A Data Driven Approach to Relevancy Recognition for Contextual Question Answering

Authors: Fan Yang (OGI)
Junlan Feng
Giuseppe Di Fabbrizio

Speaker: Junlan Feng

Outline

- ***Motivations***
- **Previous research work**
- **A data driven approach**
- **Results**
  - Results on TREC data
  - Results on HandQA data
- **Preliminary contextual information fusion**
- **Future work**

Motivations

- **WebTalk, a Research Project at AT&T Labs**
  - WebTalk is a system for analyzing unstructured information from company websites to support automatic creation of customer care dialog applications.
  - Question Answering is a key component technology.
    - Users often ask questions naturally as part of contextualized interaction.
    - Many questions that users frequently want answers for cannot be satisfied with a simple answer. By the nature, the question initiates a dialog.
- **Most available QA systems and QA technologies are limited to answer questions in isolation.**
- **Contextual question answering (QA), in which users' information needs are satisfied through an interactive QA dialogue.**

Research Purpose

- **To develop techniques for contextual QA**
  - Relevancy recognition
    - Determine whether a question is relevant to the previous interaction context
  - Contextual information fusion:
    - Use contextual information to help retrieve answers

Outline

- **Introduction**
- ***Related research work***
- **A data driven approach**
- **Results**
  - Results on TREC data
  - Results on HandQA data
- **Preliminary contextual information fusion**
- **Future work**

A Rule Based Algorithm

Marco De Boni and Suresh Manandhar. 2005.

Given a sequence of questions $Q_1, \ldots, Q_i$

Syntactic Rules:

1. If Q_i has a pronoun or possessive adjective, which has no references in the current question, Q_i is a follow-up question.
2. If Q_i has cue words such as "precisely" or "exactly", Q_i is a follow-up question.
3. If Q_i does not contain any verbs, Q_i is a follow-up question.

Semantic Rules:

4. Otherwise, calculate the semantic similarity measure of Q_i as SimilarityMeasure(Q_i) = max f(j)*SentenceSimilarity(Q_i; Q_{i-j})
   Here f(j) is a decay function. If the similarity measure is higher than a certain threshold, Q_i is a follow-up question.
5. Otherwise, if answer is available, calculate the semantic similarity between Q_i and the immediate previous answer A_{i-1}: SentenceSimilarity(Q_i; A_{i-1}). If it is higher than a certain threshold, Q_i is a follow-up question that is related to the previous answer.

6. Otherwise, Q_i begins a new topic.

Outline

- Introduction
- Previous research work
- ➢ ***<u>A data driven approach</u>***
- Results
  - Results on TREC data
  - Results on HandQA data
- Preliminary contextual information fusion
- Future work

Feature Extraction

- For each question, we extract
  - Syntactic features (binary)
    - Pronoun: (exception "I", "our", "yours" …)
    - Noun:
    - Proper Noun:
    - Verb:
  - Semantic Similarity between Q and the context -..Word similarity measures:
    - PATH: noun and verb
    - WUP: noun and verb [Wu & Palmer 1994]
    - LIN: noun and verb [Lin 1998]
    - VECTOR: noun, verb, and adjective [Patwardhan 2003]

Semantic Similarity Between Questions

- **Questions:**

  Current question: $Q = \{w_1, w_2, \ldots w_n\}$

  A previous question: $Q' = \{w'_1, w'_2, \ldots w'_m\}$

- **Sentence-sentence similarity**

$$SentenceSimilarity(Q, Q') = \frac{1}{n} \sum_{1 \le j \le n} \left(\max_{1 \le i \le m} WordSimilarity(w_j, w'_i) \right)$$

