Comparative Web Search Questions

Alexander Bondarenko Pavel Braslavski Michael Völske Rami Aly
Maik Fröbe Alexander Panchenko Chris Biemann Benno Stein Matthias Hagen

Martin-Luther-Universität Halle-Wittenberg

www.webis.de

13th ACM International Conference on Web Search and Data Mining (WSDM 2020); Houston, Texas, USA, 2020

Comparative Web Search Questions

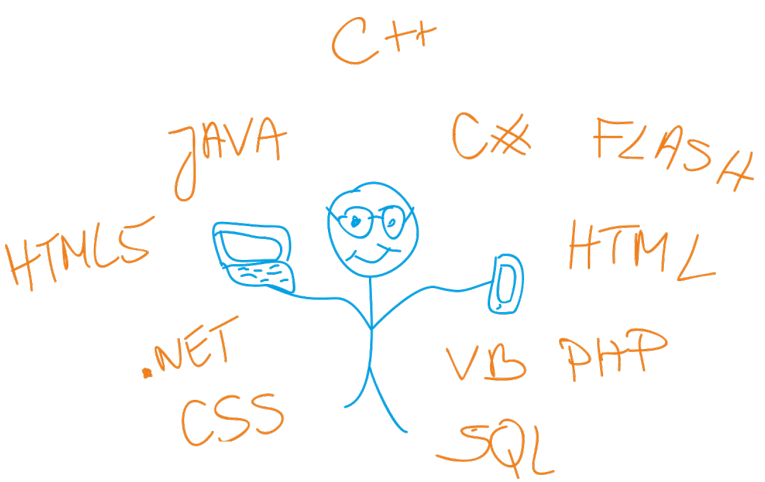
Which is a better pet, a cat or a dog?

- Comparative information need addresses a choice problem
- Informed choice needs arguments for and against

photo credit pixabay.com

Comparative Web Search Questions

- People make choices every day
- Post on Stack Exchange, but also submit to search engines
- Search engines do not show a pro/con result presentation



Which programming language should I learn first?

photo credit pixabay.com

Comparative Web Search Questions

Research goals:

- Identify comparative questions
- Study the underlying information needs
- Step towards showing a pro/con result presentation

Comparative Web Search Questions

1.5 billion question queries from 2012

11 million questions from 2012

- 50,000 Yandex and 12,500 Otvety questions labeled as comparative or not
- Comparative questions labeled with more fine-grained subclasses

