Automatic Feature Engineering for **Answer Selection** and **Extraction**

Aliaksei Severyn[1], Alessandro Moschitti[1,2]
[1]DISI, University of Trento, Italy
[2]QCRI, Qatar Foundation, Doha, Qatar

Abstract

Goal:

- Reduce effort put into tedious feature engineering for Question Answering

Tasks:

- **Answer sentence selection**
- **Answer extraction**

Previous work:

- Represent each q/a pair with a large number of **features** capturing lexical, syntactic, semantic **similarities** between question and its answer
- State-of-the-art methods rely on powerful syntactic features from Tree Edit Distance features [Wang et al. 2007, Yao et al. 2013]

Our approach:

- Model q/a pairs **explicitly** as linguistic structures (shallow syntactic and dependency trees) [CoNLL,2013]
- Rely on Kernel Learning to **automatically extract** and **learn** powerful discriminative syntactic patterns
- Exploit **semantically motivated relational linking** between question and answer passages trhough Question type and Focus classifiers + NER
- Treat **answer extraction** as a classification task

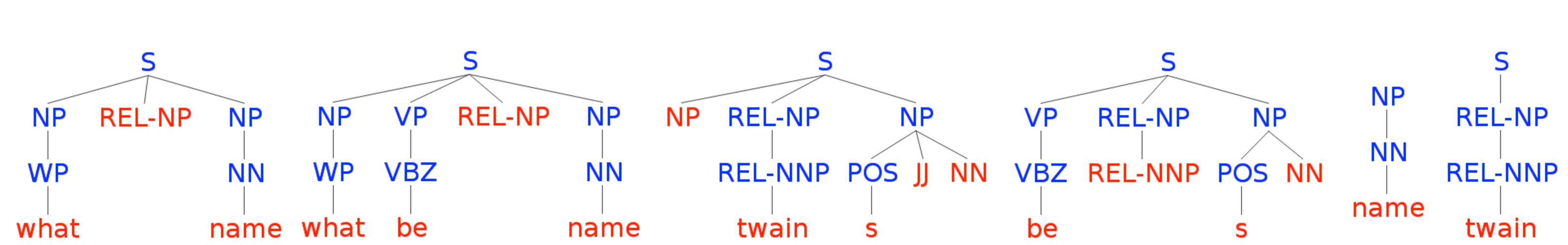
QA Pair classification with kernels

- Represent each QA pair as a triple of question and answer sentence trees and traditional similarity feature vector:
- Define QA similarity kernel between as follows: $\boldsymbol{x} = \langle \boldsymbol{T}_q, \boldsymbol{T}_s, \boldsymbol{v} \rangle$

$$\begin{aligned} K(\boldsymbol{x}^i,\boldsymbol{x}^j) &= K_{\mathrm{TK}}(\boldsymbol{T}_q^i,\boldsymbol{T}_q^j) \\ &+ K_{\mathrm{TK}}(\boldsymbol{T}_s^i,\boldsymbol{T}_s^j) \\ &+ K_{\mathrm{v}}(\boldsymbol{v}^i,\boldsymbol{v}^j) \end{aligned}$$

Tree Kernels

- Syntactic (STK) and Partial Tree Kernel (PTK) (Moschitti, 2006)
- PTK **generalizes STK** (Collins and Duffy, 2002) to generate more general tree fragments
- PTK is suitable for **constituency** and **dependency** structures

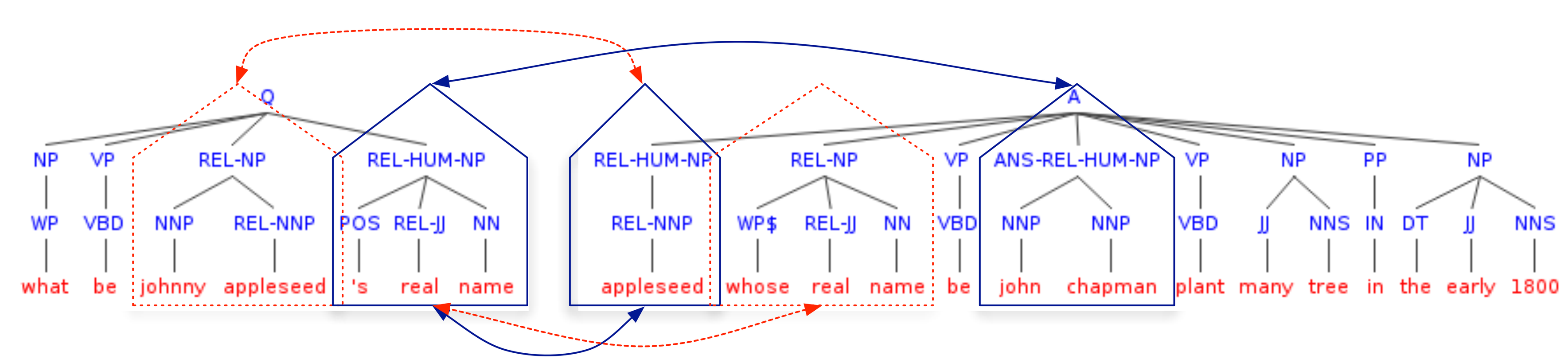


Semantic Linking

- Use Question Category (QC) and Focus Classifiers (FC) to find question category and focus word
- Run **NER** on the answer passage text
- Connect **focus word** with related NERs (according to the question category) in the answer

