Social Media & Text Analysis
lecture 7 - learn large-scale paraphrase from Twitter

Instructor: Wei Xu
Website: socialmedia-class.org
… the forced resignation of the CEO of Boeing, Harry Stonecipher, for ...

wealthy

word

rich

the king’s speech

phrase

His Majesty’s address

… after Boeing Co. Chief Executive Harry Stonecipher was ousted from ...
... the forced resignation of the CEO of Boeing, Harry Stonecipher, for ...

... after Boeing Co. Chief Executive Harry Stonecipher was ousted from ...

Who is the CEO stepping down from Boeing?

… the forced resignation of the CEO of Boeing, Harry Stonecipher, for …

… after Boeing Co. Chief Executive Harry Stonecipher was ousted from …
Watson leverages multiple algorithms to perform deeper analysis

[Question]
In May 1898 Portugal celebrated the 400th anniversary of this explorer’s arrival in India.

[Supporting Evidence]
On the 27th of May 1498, Vasco da Gama landed in Kappad Beach

Legend
- Temporal Reasoning
- Statistical Paraphrasing
- GeoSpatial Reasoning
- Reference Text
- Answer

Stronger evidence can be much harder to find and score...
- Search far and wide
- Explore many hypotheses
- Find judge evidence
- Many inference algorithms
They are culturally akin to the coastal peoples of Papua New Guinea.

Their culture is like that of the coastal peoples of Papua New Guinea.
Palpatine:
*If you will not be turned, you will be destroyed!*

If you will not be turn’d, you will be undone!

Luke:
*Father, please! Help me!*

Father, I pray you! Help me!
Paraphrase Data

80s WordNet
- Barzilay McKeown

'01 Novels
- Dolan Quirk Brockett

'04 News
- Callison-Burch Ganitkevitch Van Durme Bernard

'05 '13 Bi-Text
- Chen Dolan

'11 Video Shakespeare
- Xu Ritter Dolan Grishman Cherry

'12

(This Work)
Twitter
- Xu Ritter Callison-Burch Dolan Ji

Wei Xu · socialmedia-class.org
Paraphrase Research

- WordNet
- Novel
- News
- Bi-Text
- Video
- Shakespeare

information extraction
question answering
semantic similarity
semantic parsing
text-to-text generation
...

but, primarily for formal language usage and well-edited text
Previous Work

multiple English translations of novels

(Barzilay and McKeown, 2001)
Previous Work

only a few hundreds news agencies report big events using formal language

(Dolan, Quirk and Brockett, 2004; Dolan and Brockett, 2005; Brockett and Dolan, 2005)
Previous Work

ask dozens of annotators to write one sentence for a short video (<= 10 seconds)

(Chen and Dolan, 2011)
Previous Work

... 5 farmers were thrown into jail in Ireland ...

... fünf Landwirte festgenommen, weil ...

... oder wurden festgenommen, gefoltert ...

... or have been imprisoned, tortured ...

pivoting through bilingual text from European Parliament proceedings, multilingual websites etc.

(Bannard and Callison-Burch, 2005; Ganitkevitch, Van Durme and Callison-Burch, 2013)
Twitter as a new resource

Rep. Stacey Newman @staceynewman · 5h
So sad to hear today of former WH Press Sec James Brady’s passing. @bradybuzz & family will carry on his legacy of #gunsense.

Jim Sciutto @jimsciutto · 4h
Breaking: Fmr. WH Press Sec. James Brady has died at 73, crusader for gun control after wounded in ’81 Reagan assassination attempt

NBC News @NBCNews · 2h
James Brady, President Reagan’s press secretary shot in 1981 assassination attempt, dead at 73 nbcnews.to/WX1Btq pic.twitter.com/1ZtuEakRd9

Wei Xu, Alan Ritter, Ralph Grishman.
“A Preliminary Study of Tweet Summarization using Information Extraction” in LASM (2014)
Twitter as a powerful resource

thousands of users
talk about both big and micro events
using formal, informal, erroneous language

Very diverse!
Enables new applications

Noisy Text Normalization

\[
\text{oscar nom’d doc} \quad \text{Oscar-nominated documentary} \\
\text{don’t want for} \quad \text{don’t wait for}
\]
Enables new applications

Human-computer Interaction

who wants to get a beer?

want to get a beer?
who else wants to get a beer?
who wants to go get a beer?
who wants to buy a beer?
who else wants to get a beer?
trying to get a beer?

… (21 different ways)

Wei Xu, Alan Ritter, Ralph Grishman.

