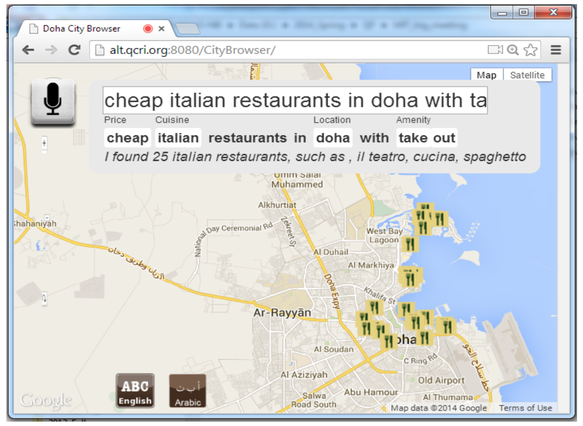
A Study of using Syntactic and Semantic Structures for Concept Segmentation and Labeling

(ALT-QCRI) *Iman Saleh, Shafiq Joty, Lluís Màrquez, Alessandro Moschitti, Preslav Nakov*
Scott Cyphers, James Glass (CSAIL-MIT)

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
Member of Qatar Foundation عضو في مؤسسة قطر

CSAIL

Motivation



How do we convert a spoken request like

"cheap italian restaurants in doha with take out"

into a database query?

Processing Steps

Text/Speech

"cheap italian restaurants in doha with take out"

↓ Semantic tagger

Semantically segmented text

(cheap) (italian) restaurants in (doha) with (take out)

<Price> <Cuisine> <City> <Amenity>

↓ Rule normalizer

Semantic representation

```
{"Price" : "low",
 "Amenity" : "take out",
 "City" : "doha",
 "Cuisine" : "italian"}
```

↓ Query builder

DB query

```
{$and [{"cuisine":{"$regex":"italian"}},
       {"city":{"$regex":"doha"}},
       {"price":{"$regex":"low"}},
       {"amenity":{"$regex":"carry out"}}]}
```

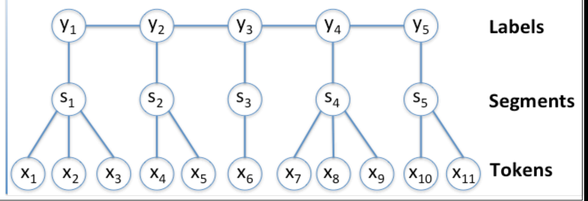
Computation Models for Semantic Tagging

State-of-the-art system

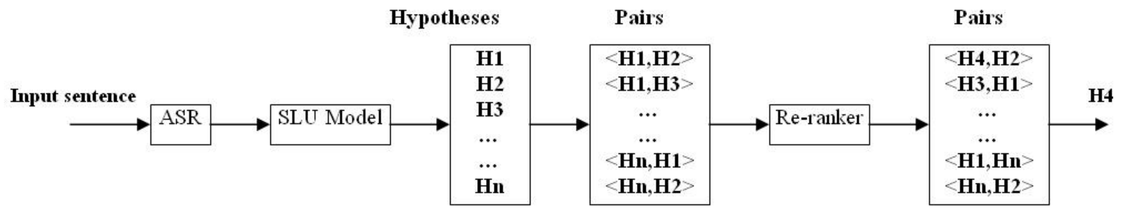
Joint sequential segmentation/classification

- **Semi-Markov CRFs** (Sarawagi & Cohen 04)
- Discriminative probabilistic sequential model
- Undirected graphical model

$$P(s \mid x) = \frac{1}{Z_\lambda(x)} \exp\{\sum_j \lambda_j f(y_{j-1}, s_j, x)\}$$

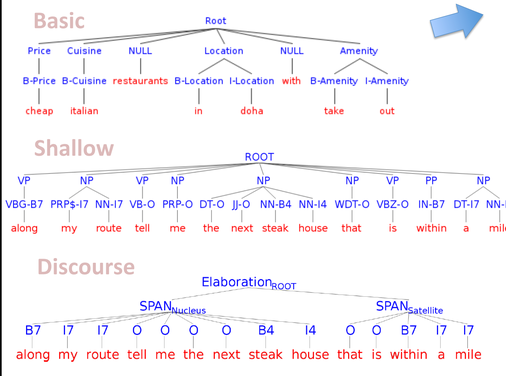


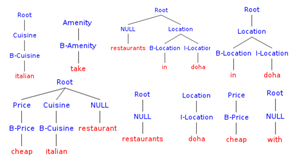
Our Approach: Reranking with Kernel Machines