EAT	Named Entity types
HUM	Person
LOCATION	Location
ENTITY	Organization, Person, Misc
DATE	Date, Time, Number
QUANTITY	Number, Percentage
CURRENCY	Money, Number

Semantic linking with QC + FC + NER



Answer Extraction as Classification

- Consider a set **S** of all NP chunks as anwer candidates
- For each chunk in **S** generate a tree with a subtree containing c tagged with ANS tag
- TRAIN: Chunks spanning correct answers create +1 examples and the others are -1 -- can train a simply binary classifier
- TEST: predict the answer by taking the tree with the highest prediction score

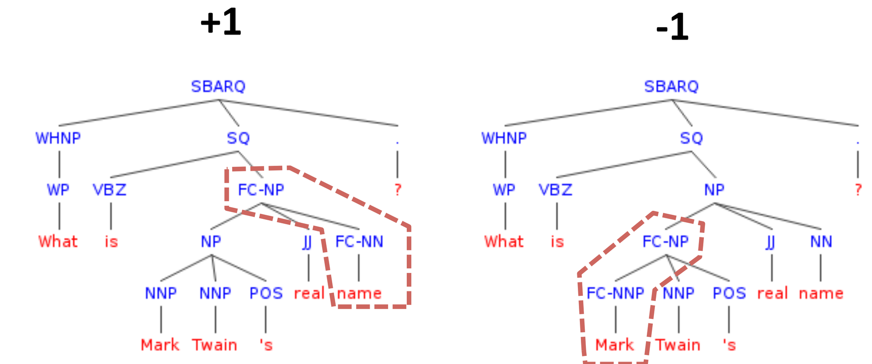
Question classifier

- Tree kernel SVM multi-classifier
- Data
  - 5500 questions from UIUIC
- 6 coarse classes:
  - ABBR, DESC, ENTY, HUM, LOC, NUM

Dataset	STK	PTK
UIUIC	**86.1**	82.2
TREC test	**79.3**	78.1

Focus classifier

+1 -1



Dataset	ST	STK	PTK
Mooney	73.0	**81.9**	80.5
SeCo-600	90.0	**94.5**	90.0
Bunescu	89.7	**98.3**	96.9

Answer sentence reranking on TREC 13

- Test on public benchmark TREC 13

System	MAP	MRR
Wang et al. (2007)	0.6029	0.6852
Heilman & Smith (2010)	0.6091	0.6917
Wang & Manning (2010)	0.5951	0.6951
Yao et al. (2013)	0.6319	0.7270
+ WN	0.6371	0.7301
shallow tree (S&M, 2012)	0.6485	0.7244
+ semantic tagging	**0.6781**	**0.7358**

Answer Extraction (simple majority voting)

- Replicate the setting in Yao et al., (2013) on TREC 13
- Best answer selection -- major voting scheme

system	set	P	R	F1
Yao et al. (2013)	TRAIN	55.7	43.8	49.1
+ forced		54.5	53.9	54.2
+ WN		55.2	53.9	54.5
this work		**66.2**	**66.2**	**66.2**
Yao et al. (2013)	ALL	67.2	50.6	57.7
+ forced		60.9	59.6	60.2
+ WN		63.6	62.9	63.3
this work		**70.8**	**70.8**	**70.8**

Sources of error

- **Chunking**: each answer is assumed to be spanned by a single chunk
- **Semantic linking**: relies on QC and NER
- **Candidate answer classification**: high recall, low precision introduce more semantic constraints on possible answers
- **Best answer selection**: naive major voting, ignores global view on the candidates

Conclusions

- Treat q/a pairs **directly** encoding them into linguistic structures augmented with semantic information
- Structural kernel technology to **automatically extract** and **learn** syntactic/semantic features
- **Semantic linking** using question and focus classifiers (trained with same tree kernel technology) and NERs
- Automatic feature engineering vs. previous TED approaches
- State-of-the-art results on two important QA tasks: **Answer Sentence Selection** and **Answer Extraction**

Acknowledgments

This research has been supported by the European Community's Seventh Framework Program (FP7/2007-2013) under the #288024 LIMOSINE project. Aliaksei Severyn is supported by Google Europe Doctoral Fellowship in Machine Learning.