Towards Lexical Semantic Analysis of Tweets

Paul Cook

Department of Computing and Information Systems
The University of Melbourne

Joint work with Timothy Baldwin, Spandana Gella, and Bo Han

Research Overview

- Computational lexicography
- Neologisms and new word-senses
- Word sense induction
- Web corpus construction and evaluation
- Multiword expressions
- Social media processing

NLP for Social Media

- Lots of research in NLP for Twitter
  - Event detection, sentiment analysis, text-based geolocation
  - POS tagging, NER, lexical normalisation

NLP for Social Media

- Lots of research in NLP for Twitter
  - Event detection, sentiment analysis, text-based geolocation
  - POS tagging, NER, lexical normalisation
- Lexical semantics?

Lexical Semantics on Twitter

- Little work to-date
- Short, noisy text: Challenging for traditional approaches
- Lexical semantics might benefit applications
- Learn how social media and conventional text differ

This Talk

1. Word usage patterns on Twitter
2. Word usage similarity for Twitter

Word Usage Patterns

Conventional text

- One sense per discourse
- First-sense heuristic

Word Usage Patterns

Conventional text

- One sense per discourse
- First-sense heuristic

Twitter

- One sense per tweeter?

Word Usage Patterns

Conventional text

- One sense per discourse
- First-sense heuristic

Twitter

- One sense per tweeter?
- First-sense heuristic?

One Sense per Tweeter?

User1

- wrap your hands around your elbows and maintain the **position** for a few seconds
- Look at this elegant yoga **position**: URL

User2

- Growing up I somehow knew I would be in this **position**.
- check out the programmer analyst (west corporation) **position** URL

First-sense Heuristic for Twitter?

- Lmao like I don't do that to every **paper** lol smh
- We loved making **paper** boats as kids. I feel sorry for kids now a days who don't know how to make one.
- Will Raptors easy playoff stretch run on **paper** result in third place in NBA East? URL #raptors

Motivation

One sense per tweeter

- Lack of context ($\leq$ 140 characters)
- Address this via user-level sense priors
- Understanding of user-level word usage

First-sense heuristic

- Effectiveness of first-sense WSD

Resources

- Sense inventory: Macmillan Dictionary
  - Coarse-grained senses
  - Regularly updated
- Target lemmas: 20 nouns
  - High-to-mid frequency
  - Medium polysemy: ≥ 3 senses (ave. 5.5)

Datasets

- 4 datasets:
  $\{\textsc{Twitter}, \textsc{ukWaC}\} \times \{\textsc{rand}, \textsc{user}\}$
- ukWaC: More-conventional (web) text
- RAND: Random sample of usages
- USER: 5 usages from each user/document
- 2000 items each: 100 usages of each noun

Annotation

- Amazon Mechanical Turk
- For each usage, pick the most appropriate sense(s), or "other"
- Quality control
  - Included some gold-standard Macmillan example sentences in each HIT
  - Filtered annotations based on accuracy over these items
- Fleiss' Kappa: 0.47–0.71
- Final annotation via unweighted voting

Analysis

Average proportion of users/documents using a noun in the same sense across all 5 usages

- TWITTER$_{\text{USER}}$: 65%
- UKWAC$_{\text{DOC}}$: 63%

One sense per tweeter heuristic is as strong as one sense per discourse

Analysis: Pairwise Agreement

	Partition	Agreement (%)
Gale et al. (1992)	document	94.4
TWITTER$_{\text{USER}}$	user	95.4
TWITTER$_{\text{USER}}$	—	62.9
TWITTER$_{\text{RAND}}$	—	55.1
UKWAC$_{\text{DOC}}$	document	94.2
UKWAC$_{\text{DOC}}$	—	65.9
UKWAC$_{\text{RAND}}$	—	60.2

Analysis: Pairwise Agreement

	Partition	Agreement (%)
Gale et al. (1992)	document	94.4
TWITTER$_{USER}$	user	95.4
TWITTER$_{USER}$	—	62.9
TWITTER$_{RAND}$	—	55.1
UKWAC$_{DOC}$	document	94.2
UKWAC$_{DOC}$	—	65.9
UKWAC$_{RAND}$	—	60.2

Analysis: Pairwise Agreement

	Partition	Agreement (%)
Gale et al. (1992)	document	94.4
TWITTER$_{USER}$	user	95.4
TWITTER$_{USER}$	—	62.9
TWITTER$_{RAND}$	—	55.1
UKWAC$_{DOC}$	document	94.2
UKWAC$_{DOC}$	—	65.9
UKWAC$_{RAND}$	—	60.2