Comparative Web Search Questions

Comparative Yandex question queries

- Opinionated/argumentative: 66%

  `Which is a better pet, a cat or a dog?`

- Reason/factoid: 33%

  `Which river is longer, Mississippi or Nile?`

- Request preference: 70%

  `What is the best pet for me?`

Comparative Web Search Questions

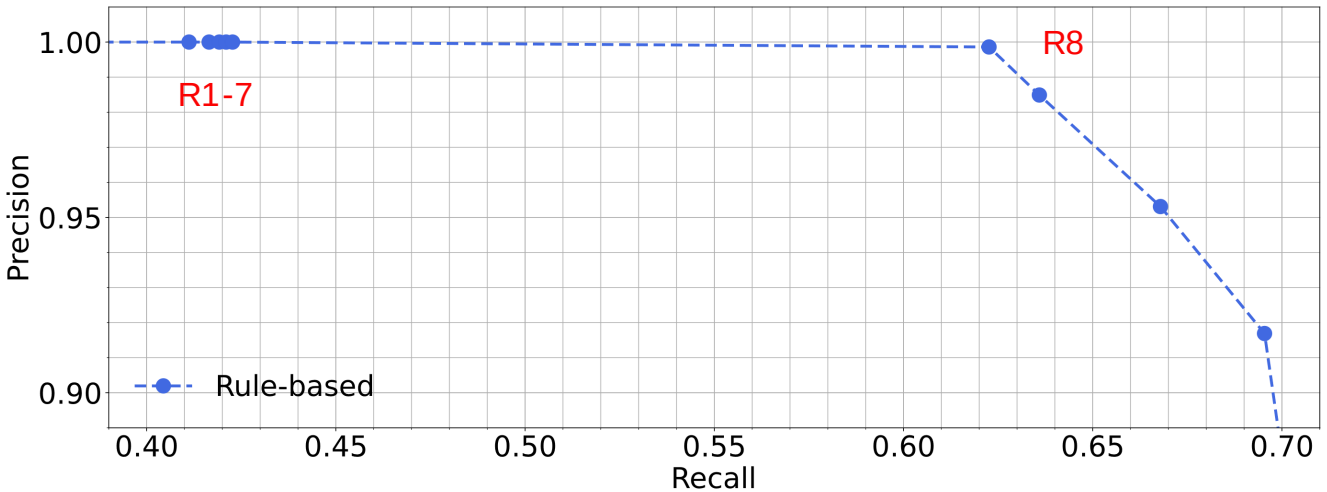
Identifying Comparative Questions

- Change a search engine's result presentation: intricate decision
- Identifying comparative questions: precision-oriented task
- Ensemble classifier: rules + traditional + neural
- Goal: Maximize recall at precision of 1.0

`Which is a better pet, a cat or a dog?`	comparative
`What kind of mammals are cats and dogs?`	non-comparative

Comparative Web Search Questions

Identifying Comparative Questions



(R1) `[better] ∧ ¬[how]`

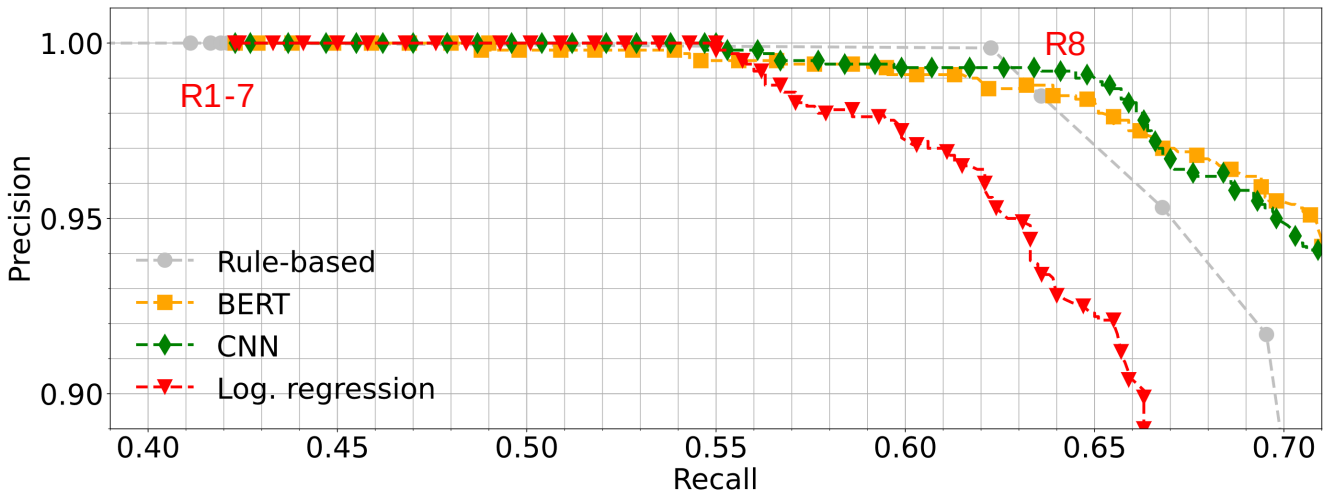
Which is a better pet, a cat or a dog? comparative

