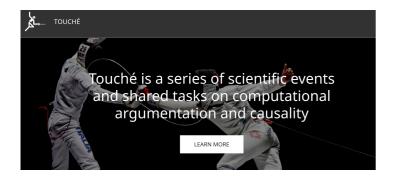
Touché: Argument Retrieval



- Platform for argument retrieval and argument analysis research.
- ☐ Argument relevance / quality / stance corpora and rankings.
- Organizing workshops at CLEF since 2020.

43 © Alexander Bondarenko 2023

Touché: Argument Retrieval



- ☐ Platform for argument retrieval and argument analysis research.
- ☐ Argument relevance / quality / stance corpora and rankings.
- Organizing workshops at CLEF since 2020.

Argument Retrieval for Comparative Questions:

□ Task: Given a comparative question, retrieve and rank documents of high arg. quality, detect the stance.

Touché: Argument Retrieval

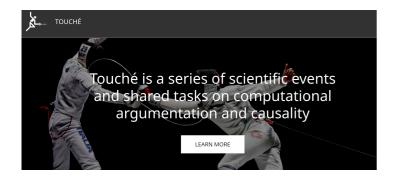


- Platform for argument retrieval and argument analysis research.
- ☐ Argument relevance / quality / stance corpora and rankings.
- Organizing workshops at CLEF since 2020.

#### Argument Retrieval for Comparative Questions:

- □ Task: Given a comparative question, retrieve and rank documents of high arg. quality, detect the stance.
- □ 16 teams, 55 runs, 6 000 manual judgments.

Touché: Argument Retrieval



- Platform for argument retrieval and argument analysis research.
- ☐ Argument relevance / quality / stance corpora and rankings.
- Organizing workshops at CLEF since 2020.

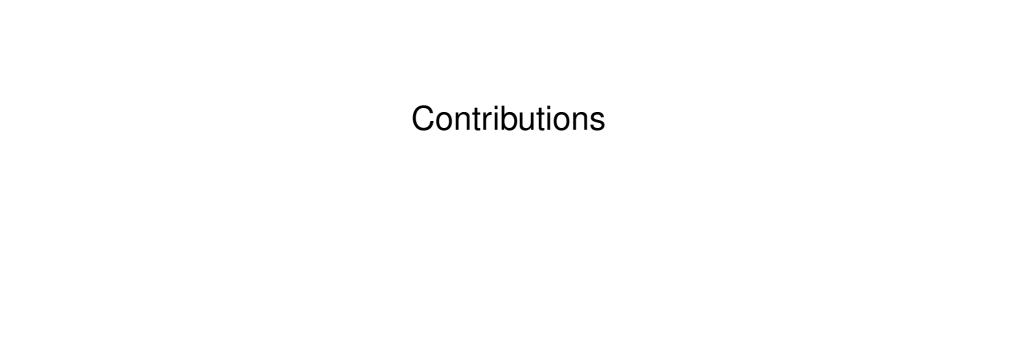
#### Argument Retrieval for Comparative Questions:

- ☐ Task: Given a comparative question, retrieve and rank documents of high arg. quality, detect the stance.
- □ 16 teams, 55 runs, 6 000 manual judgments.

#### Findings:

- □ From sparse to dense retrieval.
- Stance detection is hard.
- □ Best so far: retrieval pipelines that include argument mining and arg. quality estimation.

46



# Contributions

- High-precision classifiers for identifying comparative questions.
- A taxonomy of comparative information needs.
- Classifiers for detecting components in comparative questions.
- Stance detectors for potential answers.
- Shared tasks on argument retrieval and analysis.
- User study on clarifying comparative questions.

#### Contributions

- High-precision classifiers for identifying comparative questions.
- A taxonomy of comparative information needs.
- Classifiers for detecting components in comparative questions.
- Stance detectors for potential answers.
- Shared tasks on argument retrieval and analysis.
- User study on clarifying comparative questions.

thank you!