Analysis: Pairwise Agreement

	Partition	Agreement (%)
Gale et al. (1992)	document	94.4
TWITTER$_{USER}$	user	95.4
TWITTER$_{USER}$	—	62.9
TWITTER$_{RAND}$	—	55.1
UKWAC$_{DOC}$	document	94.2
UKWAC$_{DOC}$	—	65.9
UKWAC$_{RAND}$	—	60.2

Analysis: First-sense Heuristic

Accuracy of an oracle first-sense WSD system

- TWITTER$_{\text{RAND}}$: 45.3%
- UKWAC$_{\text{RAND}}$: 55.4%

First-sense tagging is less accurate in Twitter data

Further Analysis

Comparing $\textsc{Twitter}_{\textsc{rand}}$ and $\textsc{ukWaC}_{\textsc{rand}}$

- Sense distributions are less skewed on Twitter
  - Sense entropy lower for $\textsc{ukWaC}_{\textsc{rand}}$ for 15 nouns
- 8/20 nouns have different first senses
- More "Other" senses in Twitter data
  - $\textsc{Twitter}_{\textsc{rand}}$: 12.3%
  - $\textsc{ukWaC}_{\textsc{rand}}$: 6.6%

Future Analyses

- One sense per conversation?
- Impact of time?
- Impact of geospatial and sociolinguistic factors?

Summary so far

- One sense per tweeter?
    - At least as strong as one sense per discourse
- First-sense heuristic?
    - First sense tagging is less accurate for Twitter
- Annotated dataset available soon

Further Analysis

Comparing TWITTER$_{\text{RAND}}$ and UKWAC$_{\text{RAND}}$

- Sense distributions are less skewed on Twitter
  - Sense entropy lower for UKWAC$_{\text{RAND}}$ for 15 nouns
- 8/20 nouns have different first senses
- **More "Other" senses in Twitter data**
  - TWITTER$_{\text{RAND}}$: 12.3%
  - UKWAC$_{\text{RAND}}$: 6.6%

This Talk

1. Word usage patterns on Twitter
2. Word usage similarity for Twitter

Word Sense Disambiguation (WSD)

Given a word in context, select the best-fitting sense from a sense inventory

ne1 headin to blue boyz footy match this weekend? #iamcarlton

Word Sense Disambiguation (WSD)

Given a word in context, select the best-fitting sense from a sense inventory

ne1 headin to blue boyz footy match this weekend? #iamcarlton

- a formal contest in which two or more persons or teams compete

Word Sense Disambiguation (WSD)

Given a word in context, select the best-fitting sense from a sense inventory

I think they are a perfect match! #cute #xoxo

Word Sense Disambiguation (WSD)

Given a word in context, select the best-fitting sense from a sense inventory

I think they are a perfect match! #cute #xoxo

- something that resembles or harmonizes with; "that tie makes a good match with your jacket"

Word Sense Disambiguation (WSD)

Given a word in context, select the best-fitting sense from a sense inventory

I think they are a perfect match! #cute #xoxo

- something that resembles or harmonizes with; "that tie makes a good match with your jacket"
- a pair of people who live together; "a married couple from Chicago"

Issues with WSD

- Choice of sense inventory
  - *match.n*: WordNet: 9 senses, Macmillan: 4 senses
- Cannot capture novel usage patterns
- Assumes a single sense per usage
- Existence of word senses has been questioned

Issues with WSD for Social Media

- Few sense tagged resources for social media
- Short, noisy, non-standard syntax
- More usages that don't match senses from conventional inventories

Issues with WSD for Social Media

- Few sense tagged resources for social media
- Short, noisy, non-standard syntax
- More usages that don't match senses from conventional inventories

Let's just do away with sense inventories

Usage Similarity (Usim)

- Manual task of rating the similarity of a pair of usages of a word (Erk and McCarthy 2009)
- Similarity on a graded scale (1 – 5)
- No more senses; independent of sense inventory
- Novel usages: Rate similarity to other usages

Usim Example

- *Setting goals for myself this year, figured if it's on <u>paper</u> I'll be more inspired to work harder.*
- *This is very unsmart of me to get tipsy and then have to go home and write a <u>paper</u>.*

Annotators' judgement: 3.2

Usim-tweet Dataset

- 10 nouns (from original Usim study)
- 55 pairs of tweets annotated per lemma
- Annotation via Amazon Mechanical Turk
- Rate similarity of usage pairs: 1–5, or unknown
- Quality control based on correlation with all other annotators