(R8) `[difference(s)?|distinguish] ∧ [and|from|or|vs]`

How to teach a dog to distinguish A and B? non-comparative

Comparative Web Search Questions

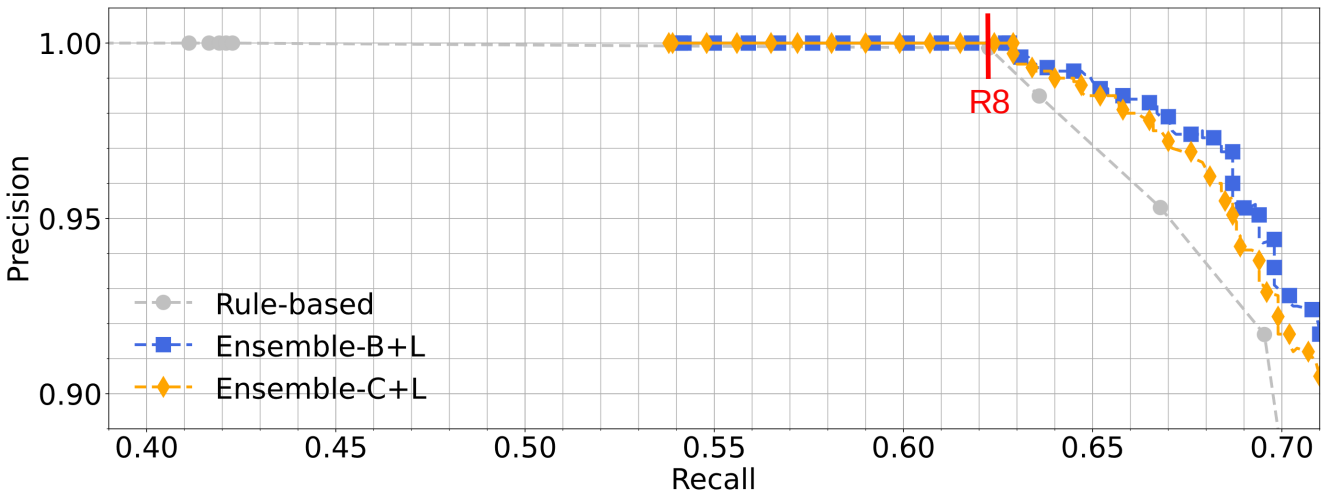
Identifying Comparative Questions



- Tune classifiers for a maximum precision
- Supplement handcrafted rules: extend recall

Comparative Web Search Questions

Identifying Comparative Questions

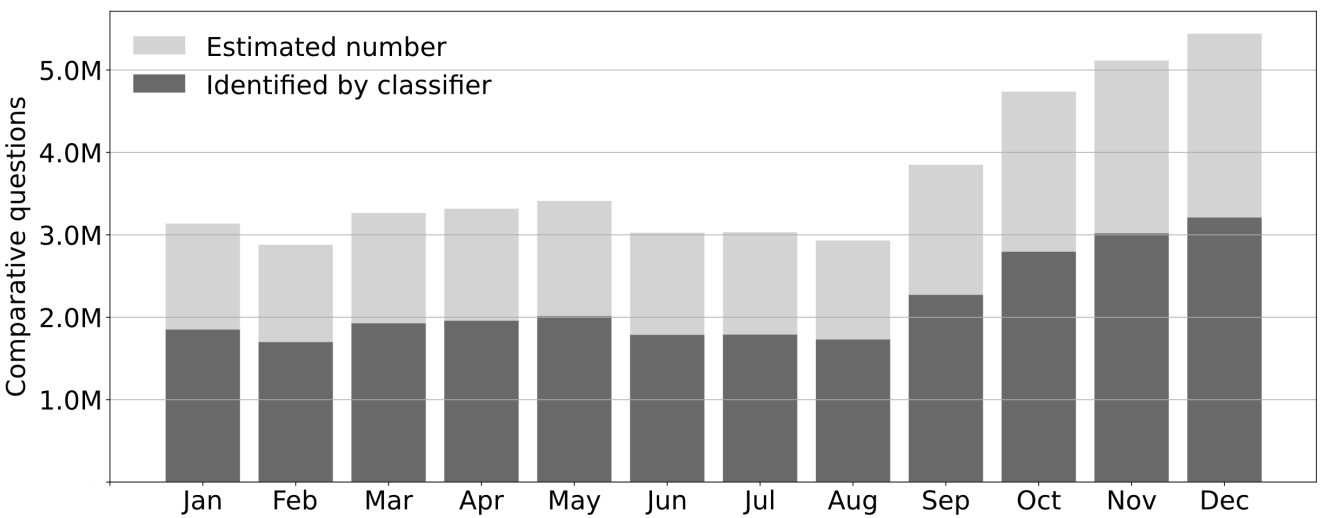


- Rules + BERT + Logistic regression
- Rules + CNN + Logistic regression
- Gradually decrease classifiers decision threshold

