Corpus-based Question Answering for *why*-Questions

Ryuichiro Higashinaka
and
Hideki Isozaki
NTT Communication Science Labs.

Overview

- Corpus-based approach for whyQA
  - Use relation-annotated corpora to automatically acquire causal expression patterns
  - Use a corpus of why-questions and answers to train an answer-candidate ranker
- Implemented system: NAZEQA
- Experimental results

WhyQA

- Answering generic “Why X?” questions by extracting causes from a text archive
  - E.g., Why do ostriches run fast?
- Few implemented systems
- Previous approach (Fukumoto, 2007)
  - Uses hand-crafted patterns to extract cause-bearing passages
  - Patterns rely on explicit causal cue words e.g., tame, node, de, kara (because/by/from)

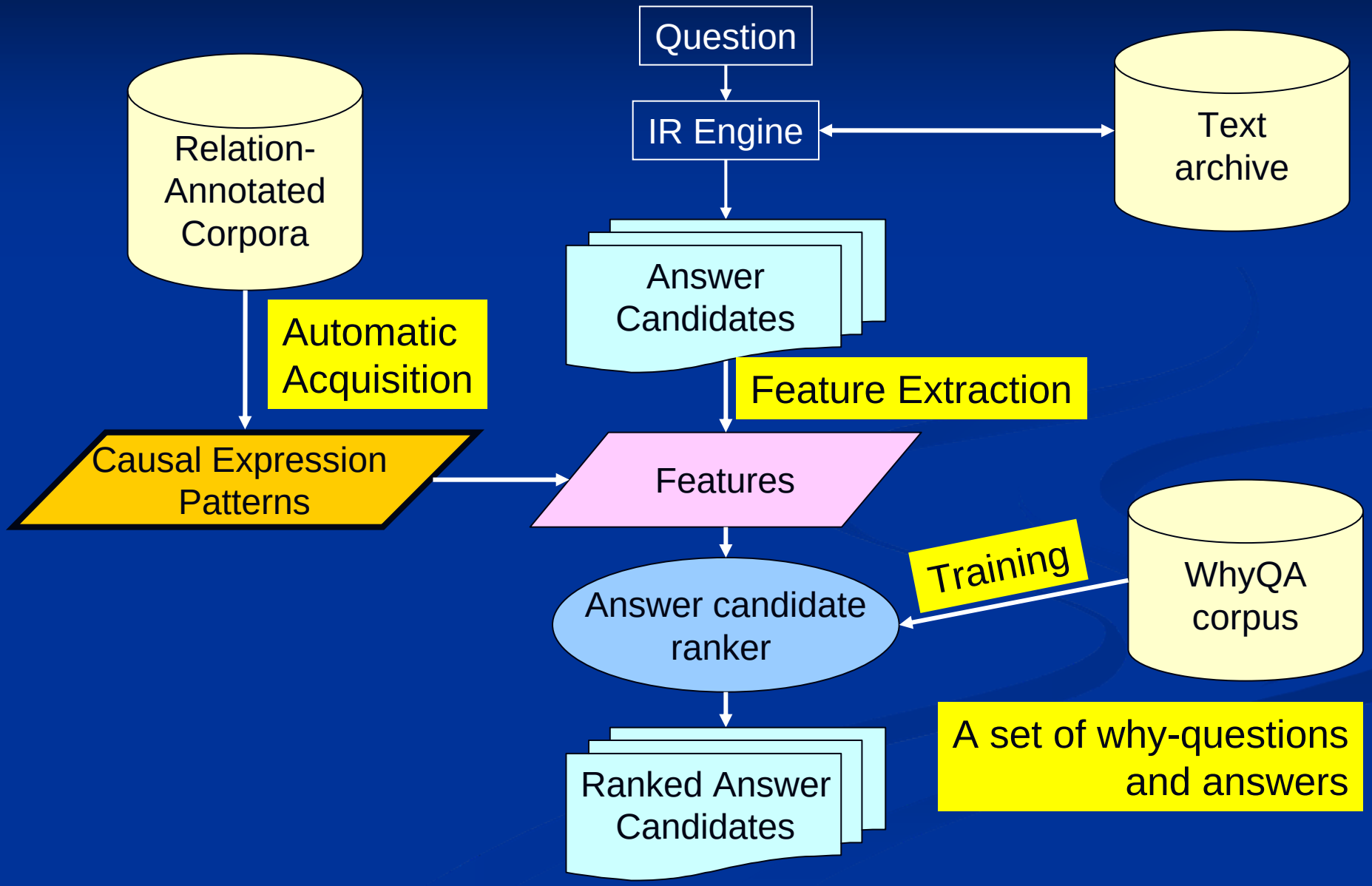
Problem

- Hard to cover causal expressions by hand
  - Causes are expressed by a wide variety of expressions
    - Half of causes are not marked by cue words (Inui and Okumura, 2005)
  - Cue words are not always reliable
    - Only 6-7 % of “de” (by) trigger causes (Abekawa and Okumura, 2004)
- Hand-crafted patterns are costly to make

Approach

- Automatically acquire causal expression patterns from relation-annotated corpora
  - FrameNet, PropBank, EDR Corpus
  - Typically annotated with a causal relation
- Use the acquired patterns to create features to represent answer candidates
- Train an answer-candidate ranker that ranks answer candidates on the basis of the features