- Semi-CRF generates *n*-best hypotheses
- SVMs and convolution tree kernels are then used to learn a reranking function
- Such function can choose the best hypothesis by exploiting structural representations

Hypothesis representations in form of semantic trees





Classification function with kernels

SVM: $H(\vec{x}) = \vec{w} \cdot \vec{x} + b = 0$

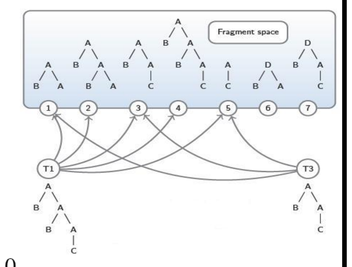
Dual: $\sum_{i=1..l} y_i \alpha_i \phi(o_i) \cdot \phi(o) + b = 0$

$K(o_i, o) = \langle \phi(o_i) \cdot \phi(o) \rangle$

Re-ranking Kernel:

$K(\langle H_1, H_2 \rangle, \langle H_1', H_2' \rangle) = S(H_1, H_1') + S(H_2, H_2') - S(H_1, H_2') - S(H_2, H_1')$

Common Tree Fragment Set



Dataset

Amazon Mechanical Turk was used to collect a corpus of sentences and gold-standard human annotations: 7,661 sentences (McGraw et al. 12)

Train	Test	Train RR	Test RR
6,922	1,521	28,482	7,605

Acknowledgements

This research is developed by the Arabic Language Technologies (ALT) group at Qatar Computing Research Institute (QCRI) within the Qatar Foundation in collaboration with MIT. It is part of the Interactive sYstems for Answer Search (**Iyas**) project.

Experiments

- N-best Oracle accuracy shows great headroom for improving on the CRF baseline

N	Oracle F1
1	83.03
2	87.76
5	92.63
10	95.23
100	98.72

- F1 scores for different tree kernels on basic tree

Subtree kernel	Syntactic tree kernel	Syntactic tree kernel + BOW	Partial tree kernel	Sequence kernel
84.08	83.19	83.20	**84.61**	82.93

- F1 scores for different feature combinations

	CRF Feat.	All Feat.	All Feat. + Basic Tree	All Feat. + Shallow Tree	All Feat. + Discourse Tree
F1	83.44	83.86	84.76	84.79	84.55
Rel. Imp	2.4%	4.8%	10.2%	10.3%	8.9%

Error analysis

	Other	Rating	Restaurant	Amenity	Cuisine	Dish	Hours	Location	Price
Other	8260	35	43	110	15	19	55	113	9
Rating	29	266	0	**14**	3	6	0	0	8
Restaurant	72	6	657	**20**	19	15	0	5	0
Amenity	117	9	10	841	**27**	**27**	7	12	7
Cuisine	36	2	12	**26**	543	**44**	3	1	0
Dish	23	0	4	**20**	**33**	324	1	4	0
Hours	61	0	1	2	6	1	426	9	1
Location	104	1	14	**20**	2	1	1	1457	0
Price	22	1	0	7	0	2	0	1	204

- Inaccuracies and inconsistencies in human annotations
- Requires lexical semantics and more coverage

Conclusions

- Structural kernels yield significant improvements.
- Partial tree kernel gives best results.
- Shallow tree is more helpful than other deep structures.
- Still large room for further improvement in the future.

References

[1] McGraw, S. Cyphers, P. Pasupat, J. Liu, and J. Glass. 2012. Automating crowd-supervised learning for spoken language systems. In Proceedings of INTERSPEECH 2012.

[2] M. Dinarelli, A. Moschitti, and G. Riccardi. Discriminative Reranking for Spoken Language Understanding. IEEE Transaction on Audio, Speech and Language Processing, 2012.

[3] S. Sarawagi, W. W. Cohen, Semi-Markov Conditional Random Fields for Information Extraction, in proceedings of NIPS, 2004 .