Question Classification for a Croatian QA System

Tomislav Lombarović, Jan Šnajder, Bojana Dalbelo Bašić

Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

BSNLP 2011
Plzeň - Chech Republic, 5 September 2011

Contents

Outline

Introduction

- Large amounts of information are available today
- The need for effective search becomes more important
- Users want targeted and precise answers to their questions

Question answering

- QA system provides answer to a user's question, rather than a list of relevant documents
- First steps in the '60 (BASEBALL, LUNAR)
- Steady increase in research (TREC QA and CLEF QA tracks)
- Recent work on QA for Slavic languages: Bulgarian (Simov & Osenova 2006), Polish (Walas & Jassem 2003) and Slovene (Čeh & Ojsterešek 2009)
- QA system can be broken down into several steps
  - question classification
  - document retrieval
  - paragraph of passage retrieval
  - answer synthesis

Question classification

- Question should be classified according to the expecting answer type
- Various methods
  - rule-based methods (regular expressions)
  - statistical language modelling
  - machine learning
- Question taxonomy
  - simple flat taxonomy
  - more complex multilevel taxonomy (fine- and coarse-grained)

Example

Question

What chocolate bar created by Frank Mars and his wife is often called a Milky Way with peanuts?

Document passage

A milk chocolate bar filled with peanut butter nougat, roasted peanuts and caramel makes Snickers the best-selling candy bar. According to Mars Incorporated, there are 16 peanuts in Snickers. The United Kingdom and Ireland sell it as the Marathon bar. The name Snickers comes from a horse owned by the Mars family.

- Classify as ENTITY-Food, retrieve passage, extract entities of the correct type

Related Work

- Classification models:
  - Early work used rule-based classification (Kwok et al. 2001)
  - Machine learning (Zhang & Lee 2003): SVM, DT
  - SVM (Haciouglu & Ward 2003; Metzler & Croft 2004)
  - SNOW (Li & Roth 2002)
- Features:
  - words and ngrams (Zhang & Lee 2003)
  - syntactic features (noun phrases, chunks, and head chunks) (Li & Roth 2002; Metzler & Croft 2004)
  - semantic features: named entities (Haciouglu & Ward 2003; Li & Roth 2002), WordNet hypernyms (Metzler & Croft 2004)
- Question taxnonomy:
  - early approaches use one-level taxonomy
  - two-level taxonomy (Li & Roth 2002)

Outline

Question Classification for Croatian

- Question taxonomy: similar to (Li & Roth 2002)
  - two level taxonomy
  - 6 coarse and 50 fine classes
- Classification models
  - support vector machines (LibSVM)
  - decision trees (Rapid Miner)
  - k-nearest neighbours
  - language models

Question taxonomy

- Coarse Question taxonomy
  1. ABBREVATION (Abbreviation, Expansion)
  2. DESCRIPTION (Definition, Description, Manner, Reason)
  3. ENTITY (Animal, Body, Color, ...22 subclasses)
  4. HUMAN (Description, Group, individual, Title)
  5. LOCATION (City, Country, Mountain, State, Other)
  6. NUMERIC (Code, Count, Date, ...13 subclasses)

Features

- Simple features
  - word forms
  - bigrams (skip bigrams)
- Lematization and feature selection
  - reduces feature space

QC test collection

- No available QC test collection for Croatian language
- We built one from scratch
- Total of 2303 questions
  - Collection C1: 1350 already classified questions translated from English (Li & Roth 2002)
  - Collection C2: 953 new question from the Croatian edition of game show *"Who Wants to Be a Millionaire?"*
  - Collection C3: C1 $\cup$ C2

Outline

Evaluation

- Test collections: C1, C2 and C3
- Four classification models
- Classification strategies:
  - fine-grained
  - coarse-grained
  - hierarchical fine-grained
- Document frequency feature selection
  - can remove 60% of features without affecting performance

Classification performance

		Coarse-grained [%]		Fine-grained [%]		Hier. Fine-grained [%]	
	Collection	Acc	F1	Acc	F1	Acc	F1
SVM	C1	85.7	77.9	70.2	36.9	69.4	36.2
	C2	75.9	62.8	69.2	21.8	66.5	21.4
	C3	83.3	78.0	69.9	39.4	69.8	39.2
DT	C1	75.6	71.6	62.8	39.4	56.2	27.2
	C2	68.5	66.2	62.4	20.8	57.4	15.7
	C3	77.1	66.2	65.6	35.3	61.5	29.6
k-NN	C1	75.9	70.4	60.8	31.2	53.7	27.9
	C2	70.9	58.6	60.5	19.0	60.3	17.3
	C3	74.6	71.9	60.7	34.0	60.8	33.7
LM	C1	66.6	60.3	55.5	29.0	53.7	26.3
	C2	60.9	52.4	53.0	17.2	50.6	16.8
	C3	60.5	54.9	52.4	30.7	47.4	27.9

Per-category performance

SVM coarse-grained classification on C1

	ABB.	ENTITY	DESC.	HUMAN	LOCATION	NUMERIC
P (%)	100.0	75.5	85.7	89.7	92.7	95.1
R (%)	38.1	85.4	88.0	84.1	83.0	92.9
F1 (%)	55.2	79.0	86.8	86.8	87.6	94.0

Word forms vs. lemmas vs. stems

SVM classification on C3

	Coarse-grained (%)				Fine-grained (%)			
	Acc	P	R	F1	Acc	P	R	F1
word forms	82.3	79.6	75.8	76.5	67.7	41.5	36.4	37.0
stems	82.9	82.9	77.7	79.2	70.0	41.1	39.7	40.1
lemmas	83.3	81.5	76.7	78.0	69.9	43.4	39.1	39.4

Croatian vs. English

SVM, collection C1

Classifier		Coarse-grained (%)				Fine-grained (%)			
		Acc	P	R	F1	Acc	P	R	F1
SVM	cro	85.7	81.0	76.8	77.9	70.2	39.1	37.3	36.9
	eng	78.7	78.8	75.0	76.1	64.2	32.4	30.2	29.9
DT	cro	75.6	73.7	70.8	71.6	62.8	45.7	38.9	39.4
	eng	70.0	78.5	65.7	69.1	58.3	37.8	31.7	32.8
k-NN	cro	75.9	71.6	71.2	70.4	60.8	32.8	33.0	31.2
	eng	70.7	70.1	68.2	68.1	57.1	29.4	30.1	27.9
LM	cro	66.6	63.1	62.1	60.3	55.5	31.9	29.7	29.0
	eng	63.1	58.7	64.6	58.0	52.2	29.6	27.9	27.3

Outline

Conclusion

- Question classification is important for QA
- We experimented with four classification methods and built a QC test collection
- SVM outperforms other classifiers
- Coarse-grained more accurate than fine-grained
- Morphological normalization increases performance
- Performance is better for Croatian than for English
- Future work:
  - improvement of the QC test collection
  - use of syntactic and semantic features

Questions