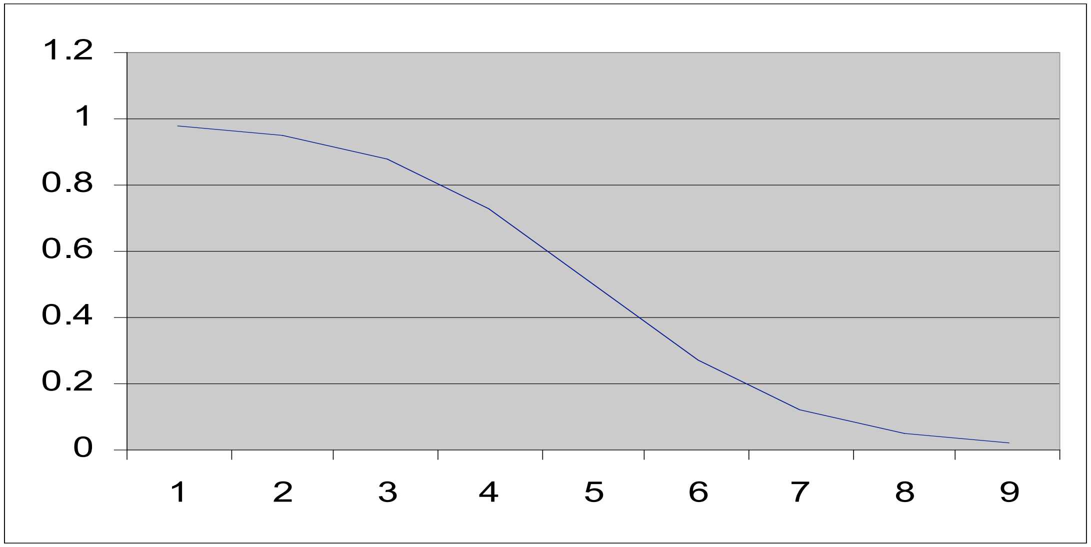
- **Word-word similarity: based on Wordnet**
  – **PATH, LIN, WUP, VECTOR**

Question Similarity Measurement

Given a sequence of questions $Q_1, \ldots, Q_{ni}$

$$Similarity(Q_i, Context) = \underset{0<j<i}{MAX}(d(j) * SS(Q_i, Q_{i-j}))$$

$d(x) = 1 - \frac{1}{1+e^{(n-x)}}$ **is a decay function**



Learning Algorithms

- A binary classification problem
  - Decision Tree (DT)
  - Adaboost

Outline

- **Introduction**
- **Previous research work**
- **A data driven approach**
- ***Results on Relevancy Recogition***
  - Results on TREC data
  - Results on HandQA data
- **Preliminary contextual information fusion**
- **Future work**

TREC Data

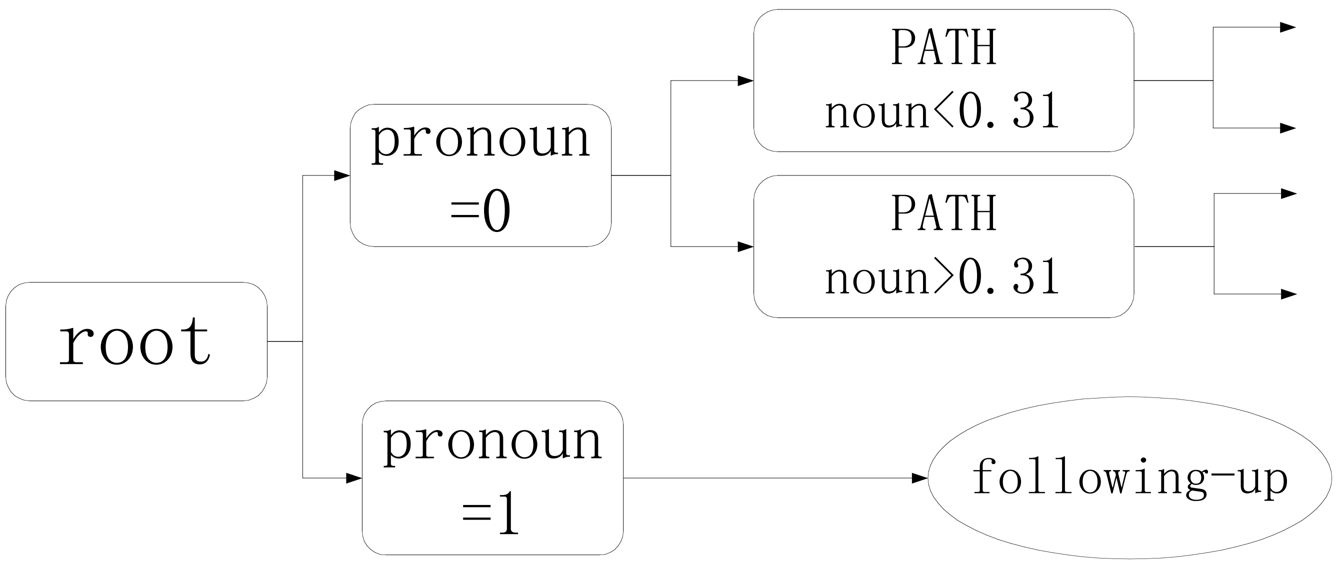
- Training set
  - TREC 2004
  - 286 questions
  - 64 series
- Testing set
  - TREC 2001 context track
  - 42 questions
  - 10 series

