# A Graph-based Semi-Supervised Learning for Question-Answering

Yi ZHU

Celikyilmaz, A. and Thint, M. and Telecom, B. and Huang, Z.
In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, 2009

January 27, 2010

# Outline

# Objective

1. Ranking candidate answering sentences for question, so we can find the best answer for one question.

# Objective

1. Ranking candidate answering sentences for question, so we can find the best answer for one question.
2. Ranking Question-Answering pairs (score of ranking).

# The Model

1. Using textual entailment analysis, we obtain entailment features between a question posed by the user and the candidate sentences returned from search engine.

# The Model

1. Using textual entailment analysis, we obtain entailment features between a question posed by the user and the candidate sentences returned from search engine.
2. Use this score as the similarity measurements, train with the Graph-based Semi-Supervised learning.

# Question-Type-Candidate Sentence NER match feature

1. answer contains the fine NER: 1

# Question-Type-Candidate Sentence NER match feature

1. answer contains the fine NER: 1
2. answer contains the coarse NER: 0.5

# Question-Type-Candidate Sentence NER match feature

1. answer contains the fine NER: 1
2. answer contains the coarse NER: 0.5
3. no NER match: 0

## Question component match features

1. example:

# Question component match features

1. example:
   - Q: When did Nixon die?

# Question component match features

1. example:
   - Q: When did Nixon die?
   - A: Richard Nixon, 37th President of USA, passed away of stroke on April 22, 1994.

# Question component match features

1. example:
   - Q: When did Nixon die?
   - A: Richard Nixon, 37th President of USA, passed away of stroke on April 22, 1994.
2. Head-Match (die $\rightarrow$ passed away)

# Question component match features

1. example:
   - Q: When did Nixon die?
   - A: Richard Nixon, 37th President of USA, passed away of stroke on April 22, 1994.
2. Head-Match (die $\rightarrow$ passed away)
3. Subject-Match (Nixon $\rightarrow$ Richard Nixon)

## Question component match features

1. example:
   - Q: When did Nixon die?
   - A: Richard Nixon, 37th President of USA, passed away of stroke on April 22, 1994.
2. Head-Match (die $\rightarrow$ passed away)
3. Subject-Match (Nixon $\rightarrow$ Richard Nixon)
4. Object-Match (–)

## Question component match features

1. example:
   - Q: When did Nixon die?
   - A: Richard Nixon, 37th President of USA, passed away of stroke on April 22, 1994.
2. Head-Match (die $\rightarrow$ passed away)
3. Subject-Match (Nixon $\rightarrow$ Richard Nixon)
4. Object-Match (–)
5. Modifier-Match ([X] $\rightarrow$ April 22, 1994)

## Lexico-Syntactic Alignment Features

1. Noun and verb counts;

# Lexico-Syntactic Alignment Features

1. Noun and verb counts;
2. matching synonyms in WordNet;

## Lexico-Syntactic Alignment Features

1. Noun and verb counts;
2. matching synonyms in WordNet;
3. Verb match statistics using WordNet;

## Lexico-Syntactic Alignment Features

1. Noun and verb counts;
2. matching synonyms in WordNet;
3. Verb match statistics using WordNet;
4. ...

# The Model

1. Use textual entailment analysis, we obtain entailment features between a question posed by the user and the candidate sentences returned from search engine.
2. Use this score as the similarity measurements, train with the Graph-based Semi-Supervised learning.

# Semi-Supervised learning

1. Data point: $X = \{x_1, \ldots, x_n\}$;

## Semi-Supervised learning

1. Data point: $X = \{x_1, \ldots, x_n\}$;
2. Label: $Y = \{y_1, \ldots, y_n\}$;

## Semi-Supervised learning

1. Data point: $X = \{x_1, \ldots, x_n\}$;
2. Label: $Y = \{y_1, \ldots, y_n\}$;
3. Labeled data: $L = \{(x_1, y_1), \ldots, (x_l, y_l)\}$;

## Semi-Supervised learning

1. Data point: $X = \{x_1, \ldots, x_n\}$;
2. Label: $Y = \{y_1, \ldots, y_n\}$;
3. Labeled data: $L = \{(x_1, y_1), \ldots, (x_l, y_l)\}$;
4. Unlabeled data: $U = \{x_1, \ldots, x_u\}, l + u = n$.

## Semi-Supervised learning

1. Data point: $X = \{x_1, \ldots, x_n\}$;
2. Label: $Y = \{y_1, \ldots, y_n\}$;
3. Labeled data: $L = \{(x_1, y_1), \ldots, (x_l, y_l)\}$;
4. Unlabeled data: $U = \{x_1, \ldots, x_u\}, l + u = n$.
5. Objective: estimate the true label for the unlabeled data $U$.