Comparative Web Search Questions

Yandex Log Analysis

- Test classifiers
- Ensemble-B+L recall 0.60 and Ensemble-C+L recall 0.59
- Classify 1.5 billion questions in the log: BERT days, CNN hours
- Yandex log: 2.8% are comparative questions (one per second)



Comparative Web Search Questions

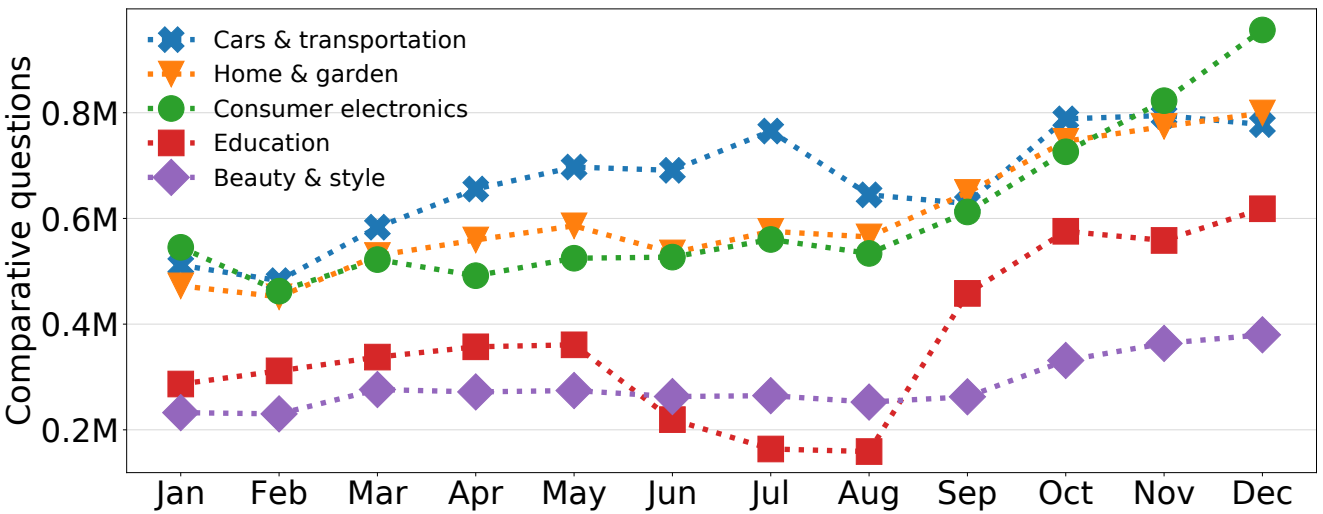
Yandex Log Analysis

Category	Quest. mln.	Comp. %	Most frequently asked question
Consum. electronics	105.4	6.3	`Which tablet is best to buy?`
Cars & transport.	143.7	5.2	`Anti-radar, which one is best?`
Home & garden	166.7	4.0	`Which vacuum cleaner is best to buy?`
Education	101.8	3.9	`Which pilot first surpassed the supersonic speed?`
Beauty & style	93.7	3.3	`When is it best to cut hair?`
...	...		