Results on TREC Data Using DT

	Training data			Testing data			
	Predicted Class			Predicted Class			
True Class	First	Follow		First	Follow	Recall	
First	63	1		9	1	**90%**	**90%**
Follow	2	220		2	30	**94%**	**78%**
	65	221		10	32		
Precision	**96.9%**	**99.5%**		**82%**	**97%**		
				56%	**96%**		
Accuracy	**99.0%**			**93%**		**81%**	

Orange: performances using the rule-based algorithm

Tree



root
pronoun
=0
pronoun
=1
PATH
noun<0.31
PATH
noun>0.31
following-up

Error Analysis

- 2 failures to recognize follow-up question
  - Lack of semantic relations in WordNet
- 1 failure to recognize the first question
  - Over-fitting of decision tree learning
  - Adaboost ?

Results Using Adaboost

	Training data			Testing data		
	Predicted Class			Predicted Class		
True Class	First	Follow		First	Follow	Recall
First	64	2		10	0	**100%**
Follow	1	220		5	27	**84%**
Total	65	222		15	27	
Precision	**98.5%**	**99%**		**67%**	**100%**	
Accuracy	**99.0%**			**88%**		

HandQA Data

- Characteristics
  - Real data collected from a customer care QA system
  - Repeat or rephrase questions. Examples:

    How to make number non published

    Non published numbers

    How to make number non listed
  - Noisy: typos, bad grammars, keywords, …
  - 5908 questions
  - 2184 series
  - 90% data used for training, 10% for testing

Results on HandQA Data

	Training data			Testing data	
Class	Recall	Precision		Recall	Precision
First	75%	68%		73%	62%
Follow	79%	84%		75%	83%