Approach (cont'd)



Implemented System: NAZEQA

- A Japanese whyQA System
- Uses the EDR corpus as a relation annotated corpus to acquire causal expression patterns
- Derives three types of features for answer candidates
- Uses Ranking SVM for ranker training
- Uses a manually constructed WhyQA corpus (WHYQA Collection)

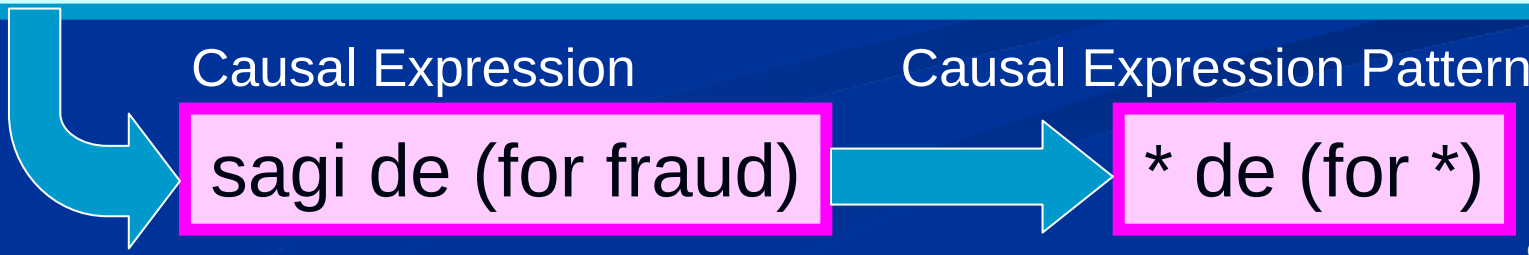
Acquiring Causal Expression Patterns from the EDR Corpus

- The EDR Corpus
  - A collection of Japanese sentences from various sources
  - Provides a semantic representation for each sentence
    - Specifies relations of content words
    - Relations include a causal relation
  - 8,474 text spans annotated with a causal relation in 207,802 sentences

Acquiring Causal Expression Patterns from the EDR Corpus

- Extract text spans annotated with a causal relation as causal expressions
- Convert the expressions into patterns by
  - Leaving only functional words
    - Auxiliary verbs, case, aspect, tense markers
  - Replacing others with wildcards “*”

Kare ha sagi de taiho sareta (He was arrested for fraud)



Acquired Causal Expression Patterns from the EDR Corpus

- Obtained 394 distinct causal expression patterns
- Examples
  - de (by/for)
  - tame (because)
  - niyoru, niyotte (because of)
  - no (-GEN) * wa (topic marker)
  - de (by/for) * wo (-ACC) * teshimai (-PERF)
  - koto niyotte (by the fact that)