- More than just comparing products to buy

Comparative Web Search Questions

Yandex Log Analysis

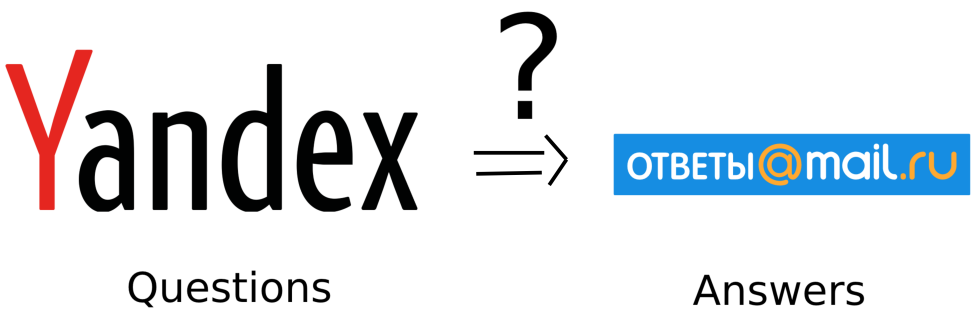


- Holidays in summer
- People compare vacation destinations starting from May
- In autumn differences between eatable and poisonous mushrooms

Comparative Web Search Questions

Answering Comparative Questions

- More than 65% of comparative questions are non-factoid
- Search for answers to Yandex questions on Otvety
- 48% have an answer on Otvety



Comparative Web Search Questions

Summary

- Yandex log: 2.8% are comparative questions (one per second)
- Comparison intents: beyond just comparing products to buy
- 65% of comparative questions are clearly non-factoid
- Half of them could be answered with CQA answers
- How can we answer non-factoid comparative web search questions?

Comparative Web Search Questions

Summary

- Yandex log: 2.8% are comparative questions (one per second)
- Comparison intents: beyond just comparing products to buy
- 65% of comparative questions are clearly non-factoid
- Half of them could be answered with CQA answers
- How can we answer non-factoid comparative web search questions?

Future Work

- Extract compared items and aspects
- Summarization for answers to the non-factoid questions

Comparative Web Search Questions

Summary

- Yandex log: 2.8% are comparative questions (one per second)
- Comparison intents: beyond just comparing products to buy
- 65% of comparative questions are clearly non-factoid
- Half of them could be answered with CQA answers
- How can we answer non-factoid comparative web search questions?

Future Work

- Extract compared items and aspects
- Summarization for answers to the non-factoid questions

thank you!

Backup

Experimental Results

Classification results on the test set. Goal: Best recall at a precision of 1.0

Individ. model	Recall	F1	Ensembles	Recall	F1
Logistic	0.54	0.70	Ens.-B+L	**0.60**	0.75
CNN	0.52	0.68	Ens.-C+L	**0.59**	0.74
BERT	0.44	0.61	Ens.-B+C+L	0.55	0.71
Rules (R1–7)	0.44	0.61			

Comparative subclass classifications on the test. Goal: Maximize F1

	CNN			BERT		
Subclass	Prec.	Rec.	F1	Prec.	Rec.	F1
Opinion/argument	0.93	0.90	0.92	0.92	0.91	0.91
Reason/factoid	0.85	0.79	0.82	0.82	0.89	0.86
Context/aspect	0.88	0.52	0.62	0.75	0.74	0.74
Method	0.79	0.80	0.79	0.75	0.82	0.78
Preference	0.97	0.98	0.97	0.96	1.00	0.97
Direct	0.95	0.96	0.96	0.95	0.98	0.97
Superlative	0.93	0.79	0.86	0.92	0.86	0.89
Micro average	0.92	0.88	0.90	0.90	0.93	0.91

Backup

Annotations

Absolute and relative frequencies of the comparative question subclasses, which are not mutually exclusive

	Yandex		Otvety	
Comparative	**1,405 (3% of all)**		**1,571 (13% of all)**	
Opinion	916	(65%)	1,469	(94%)
Argumentative	676	(48%)	586	(37%)
Reason	83	(6%)	10	(<1%)
Factoid	378	(27%)	101	(6%)
Method	106	(8%)	41	(3%)
Superlative	180	(13%)	287	(18%)
Direct	603	(43%)	893	(57%)
Aspect	302	(22%)	546	(35%)
Context	238	(17%)	405	(26%)
Preference (requested)	985	(70%)	1,281	(82%)
(stated)	18	(1%)	77	(5%)