Distant Supervision for Relation Extraction with Matrix Completion

By. **Miao Fan**
Ph.D. Candidate
Dept. of C.S.
Tsinghua University
fanmiao.cslt.thu@gmail.com

1. Introduction

- Distant Supervision for Relation Extraction

Motto	*Die Luft der Freiheit weht* (German)[1]
Motto in English	The wind of freedom blows [1]
Established	1891[2]
Type	Private
Endowment	$17.04 billion [3]
President	John L. Hennessy
Provost	John Etchemendy
Academic staff	1,995[4]
Admin. staff	10,979[5] excluding SHC
Students	15,870
Undergraduates	6,999[6]
Postgraduates	8,871[6]
Location	Stanford, California, U.S.
Campus	Suburban, 8,180 acres (3,310 ha)[7]
Colors	Cardinal and white
Athletics	NCAA Division I (FBS) Pac-12
Nickname	Cardinal
Mascot	Stanford Tree (unofficial)
Website	Stanford.edu

Stanford University

Structured Data (Knowledge base)

Alignment

Leland Stanford Junior University, commonly referred to as **Stanford U** in Stanford, California in the northwestern Silicon Valley on an 8,180-acre universities in the world.[10][11][12][13]

Leland Stanford, governor of and U.S. senator from California and leading 1891 in memory of their son, Leland Stanford, Jr., who died of typhoid two coeducational and nondenominational institution. Tuition was free until the death and after much of the campus was damaged by the 1906 San Franc faculty and graduates' entrepreneurialism to build self-sufficient local indus to a linear accelerator, and was one of the original four ARPANET nodes (

Today, the University comprises various academic components and has n academic schools of Humanities and Sciences and Earth Sciences, as we Medicine, with a student body of approximately 7,000 undergraduates and university.[4] Moreover, it has produced the largest number of Turing Awar billionaires and 17 astronauts, and is one of the leading producers of mem prominent companies including Google, Hewlett-Packard, Nike, Sun Micro more than $2.7 trillion in annual revenue, equivalent to the 10th-largest ec Luther King, Jr.[18] Its most recent acceptance rate, 5.69% for the Class of

Stanford competes in 34 varsity sports and is one of two private universitie championships (the second-most for a university), including at least one fo Directors' Cup every year since 1995.[20][21] Stanford athletes have won me 129 of them gold. In the 2008 Beijing Olympic Games, Stanford won more total medals won, would have tied with Japan for 11th place.[22][23][24]

Free text (unstructured data)

1. Introduction

- **Basic Assumption:**
  - If **two entities** (PER, ORG, LOC) participating in **a relation** in the knowledge base (**Freebase, Wikipedia Info-box**) **co-occur** in the same sentence. We believe that the sentence expresses that relation to some extend.
  - I.E.
    - **Free text: Barack Obama is the 44th president of U.S.A**

    - **Knowledge base: President of (Barack Obama, U.S.A)**

    - **Feature:** is the 44th president of (Token,POS, DEP)
    - **Label:** President of

1. Introduction

Multi-label Classification Problem with Sparse and Noise features!!!

Entity Pair	<Barack Obama, U.S.>
Relation Instances (Knowledge base)	1. President of (Barack Obama, U.S.) 2. Born in (Barack Obama, U.S.)
Relation Mentions (Free text)	1. *Barack Obama* is the 44th and current President of the *U.S.* (President of) 2. *Barack Obama* ended *U.S.* military involvement in the Iraq War. (-) 3. *Barack Obama* was born in Honolulu, Hawaii, *U.S.* (Born in) 4. *Barack Obama* ran for the *U.S.* Senate in 2004. (Senate of)

1. Introduction

Problem Formulation

d dimension of features(POS , DEP, NER)

	Feature_1	...	Feature_d
Item_1 (Obama, U.S.A)	1(was born in)	0	1(the 44th president of)
Item_2	...		
Item_3			
...	1		0
Item_n			

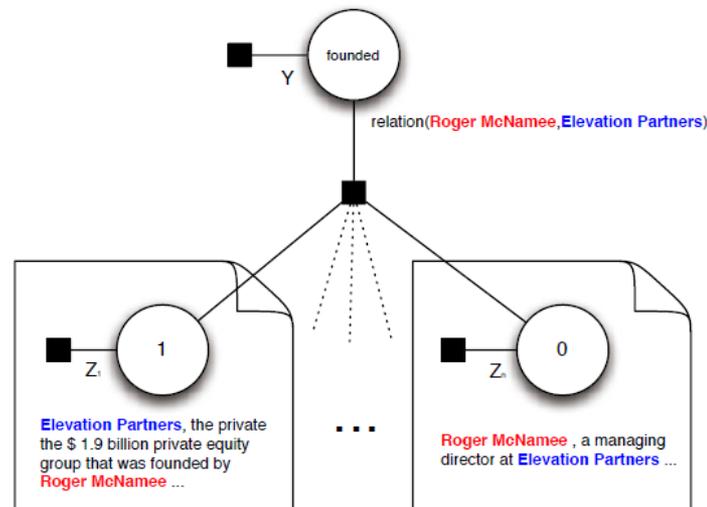| Label_1 (born in)	...	Label_t(President)
1	0	1
1		
0		
1		
1		

t relation labels

1. Introduction

- Matrix Completion

  - Suitable to deal with the problem!