Feature Extraction

- For an answer candidate to be the correct answer, it should
  - have a causal expression
  - be similar to the question in content
  - show some causal relation to the question
- Three types of features to represent how each candidate satisfies each condition
  - Causal Expression Features
  - Content Similarity Features
  - Causal Relation Features

Based on the acquired patterns

Causal Expression Features

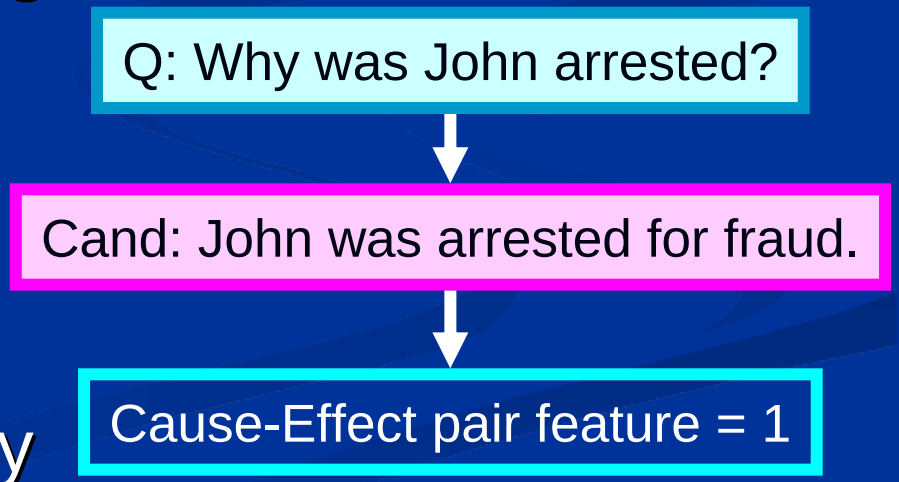
- AUTO-Causal Expression Features
  - 394 binary features representing if each acquired pattern matches the answer candidate
- MAN-Causal Expression Feature
  - A binary feature representing if the answer candidate is matched by existing hand-crafted patterns (Fukumoto, 2007)

Content Similarity Features

- Question-Cand Cosine Similarity feature
  - Cosine similarity of word frequency vectors
- Document-Question Relevance feature
  - Similarity between the question and the document where the candidate is found
  - Inverse rank of the retrieved document
- Synonym Pair feature
  - A binary feature indicating if synonyms are found in the question and answer candidate

Causal Relation Features

- Cause-Effect Pair feature
  - A binary feature representing if a cause-effect word pair is found in the answer candidate and the question
  - Cause-effect word pairs
    - Explode → Die
    - Murder → Arrest
    - Fraud → Arrest
  - Generated from the EDR concept dictionary



Q: Why was John arrested?

Cand: John was arrested for suspicion of fraud.

EDR Corpus → Causal Expression Patterns

Pattern1 for * of
Pattern2 by * of
Pattern3 due to
Pattern394 because

Matches pattern1 1
Matches pattern2 0
Matches pattern3 0
Matches pattern394 0

Causal Expression Features

Hand-crafted patterns (Fukumoto, 2007)

Matches hand-crafted patterns 0

Question-Candidate Cosine Similarity

0.3

Document-Question Relevance

0.1

EDR Concept Dictionary → Thesaurus

Examples
arrest ⇔ apprehend
arrest ⇔ capture
arrest ⇔ nip

Q and Cand have synonyms 1

Content Similarity Features

EDR Word Dictionary → Cause-Effect word pairs

Examples
explode ➔ die
murder ➔ arrest
fraud ➔ arrest

Causal Relation Feature

Q and Cand have A cause-effect word pair 1

WHYQA Collection

- 1,000 sets of why-questions and answers
- Procedure
  - Create a why-question
  - Retrieve top-20 documents for the question
  - Extract all sentences in the top-20 documents
  - Create sentence-level answers by selecting sentences that contain answers
  - Create paragraph-level answers by selecting paragraphs that contain answer sentences