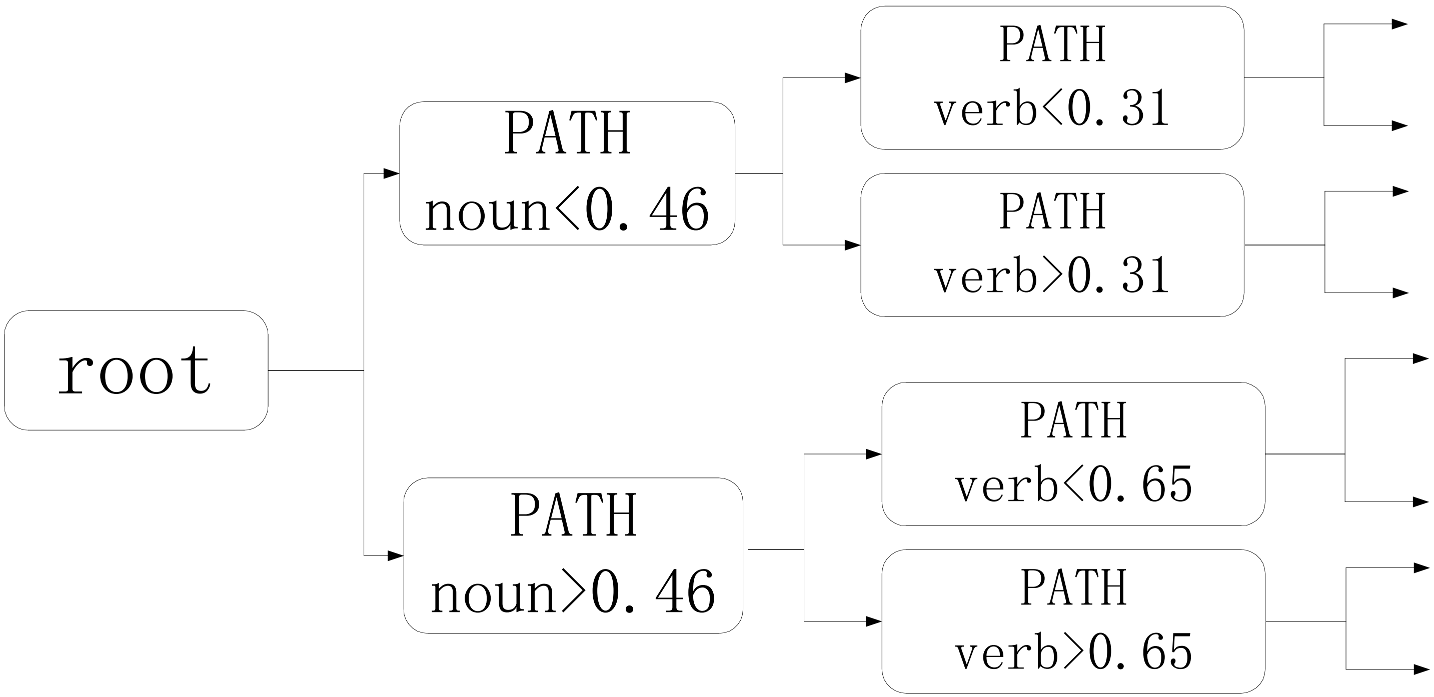
- **Overall Accuracy: 74%**

Experimental Analysis with HandQA Data

- Syntactic features
  - Not helpful
  - Not reliable
    - Typos, grammars, capitalization, punctuation, …

- Semantic features
  - More important due to characteristics of data
  - Similar topics in consecutive series

Tree


root
PATH
noun<0. 46
PATH
noun>0. 46
PATH
verb<0. 31
PATH
verb>0. 31
PATH
verb<0. 65
PATH
verb>0. 65

Summary of Results

- Machine learning approach
  - Flexible, different rules for different data sets
    - Pronoun for TREC; PATH for HandQA
  - Better results
  - Describe the data better

- Semantic similarity
  - PATH is one of the dominating features

Outline

- **Introduction**
- **Previous research work**
- **A data driven approach**
- **Results**
  - Results on TREC data
  - Results on HandQA data
- ***Preliminary contextual information fusion***
- **Future work**

TREC 2004 Data

<target id="2" text="Fred Durst">　Topic Words

<q id="2.1" type="factoid">

What is the name of Durst's group? </q>

<q id="2.2" type="factoid">

What record company is he with? </q>

<q id="2.3" type="list">

What are titles of the group's releases? </q>

<q id="2.4" type="factoid">

Where was Durst born? </q>

</target>

Approaches of TREC Participants

- Search in topic words docs

- Topic words attachment
  - Attach topic words to each question

- Anaphoric replacement
  - Replace pronouns with topic words

- Deep anaphora analysis
  - Trying to find the true referent for pronouns

Is Context Information Useful?

- Context info doesn't not help. [Winikoff and Kosseim, 2004]
  - First run: original questions (contextual questions)
  - Second run: manually pronoun replacement
    (independent questions)
  - Results: not improved
  - Explanation: poor performance of the QA system?

Context Info in DR

- **Is context necessary?**
  - **- Let's use document retrieval**

The top n documents (n=?)	50	1000
Question	20%	39%
Topic words	85%	96%
Topic words+ Question	87%	96%

Use questions in previous turns

The top n documents (n=?)	50	1000
Question	20%	39%
Topic words *(not available in reality)*	85%	96%
Noun (first question)	**81%**	**92%**
First question	**76%**	**93%**
Topic words + Question	87%	96%
PN (first question) + Question	**77%**	**92%**
Noun (first question) + Question	**84%**	**94%**
First question + Question	**82%**	**94%**

Use questions in previous turns (Cont.)

The top n documents (n=?)	50	1000
Questions Only	20%	39%
Topic words	85%	96%
Noun	81%	92%
First question	76%	93%
Topic words + Question	87%	96%
PN (first question) + Question	77%	92%
Noun (first question) + Question	84%	94%
Incremental Noun	**87%**	**94%**
First question + Question	82%	94%

Incremental Nouns: Nouns in previous questions with the Semantic Similarity PATH > 0.08

Make use of answers in the context

Mode	50	1000
Question	20%	39%
Topic words	85%	96%
Noun	81%	92%
First question	76%	93%
Topic words + Question	87%	96%
PN (first question) + Question	77%	92%
Noun (first question) + Question	84%	94%
Noun(first question) + Question + Answer	**86%**	**95%**
First question + Question	82%	94%
First question + Question + Answer	**86%**	**95%**

Outline

- **Introduction**
- **Previous research work**
- **A data driven approach**
- **Results**
  - Results on TREC data
  - Results on HandQA data
- **Preliminary contextual information fusion**
- ***Future work***

Future Work

- More research to improve performance
  - Integrate context in QA
  - Evaluate context in QA
- Dialogue-based QA
  - [Small et al. 2004]
- Implement into a QA system