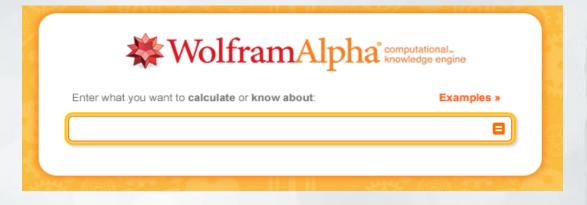# Question Classification in Question Answering

Brennen Byrne and Dustin Rodrigues

Introduction

- Question Answering - Why do we care?
- Current examples:

Introduction

- Question classification can help narrow the search space
- Only using the wh-word is not sufficient
- Building on the work of Li and Roth from UIUC
- Extensive use of Machine Learning to train on question and question answer features

Maximum Entropy Models

- Log-linear Models
- Convex
- Works Well for NLP
- Features Correspond to Constraints on Model

$$\log P(C|D, \lambda) = \sum_{(c,d) \in (C,D)} \log \frac{\exp \sum_i \lambda_i f_i(c, d)}{\sum_{c'} \exp \sum_j \lambda_j f_i(c, d)},$$

(1)

Question Classification Features

- Machine learning approach using SNoW (Sparse Network of Winnows) learning architecture
- Using the UIUC question classification dataset (5500 training and 500 test questions)
- Manually labeled with 6 coarse and 50 fine question types

Question Classification Features

Table 1: 6 coarse and 50 fine Question types defined in UIUC question dataset.

ABBR	letter	desc	NUM
abb	other	manner	code
exp	plant	reason	count
ENTITY	product	HUMAN	date
animal	religion	group	distance
body	sport	individual	money
color	substance	title	order
creative	symbol	desc	other
currency	technique	LOC	period
dis.med.	term	city	percent
event	vehicle	country	speed
food	word	mountain	temp
instrument	DESC	other	size
lang	definition	state	weight

UIUC Question Types

Question Classification Features

- Question wh-word
- Head word
  - uses syntactic parser and Collins rules
- WordNet Hypernym
  - uses Lesk algorithm
- Unigram words
- Word shape
  - 5 categories

Question Classification Experiments

- UIUC Training and Test Set
- Tried Coarse Classes and Fine
- Added one feature set at a time
- Removed feature sets individually
- More compact, more informative feature set
- 10 cross validation

	6 class	50 class
wh-word	46.0	46.8
+ head word	92.2	82.0
+ hypernym	91.8	85.6
+ unigram	93.0	88.4
+ word shape	93.6	89.0

	6 class	50 class
overall	93.6	89.0
- wh-word	93.6	89.0
- head word	92.8	88.2
- hypernym	90.8	84.2
- unigram	93.6	86.8
- word shape	93.0	88.4

Question Answer Features

- Extract binary features from question and candidate answer pair
- Named Entities (NE) presence
  - Used Stanford NER
  - Only four entities
    - person, location, organization, misc
- Numerical Entities (NUM) presence
  - 13 compiled patterns - one for each type of numerical answer

Question Answer Features

- Dictionary Entities (DIC) presence
  - 101 dictionary files from Ephyra project
  - Checks if the head word is the name of a dictionary
  - Star Wars: Episode I – The Phantom Menace vs Star Wars Episode I: The Phantom Menace
    - edit distance solves this problem

Question Answer Features

- Specific Features (SPE)
  - 8 features for particular questions
- Dependency Validity Features (DEP)
  - looks at grammatical paths from question word and candidate answer to common word

Question Answer Experiments

- How important is the question classifier?
- To Test, Get 20 Results
- Create Data Point
- True or False
- Use Data to Create Maximum Entropy Model
- Evaluate With:
  - Mean Reciprocal Rank
  - Top 1 Prediction Accuracy
  - Top 5 Prediction Accuracy

Question Answer Experiments

- Create 3 Question Classifiers (QC1, QC2, QC3)
- Test Rankings of Individual Features
- Test Classifiers against One Another

Table 6: Performance of individual and incremental feature sets for three question classifiers.

Feature	MRR			Top1			Top5		
	QC1	QC2	QC3	QC1	QC2	QC3	QC1	QC2	QC3
Individual									
Baseline	49.9	49.9	49.9	40.1	40.1	40.1	59.4	59.4	59.4
DIC	49.5	49.5	49.5	42.6	42.6	42.6	60.4	60.4	60.4
NE	48.5	47.5	50.0	40.6	40.6	42.6	61.9	60.9	63.4
NE-4	49.5	48.5	51.5	41.6	42.1	44.6	62.4	61.9	64.4
REG	52.0	54.0	54.0	44.1	47.0	47.5	64.4	65.3	65.3
SPE	55.0	55.0	55.0	48.5	48.5	48.5	64.4	64.4	64.4
DEP	51.0	51.5	52.0	43.6	44.1	44.6	65.3	65.8	65.8
Incremental									
Baseline	49.9	49.9	49.9	40.1	40.1	40.1	59.4	59.4	59.4
+DIC	49.5	49.5	49.5	42.6	42.6	42.6	60.4	60.4	60.4
+NE	50.0	48.5	51.0	43.1	42.1	44.6	62.9	61.4	64.4
+NE-4	51.5	50.0	53.0	44.1	43.6	46.0	63.4	62.9	65.8
+REG	55.0	56.9	59.9	48.0	51.0	54.0	68.3	68.8	71.8
+SPE	60.4	62.4	65.3	55.4	58.4	61.4	70.8	70.8	73.8
+DEP	61.4	62.9	**66.3**	55.9	58.4	**62.4**	71.8	71.8	**73.8**

Conclusion

Show That Incorrect Classification Makes It Harder to Answer a Question

i.e. Question Classifiers are Important to Question Answering

Also Created a Question Classifiers that Outperforms Others