ENTITY LINKING ON MICROBLOGS
WITH SPATIAL AND TEMPORAL SIGNALS

Yuan Fang *
Ming-Wei Chang

Institute for Infocomm Research, Singapore
Microsoft Research, USA

10/26/2014

* Work done while a student at Univ of Illinois at Urbana-Champaign and intern at Microsoft Research.
Problem

**Entity Linking in Microblogs:** Map entity mentions in a short message (e.g. a tweet, facebook messages) into predefined entities (e.g. entries in Wikipedia).

US secretary of state *Clinton* is hospitalized due to …

http://en.wikipedia.org/wiki/Hillary_Rodham_Clinton
http://en.wikipedia.org/wiki/United_States
Why is entity linking in microblogs important?

- Motivation: intelligence gathering (market/disaster/politics)
- But word-based matching is ineffective due to **ambiguity**
  - Noisy & informal: in-depth NLP analysis is difficult
  - Short: insufficient contexts

“Washington”?

“Spurs”? 
Why is entity linking in microblogs important?

• Motivation: intelligence gathering (market/disaster/politics)
• But word-based matching is ineffective due to **ambiguity**
  • Noisy & informal: in-depth NLP analysis is difficult
  • Short: insufficient contexts

Which "washington"?

1) Different peaks → Different entities?
2) A single peak → A mixture of entities?
Proposed Approach

Leveraging spatiotemporal signals to improve entity linking
Observation & Intuition

• Intuition 1: Spatiotemporal signals
  • Entity prior changes over time or space

• Intuition 2: Easier surface forms
  • Inter-tweet interactions

“spurs” → SA Spurs
91% in US vs. 8% in UK

“Clinton” vs. “Hillary Clinton”
Proposal: Spatiotemporal entity linking

\[ e^* = \arg \max_{e \in E} p(e|m,a,t,l) \]

\[ = \arg \max_{e \in E} p(e,m,a,t,l) \]

\[ = \arg \max_{e \in E} p(m,a|t,l,e)p(t,l,e) \]

\[ = \arg \max_{e \in E} p(m,a|e)p(e|t,l) \]

\[ = \arg \max_{e \in E} p(e|m,a)p(e|t,l)/p(e) \]

\( m \): target message (e.g. a tweet)
\( a \): anchor text (surface form)
\( t \): time – when \( m \) was published
\( l \): location – where \( m \) was published

Cond. Indep. Assumption

Intuition: update entity priors

if \( e \)'s prior at \( t, l \) is higher than its unconditioned prior, we make \( e^* = e \) more likely.
Predicting the entity

\[ e^* = \arg \max_{e \in E} \frac{p(e|m, a)p(e|t, l)}{p(e)} \]

- \( m \): target message (e.g. a tweet)
- \( a \): anchor text (surface form)
- \( t \): time – when \( m \) was published
- \( l \): location – where \( m \) was published

some existing model without using spatiotemporal signals

Wikipedia pageview statistics
Challenges: Estimating $p(e|t, l)$

Challenge 1
Lack of large-scale entity annotations

- Use an existing model to tag **unlabeled tweets** (with time/location)
- **Aggregate tweets** tagged with $e$ at time $t$/location $l$
- **Update prior** $p(e|t, l)$ based on the aggregated tweets
- **Update the model** with the estimated $p(e|t, l)$
Challenges: Estimating $p(e|t, l)$

**Challenge 2**
How to handle continuous $t, l$?

We **discretize** $t, l$ into bins over time and location
- Time bins: some fixed interval (per day, hour, etc.)
- Location bins: latitude / longitude grids

**Granularity** of bins
- Too small $\rightarrow$ not enough samples in a bin
- Too large $\rightarrow$ spatiotemporal signals become less helpful

Solution: fine granularity + **smoothing**
Smoothing over bins

- Study how a tweet is written
  - There is an \( \epsilon \) probability to spontaneously write a tweet
  - There is an \( 1 - \epsilon \) chance of imitate a tweet in a near by time/location bin
  - Imitating from which time/location bin follows a polynomial decay

\[
p(e|\delta) = \epsilon \cdot \rho_{e\delta} + (1 - \epsilon) \sum_{\delta'} \beta_{\delta'|\delta} p(e|\delta')
\]

\( \rho_{e\delta} \): estimate with existing algorithm in bin \( \delta \)