# Graph-based Semi-Supervised learning

1. Suppose we construct a graph $G(V, E)$;

# Graph-based Semi-Supervised learning

1. Suppose we construct a graph $G(V, E)$;
2. Each node in $V$ represents a data point $x$;

## Graph-based Semi-Supervised learning

1. Suppose we construct a graph $G(V,E)$;
2. Each node in $V$ represents a data point $x$;
3. If two data points $x_i$ and $x_j$ are similar with each other, we connect them with an edge $e_{ij}$

## Graph-based Semi-Supervised learning

1. Suppose we construct a graph $G(V,E)$;
2. Each node in $V$ represents a data point $x$;
3. If two data points $x_i$ and $x_j$ are similar with each other, we connect them with an edge $e_{ij}$
4. The weight $w_{ij}$ on the edge $e_{ij}$ represent the similarity between $x_i$ and $x_j$;

## Harmonic Function

1. Our goal is to find out a label function $f(x)$ to predict the label of the unlabeled data;

# Harmonic Function

1. Our goal is to find out a label function $f(x)$ to predict the label of the unlabeled data;
2. This function need to satisfy two condition:

# Harmonic Function

1. Our goal is to find out a label function $f(x)$ to predict the label of the unlabeled data;
2. This function need to satisfy two condition:
   - $f$ is close to the labeled data: $f(x_i) \approx y_i, x_i \in L$;

# Harmonic Function

1. Our goal is to find out a label function $f(x)$ to predict the label of the unlabeled data;
2. This function need to satisfy two condition:
   - $f$ is close to the labeled data: $f(x_i) \approx y_i, x_i \in L$;
   - $f$ is smooth on the graph.

# Harmonic Function

1. So the objective function is:

$$\min_{f:f(x)\in\Re} \infty \sum_{i=1}^{l} (y_i - f(x_i))^2 + \sum_{i,j=1}^{l+u} w_{ij}(f(x_i) - f(x_j))^2;$$

## Harmonic Function

1. To solve this problem, we could use the Laplacian matrix of the graph: $\Delta = D - W$;

## Harmonic Function

1. To solve this problem, we could use the Laplacian matrix of the graph: $\Delta = D - W$;
2. $W$ is the weight matrix,
   $D$ is a diagonal matrix : $D_{ii} = \sum_{j=1}^{n} w_{ij}$;

## Harmonic Function

1. To solve this problem, we could use the Laplacian matrix of the graph: $\Delta = D - W$;
2. $W$ is the weight matrix,
   $D$ is a diagonal matrix : $D_{ii} = \sum_{j=1}^{n} w_{ij}$;
3. Denote $\Delta = \begin{pmatrix} \Delta_{LL} & \Delta_{UL} \\ \Delta_{LU} & \Delta_{UU} \end{pmatrix}$

## Harmonic Function

1. To solve this problem, we could use the Laplacian matrix of the graph: $\Delta = D - W$;
2. $W$ is the weight matrix,
   $D$ is a diagonal matrix : $D_{ii} = \sum_{j=1}^{n} w_{ij}$;
3. Denote $\Delta = \begin{pmatrix} \Delta_{LL} & \Delta_{UL} \\ \Delta_{LU} & \Delta_{UU} \end{pmatrix}$
4. Solution: $\begin{cases} f_L = y_L, \\ f_U = -{\Delta_{UU}}^{-1} \Delta_{UL} y_l. \end{cases}$

# Our case

1. Label: the task achievement score;

# Our case

1. Label: the task achievement score;
2. Data point: one Q/A pair with previous mentioned features;

# Our case

1. Label: the task achievement score;
2. Data point: one Q/A pair with previous mentioned features;
3. Labeled data: label manually;

## Our case

1. Label: the task achievement score;
2. Data point: one Q/A pair with previous mentioned features;
3. Labeled data: label manually;
4. Similarity: $w_{ij} = 1 - \sum_{q=1}^{d} \frac{|x_{iq} - x_{jq}|}{d}$