Training an Answer-candidate Ranker

- Ranking SVM (Joachims, 2002)
  - Learns ranking by minimizing pairwise ranking error
- Training data
  - WHYQA Collection
  - Train a ranker so that answers (e.g., sentences selected as answers) get ranked higher than non-answers
- Trained answer-candidate rankers for sentence and paragraph-levels

Experiment

- Compare NAZEQA with two baselines
  - Baseline-1 (COS)
    - Uses cosine similarity to rank answer candidates
  - Baseline-2 (FK) (Fukumoto, 2007)
    - Selects answer candidates by hand-crafted patterns
    - Ranks the candidates by cosine similarity
- NAZEQA and baselines process the same answer candidates
  - All sentences/paragraphs in top-20 documents retrieved by an IR engine

Experiment (cont'd)

- Question set
  - All questions in the WHYQA Collection
  - 10-fold cross validation to evaluate NAZEQA
- Evaluation criteria
  - Mean Reciprocal Rank (MRR)
  - Coverage
    (Rate of questions correctly answered by top-N answers)

Results

	MRR			Coverage		
Top-N	COS	FK	NAZEQA	COS	FK	NAZEQA
Sentences as answer candidates						
Top-1	0.036	0.091++	0.113	3.6%	9.1%	11.3%
Top-5	0.086	0.139++	0.196**	19.1%	23.1%	35.4%
Top-10	0.102	0.149++	0.216**	31.3%	30.7%	50.4%
Top-20	0.115	0.152	0.227**	51.4%	35.5%	66.6%
Paragraphs as answer candidates						
Top-1	0.065	0.152++	0.186*	6.5%	15.2%	18.6%
Top-5	0.140	0.245++	0.305**	29.2%	41.6%	53.1%
Top-10	0.166	0.257++	0.328**	48.8%	50.5%	70.3%
Top-20	0.181	0.262++	0.339**	70.7%	56.4%	85.6%

** ($p<0.01$) * ($p<0.05$) NAZEQA's statistical significance over FK

++ ($p<0.01$) FK's statistical significance over COS

Results

	MRR			Coverage		
Top-N	COS	FK	NAZEQA	COS	FK	NAZEQA
Sentences as answer candidates						
Top-1	0.036	0.091++	**0.113**	3.6%	9.1%	11.3%
Top-5	0.086	0.139++	**0.196****	19.1%	23.1%	35.4%
Top-10	0.102	0.149++	**0.216****	31.3%	30.7%	50.4%
Top-20	0.115	0.152	**0.227****	51.4%	35.5%	66.6%
Paragraphs as answer candidates						
Top-1	0.065	0.152++	**0.186***	6.5%	15.2%	18.6%
Top-5	0.140	0.245++	**0.305****	29.2%	41.6%	53.1%
Top-10	0.166	0.257++	**0.328****	48.8%	50.5%	70.3%
Top-20	0.181	0.262++	**0.339****	70.7%	56.4%	85.6%

NAZEQA significantly outperforms FK

Results

	MRR			Coverage		
Top-N	COS	FK	NAZEQA	COS	FK	NAZEQA
Sentences as answer candidates						
Top-1	0.036	0.091++	0.113	3.6%	9.1%	**11.3%**
Top-5	0.086	0.139++	0.196**	19.1%	23.1%	**35.4%**
Top-10	0.102	0.149++	0.216**	31.3%	30.7%	**50.4%**
Top-20	0.115	0.152	0.227**	51.4%	35.5%	**66.6%**
Paragraphs as answer candidates						
Top-1	0.065	0.152++	0.186*	6.5%	15.2%	**18.6%**
Top-5	0.140	0.245++	0.305**	29.2%	41.6%	**53.1%**
Top-10	0.166	0.257++	0.328**	48.8%	50.5%	**70.3%**
Top-20	0.181	0.262++	0.339**	70.7%	56.4%	**85.6%**