$$Z = \left[\begin{array}{cc} X_{train} & Y_{train} \\ X_{test} & Y_{test} \end{array} \right] :$$

To be completed

2. Related Work

- We will skip the details.
    - To tolerate feature noise, **MIL** (Multi-instance Learning) is the mainstream.
    - However, the assumptions will be complex.

3. Model

- 3.1 DRMC-b

$$\underset{\mathbf{Z},\mathbf{b}}{\text{argmin}} \quad \mu||\mathbf{Z}||_* + \frac{1}{|\Omega_X|} \sum_{(i,j)\in\Omega_X} C_x(z_{ij}, x_{ij})$$

$$+ \frac{\lambda}{|\Omega_Y|} \sum_{(i,j)\in\Omega_Y} C_y(z_{i(j+d)} + b_j, y_{ij}) \tag{5}$$

3. Model

- 3.2 DRMC-1

$$\operatorname*{argmin}_{\mathbf{Z}} \ \mu \|\mathbf{Z}\|_* + \frac{1}{|\Omega_X|} \sum_{(i,j)\in\Omega_X} C_x(z_{i(j+1)}, x_{ij})$$

$$+ \frac{\lambda}{|\Omega_Y|} \sum_{(i,j)\in\Omega_Y} C_y(z_{i(j+d+1)}, y_{ij})$$

$$s.t. \qquad \mathbf{Z}(:,1) = \mathbf{1}$$

(6)

3. Model

- 3.3 Loss Function

    - For binary features and labels

$$C(u, v) = log(1 + e^{-uv})$$

3. Model

- 3.4 Predict Function

$$y_{ij} \in Y_{test}$$

$$P(r_j | ep_i) = \frac{1}{1 + e^{-y_{ij}}}$$

Then we can achieve a global ranking based on the conditional probability of r_j given entity-pair_i, P/R Curve could be drawn!

4. Algorithm

Algorithm 1 FPC algorithm for solving DRMC-b

Input:
Initial matrix $\mathbf{Z_0}$, bias $\mathbf{b_0}$; Parameters μ, λ;
Step sizes τ_z, τ_b

Set $\mathbf{Z} = \mathbf{Z_0}, \mathbf{b} = \mathbf{b_0}$
foreach $\mu = \mu_1 > \mu_2 > ... > \mu_F$ **do**
  while *relative error* $> \varepsilon$ **do**
    Gradient step:
    $\mathbf{A} = \mathbf{Z} - \tau_z g(\mathbf{Z}), \mathbf{b} = \mathbf{b} - \tau_b g(\mathbf{b})$
    Shrinkage step:
    $\mathbf{U\Sigma V^T} = SVD(\mathbf{A}),$
    $\mathbf{Z} = \mathbf{U}max(\mathbf{\Sigma} - \tau_{\mathbf{z}}\mu, 0)\mathbf{V^T}$
  end while
end foreach

Output: Completed Matrix $\mathbf{Z}$, bias $\mathbf{b}$

Algorithm 2 FPC algorithm for solving DRMC-1

Input:
Initial matrix $\mathbf{Z_0}$; Parameters μ, λ;
Step sizes τ_z

Set $\mathbf{Z} = \mathbf{Z_0}$
foreach $\mu = \mu_1 > \mu_2 > ... > \mu_F$ **do**
  while *relative error* $> \varepsilon$ **do**
    Gradient step: $\mathbf{A} = \mathbf{Z} - \tau_z g(\mathbf{Z})$
    Shrinkage step: $\mathbf{U\Sigma V^T} = SVD(\mathbf{A}),$
    $\mathbf{Z} = \mathbf{U}max(\mathbf{\Sigma} - \tau_{\mathbf{z}}\mu, 0)\mathbf{V^T}$
    Projection step: $\mathbf{Z}(:, 1) = \mathbf{1}$
  end while
end foreach

Output: Completed Matrix $\mathbf{Z}$

5. Experiments

- Datasets

- Parameter Estimation(Rank Optimization)

- Evaluation

- [..\..\科研\Drafts\Distant Supervision Relation Extraction with Matrix Completion\ACL 2014\figures](..\..\科研\Drafts\Distant Supervision Relation Extraction with Matrix Completion\ACL 2014\figures)