# Year 2007 Question

| bus line | time | time | location | location | count |
|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 209 |
| 0 | 1 | 1 | 1 | 0 | 49 |
| 0 | 1 | 0 | 0 | 0 | 719 |
| 1 | 1 | 0 | 1 | 0 | 116 |
| 1 | 0 | 0 | 0 | 0 | 2770 |
| 0 | 0 | 0 | 0 | 0 | 5163 |
| 1 | 1 | 0 | 0 | 0 | 252 |
| 1 | 0 | 0 | 1 | 0 | 198 |
| 0 | 0 | 0 | 1 | 0 | 597 |
| 0 | 1 | 1 | 0 | 0 | 96 |
| 1 | 1 | 0 | 1 | 1 | 59 |
| 1 | 1 | 1 | 0 | 0 | 65 |
| 0 | 1 | 0 | 1 | 1 | 91 |
| 1 | 0 | 0 | 1 | 1 | 96 |
| 0 | 1 | 1 | 1 | 1 | 24 |
| 0 | 0 | 0 | 1 | 1 | 192 |
| 1 | 1 | 1 | 1 | 1 | 29 |
| 1 | 1 | 1 | 1 | 0 | 37 |

# Year 2007 Answer

| bus line | time | time | location | location | count |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 2754 |
| 0 | 0 | 0 | 1 | 1 | 429 |
| 1 | 1 | 1 | 0 | 0 | 851 |
| 0 | 0 | 0 | 0 | 0 | 1186 |
| 1 | 1 | 1 | 1 | 1 | 2245 |
| 0 | 1 | 1 | 0 | 0 | 341 |
| 0 | 0 | 0 | 1 | 0 | 699 |
| 1 | 1 | 0 | 1 | 1 | 236 |
| 1 | 0 | 0 | 1 | 0 | 59 |
| 0 | 1 | 1 | 1 | 0 | 378 |
| 0 | 1 | 1 | 1 | 1 | 447 |
| 1 | 1 | 0 | 1 | 0 | 414 |
| 0 | 1 | 0 | 1 | 1 | 168 |
| 0 | 1 | 0 | 0 | 0 | 333 |
| 1 | 1 | 0 | 0 | 0 | 31 |
| 0 | 1 | 0 | 1 | 0 | 128 |
| 1 | 0 | 0 | 0 | 0 | 55 |
| 1 | 0 | 0 | 1 | 1 | 8 |

# **Year** 2007 – 2009 **Question**

| bus line | time | time | location | location | count |
|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 679 |
| 0 | 1 | 1 | 1 | 0 | 150 |
| 0 | 1 | 0 | 0 | 0 | 2718 |
| 1 | 1 | 0 | 1 | 0 | 411 |
| 1 | 0 | 0 | 0 | 0 | 11638 |
| 0 | 0 | 0 | 0 | 0 | 20772 |
| 1 | 1 | 0 | 0 | 0 | 969 |
| 1 | 0 | 0 | 1 | 0 | 926 |
| 0 | 0 | 0 | 1 | 0 | 2877 |
| 0 | 1 | 1 | 0 | 0 | 278 |
| 1 | 1 | 0 | 1 | 1 | 237 |
| 1 | 1 | 1 | 0 | 0 | 173 |
| 0 | 1 | 0 | 1 | 1 | 273 |
| 1 | 0 | 0 | 1 | 1 | 392 |
| 0 | 1 | 1 | 1 | 1 | 94 |
| 0 | 0 | 0 | 1 | 1 | 831 |
| 1 | 1 | 1 | 1 | 1 | 83 |
| 1 | 1 | 1 | 1 | 0 | 120 |

# Year 2007 – 2009 Answer

| bus line | time | time | location | location | count |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 8848 |
| 0 | 0 | 0 | 1 | 1 | 1766 |
| 1 | 1 | 1 | 0 | 0 | 3294 |
| 0 | 0 | 0 | 0 | 0 | 7007 |
| 1 | 1 | 1 | 1 | 1 | 8072 |
| 0 | 1 | 1 | 0 | 0 | 1280 |
| 0 | 0 | 0 | 1 | 0 | 3740 |
| 1 | 1 | 0 | 1 | 1 | 1280 |
| 1 | 0 | 0 | 1 | 0 | 429 |
| 0 | 1 | 1 | 1 | 0 | 1243 |
| 0 | 1 | 1 | 1 | 1 | 1573 |
| 1 | 1 | 0 | 1 | 0 | 1663 |
| 0 | 1 | 0 | 1 | 1 | 418 |
| 0 | 1 | 0 | 0 | 0 | 1881 |
| 1 | 1 | 0 | 0 | 0 | 117 |
| 0 | 1 | 0 | 1 | 0 | 447 |
| 1 | 0 | 0 | 0 | 0 | 248 |
| 1 | 0 | 0 | 1 | 1 | 315 |

# Other results

1. if $Q + A = \{1\ 1\ 1\ 1\ 1\}$, then we believe that the system answered what user want.
2. 2007 Q-A match: $2446/10762 = 0.2273$
3. $2007 - 2009$ Q-A match: $8806/43621 = 0.2019$

# Thanks

Q & A

January 27, 2010