High coverage achieved by NAZEQA

Results

	MRR			Coverage		
Top-N	COS	FK	NAZEQA	COS	FK	NAZEQA
Sentences as answer candidates						
Top-1	0.036	0.091++	0.113	3.6%	9.1%	**11.3%**
Top-5	0.086	0.139++	0.196**	19.1%	23.1%	**35.4%**
Top-10				31.3%	30.7%	**50.4%**
Top-20				51.4%	35.5%	**66.6%**
Paragra						
Top-1				6.5%	15.2%	**18.6%**
Top-5	0.140	0.245++	0.305**	29.2%	41.6%	**53.1%**
Top-10	0.166	0.257++	0.328**	48.8%	50.5%	**70.3%**
Top-20	0.181	0.262++	0.339**	70.7%	56.4%	**85.6%**

- Low coverage of FK
- Possible limitation of hand-crafted patterns

High coverage achieved by NAZEQA

Impact of features

Feature Set	Sentence (Top-5 MRR)	Paragraph (Top-5 MRR)
All features	0.181	0.287
w/o AUTO-Causal Expression	**0.138****	**0.217****
w/o MAN-Causal Expression	0.179	0.286
w/o Question-Candidate Cosine Similarity	**0.131****	**0.188****
w/o Document-Question Relevance	0.161	0.275
w/o Synonym Pair	0.180	0.282
w/o Cause-Effect Pair	0.184	0.287

MRR drops significantly without AUTO-Causal expression features ➔ Verifies effectiveness of automatically acquired causal expression patterns

Weights of features assigned by Ranking SVM

Rank	Feature Name	Weight
1	Question-Candidate Cosine Similarity	4.66
2	Exp.[*de* (by) * *wo* (-ACC) * *teshimai* (-PERF)]	1.86
3	Exp.[*no* (of) * *niyote wa* (according to)]	1.44
4	Exp.[*no* (of) * *na* (AUX) * *no* (of) * *de* (by)]	1.42
5	Exp.[*no* (of) * *ya* (or) * *niyotte* (by)]	1.35
6	Exp.[*no* (of) * *ya* (or) * *no* (of) * *de* (by)]	1.30
7	Exp.[*na* (AUX) * *niyotte* (by)]	1.23
8	Exp.[*koto niyotte* (by the fact that)]	1.22
9	Exp.[*to* (and) * *no* (of) * *niyotte* (by)]	1.20
10	Document-Question Relevance	0.89
	⋮	
27	Synonym Pair	0.40
102	MAN-Causal Expression	0.16
127	Cause-Effect Pair	0.15

Weights of features assigned by Ranking SVM

Rank	Feature Name	Weight
1	Question-Candidate Cosine Similarity	4.66
2	Exp.[*de* (by) * *wo* (-ACC) * *teshimai* (-PERF)]	1.86
3	Exp.[*no* (of) * *niyote wa* (according to)]	1.44
4	Exp.[*no* (of) * *na* (AUX) * *no* (of) * *de* (by)]	1.42
5	Exp.[*no* (of) * *ya* (or) * *niyotte* (by)]	1.35
6	Exp.[*no* (of) * *ya* (or) * *no* (of) * *de* (by)]	1.30
7	Exp.[*na* (AUX) * *niyotte* (by)]	1.23
8	Exp.[*koto niyotte* (by the fact that)]	1.22
9	Exp.[*to* (and) * *no* (of) * *niyotte* (by)]	1.20
10	Document-Question Relevance	0.89

- Many complex patterns were found to be important
- Difficult to cover such patterns by hand
  - ➔ shows effectiveness of an automatic approach

Conclusion

- Corpus-based approach for whyQA
  - Use relation-annotated corpora to acquire causal expression patterns
  - Train an answer-candidate ranker using a corpus of why-questions and answers
- Experimental results show validity of our approach
- Future work
  - Use other relations (e.g., purpose) to increase coverage of causal expressions
  - Incorporate syntactic and semantic features