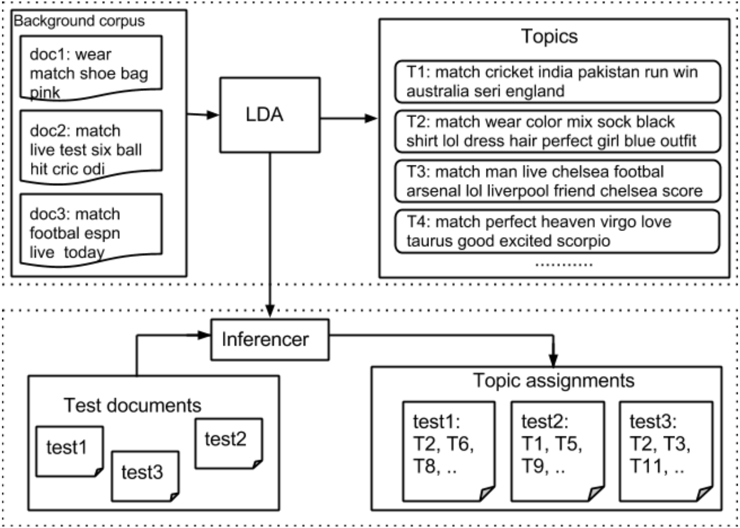
Modelling Usim for Twitter

- No large annotated training resources: unsupervised method
- No parser (yet): bag-of-words
- Methods:
  - Topic models (LDA)
  - Vector space model (VSM)
  - Weighted Textual Matrix Factorization (WTMF)
- 1 model per target word

Methods

- LDA (Our approach)
  - Represent documents as topic distribution vectors
  - Previously applied to model Usim, but not for social media (Lui et al., 2012)
- VSM (Baseline)
  - Second order co-occurrence
- WTMF (Benchmark)
  - Consider information from "missing" words related to the latent vector profile
  - State-of-the-art on a similar task

Topic Modeling — LDA



Background Corpora

- ORIGINAL: Tweets from Streaming API containing the target word as a noun
  - *gotta 3 page* **paper** *due tomorrow haven start #procrastination*
  - 17k – 299k tweets per target
- HTEXPANDED: ORIGINAL + document expansion with extra tweets containing medium-frequency hashtags
  - @USER assignment due in 1 hour 15 minutes #letsdothis #procrastination
  - 19k – 301k tweets per target
- EXPANDED: ORIGINAL + extra tweets containing the target word

Method Overview

- Build vectors using VSM, LDA, or WTMF representing each tweet
- Measure Cosine similarity of each usage pair
- Measure correlation between system similarity scores and gold-standard judgements

Results

Model	ORIGINAL	HTEXPANDED	EXPANDED
Baseline	0.09	0.08	0.09
WTMF (d)	0.03 (8)	0.10 (20)	0.09 (5)
LDA (T)	**0.20** (8)	**0.29** (5)	**0.18** (20)

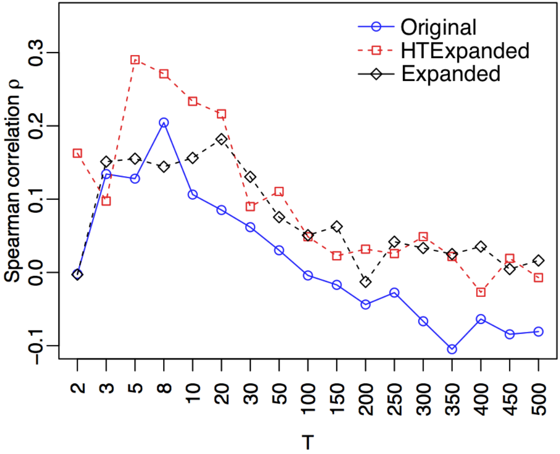
Spearman's rank correlation (ρ) for each method based on each background corpus

Results on HTEXPANDED

Lemma	Best T		$T = 5$
	ρ	T	ρ
bar	**0.35**	50	0.10
charge	**0.33**	20	-0.08
execution	**0.58**	5	**0.58**
field	**0.53**	10	0.32
figure	0.24	250	0.14
function	**0.40**	10	**0.27**
investigator	**0.50**	5	**0.50**
match	**0.45**	5	**0.45**
paper	0.32	30	0.22
post	0.20	30	-0.01
Overall	**0.39**	5	**0.29**

ρ values that are significant ($p > 0.05$) are shown in bold

Results Varying T



ρ versus number of topics (T)

Future Work

- Alternative document expansion: E.g., author-based
- Richer context representation: POS, positional word features, etc.
- Non-parametric topic modelling (e.g., HDP)

Usim Summary

- Proposed a computational model of usage similarity for Twitter
- LDA approach out-performed a baseline and benchmark
- Hashtag-based document expansion improved LDA

Summary

- Little work on lexical semantics for social media
- Social media contains many non-standard usages
- Can exploit user-level priors and social media structure
- Potential for new models of lexical semantics

Thanks