\( \beta_{\delta'|\delta} \propto (d + |\delta - \delta'|)^\lambda \) (polynomial decay)
Conditional independence assumption

• Data scarcity more severe if we use bins over \((t, l)\) jointly
• Assume conditional independence
  • Binning over time / location independently

\[ e^* = \arg \max_{e \in E} \frac{p(e|t) p(e|l)}{p(e)} \frac{p(e|m, a)}{p(e|m, a)} \]
Empirical Study

Quantitative Results and Case Study
Dataset

• Tweets
  • One month: Dec 2012
  • Focus on tweets from verified users
  • Only keep tweets in English and with locations in the United States
  • Discard retweets

• 1.8 million tweets in total
  • Entity priors over time/locations are bootstrapped from them
Evaluation methodology

• **IE-driven evaluation**
  • Uniformly sample 500 tweets (250 dev + 250 test)
  • Metric: macro F-score [NAACL13]

• **IR-driven evaluation**
  • Important for many applications
    • e.g. sentiment analysis for a product
  • Select ten query entities
    • Sample 100 tweets for each query entity
    • Total 1000 tweets
    • Labeled each to indicate whether it mentions the query entity or not
  • Metric: macro F-score, but only consider the query entity

| Ten entities                                |
|---------------------------------------------|
| Newtown, Connecticut                       |
| Big Bang (South Korean band)                |
| Les Misérables (2012 film)                  |
| Winter solstice                             |
| San Antonio Spurs                           |
| Hillary Rodham Clinton                      |
| Catherine, Duchess of Cambridge             |
| Washington (state)                          |
| Hanukkah                                    |
| Django unchained (2012 film)                |
Algorithm settings

• Baseline: E2E [NAACL 2013]
  • State-of-the-art
  • Learn to jointly detect mention and disambiguate entities
  • SVM trained with independent data
  • Convert output to probability by minimizing cross entropy on dev set

• Baseline: LP (link probability)
  • Link probability in Wikipedia articles
  • Choose mention detection threshold by minimizing cross entropy on dev set

• Our algorithm
  • Tune parameters on dev set
A) Are the baselines good enough?

|       | Precision | Recall | F1  |
|-------|-----------|--------|-----|
| Wikiminer | 78.9      | 24.7   | 37.6|
| Illinois | 77.3      | 34.9   | 48.1|
| LP      | 49.7      | 47.0   | 48.3|
| E2E     | 85.5      | 42.8   | 57.0|
B) Are spatiotemporal signals useful?

|                | IE-driven | IR-driven | 64.9 | 71.4 | 65.0 | 76.1 | 68.6 | 79.0 |
|----------------|-----------|-----------|------|------|------|------|------|------|
| E2E            | 57.0      | 58.4      |      |      |      |      |      |      |
| + Time         | 64.9      | 71.4      |      |      |      |      |      |      |
| + Location     | 65.0      | 76.1      |      |      |      |      |      |      |
| + Both         | **68.6**  | **79.0**  |      |      |      |      |      |      |

|                | IE-driven | IR-driven | 48.3 | 48.5 | 52.4 | 59.7 | 50.3 | 61.8 | 49.0 | 53.3 |
|----------------|-----------|-----------|------|------|------|------|------|------|------|------|
| LP             | 48.3      | 48.5      |      |      |      |      |      |      |
| + Time         | **52.4**  | **59.7**  |      |      |      |      |      |      |
| + Location     | 50.3      | **61.8**  |      |      |      |      |      |      |
| + Both         | 49.0      | 53.3      |      |      |      |      |      |      |

(a) Macro F-scores
C) Graph-based smoothing

(a) Base system: E2E

(b) Base system: LP
D) Case Study: More informative time profiling

Target entity: Washington (state)

Time profiling for keyword "washington"

Are all these peaks for Washington state?

Time profiling for “washington” entities

(1) Washington (state): legalization of marijuana
(2) Washington, D.C.: fiscal cliff + winter weather alert
(3) Washington Redskins: Game for division title
Conclusion & future work

• We demonstrated that
  • Spatiotemporal signals are critical in advancing entity linking
  • Aggregation of many (individually) noisy tweets help

• Future work
  • A more general framework to incorporate more non-text meta data
  • Online updating of spatiotemporal model
  • Of course, improve the base model!

We made some improvement to the base model $p(e|m, a)$