“Gathering and Generating Paraphrases from Twitter with Application to Normalization” In BUCC (2013)
Enables new applications

Language Education

Aaaaaaaaaand stephen curry is on fire

What a incredible performance from Stephen Curry

Listen & Speak
Like a Native Speaker

(ONGOING WORK WITH QUANZE CHEN AND CHRIS CALLISON-BURCH)
Enables new applications

Sentiment Analysis

This nets vs bulls game is great
This Nets vs Bulls game is nuts
Wowsers to this nets bulls game
this Nets vs Bulls game is too live
This Nets and Bulls game is a good game
This netsbulls game is too good
This NetsBulls series is intense
Learn Paraphrases
Learn Paraphrases

identify parallel sentences automatically from Twitter’s big data stream

- Mancini has been sacked by Manchester City
- Mancini gets the boot from Man City

- WORLD OF JENKS IS ON AT 11
- World of Jenks is my favorite show on tv
Early Attempts on Twitter

- 1242 tweet pairs, tracking celebrity & hashtags (Zanzotto, Pennacchiotti, Tsioutsioulakis, 2011)

- named entity + date (Xu, Ritter, Grishman, 2013)

- bilingual posts (Ling, Dyer, Black, Trancoso, 2013)
Named Entity + Time

Tyler Anderson
@tylerjanderson

From **January 16**, **Instagram** can sell your photos without permission
google.com/articles/geek-…

Jeff Clutter
@pibbbs

**Instagram** can sell your photos without consent starting **January 16th**.
Self-translation

La neige est annoncée pour demain
Snow is announced for tmrw!
bit.ly/1t85YQ5 #neige #snow
#winteriscoming #valloire

The snow is announced for tomorrow

(Ling, Dyer, Black and Trancoso, 2013)
Design a Model

Train it on data
A Challenge

Mancini has been sacked by Manchester City

Mancini gets the boot from Man City

very short
lexically divergent

(less word overlap, even in high-dimensional space)
## Previous Methods

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Reference</th>
<th>Description</th>
<th>Supervision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector Based</td>
<td>Mihalcea et al. (2006)</td>
<td>cosine similarity with tf-idf weighting</td>
<td>unsupervised</td>
</tr>
<tr>
<td>ESA</td>
<td>Hassan (2011)</td>
<td>explicit semantic space</td>
<td>unsupervised</td>
</tr>
<tr>
<td>KM</td>
<td>Kozareva and Montoyo (2006)</td>
<td>combination of lexical and semantic features</td>
<td>supervised</td>
</tr>
<tr>
<td>LSA</td>
<td>Hassan (2011)</td>
<td>latent semantic space</td>
<td>unsupervised</td>
</tr>
<tr>
<td>RMLMG</td>
<td>Rus et al. (2008)</td>
<td>graph subsumption</td>
<td>unsupervised</td>
</tr>
<tr>
<td>MCS</td>
<td>Mihalcea et al. (2006)</td>
<td>combination of several word similarity measures</td>
<td>unsupervised</td>
</tr>
<tr>
<td>WTMF</td>
<td>Guo and Diab (2012)</td>
<td>latent semantics model of missing words</td>
<td>unsupervised</td>
</tr>
<tr>
<td>STS</td>
<td>Islam and Inkpen (2007)</td>
<td>combination of semantic and string similarity</td>
<td>unsupervised</td>
</tr>
<tr>
<td>SSA</td>
<td>Hassan (2011)</td>
<td>salient semantic space</td>
<td>unsupervised</td>
</tr>
<tr>
<td>QKC</td>
<td>Qiu et al. (2006)</td>
<td>sentence dissimilarity classification</td>
<td>supervised</td>
</tr>
<tr>
<td>ParaDetect</td>
<td>Zia and Wasif (2012)</td>
<td>PI using semantic heuristic features</td>
<td>supervised</td>
</tr>
<tr>
<td>SDS</td>
<td>Blacoe and Lapata (2012)</td>
<td>simple distributional semantic space</td>
<td>supervised</td>
</tr>
<tr>
<td>matrixJcn</td>
<td>Fernando and Stevenson (2008)</td>
<td>JCN WordNet similarity with matrix</td>
<td>unsupervised</td>
</tr>
<tr>
<td>FHS</td>
<td>Finch et al. (2005)</td>
<td>combination of MT evaluation measures as features</td>
<td>supervised</td>
</tr>
<tr>
<td>PE</td>
<td>Das and Smith (2009)</td>
<td>product of experts</td>
<td>supervised</td>
</tr>
<tr>
<td>WDDP</td>
<td>Wan et al. (2006)</td>
<td>dependency-based features</td>
<td>supervised</td>
</tr>
<tr>
<td>SHPNM</td>
<td>Socher et al. (2011)</td>
<td>recursive autoencoder with dynamic pooling</td>
<td>supervised</td>
</tr>
<tr>
<td>MTMETERSICS</td>
<td>Madnani et al. (2012)</td>
<td>combination of eight machine translation metrics</td>
<td>supervised</td>
</tr>
<tr>
<td>LEXLATENT</td>
<td>Ji and Eienstein (2013)</td>
<td>combination of latent space and lexical features</td>
<td>supervised</td>
</tr>
</tbody>
</table>

mostly based on sentence similarity of surface words or latent semantics
Design a Model

At-least-one-anchor Assumption

two sentences about the same topic are paraphrases if and only if they contain at least one word pair that is a paraphrase anchor

That boy Brook Lopez with a deep 3
brook lopez hit a 3

Yes!

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji. “Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Another Challenge

not every word pair of similar meaning indicates sentence-level paraphrase

\[
\text{Iron Man 3} \text{ was brilliant fun} \\
\text{Iron Man 3 tonight see what this is like}
\]

No!

Solution:
a discriminative model using features at word-level

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji. “Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Features

- **String Features:**
  - words, stemmed forms, normalized forms
  - same, similar or dissimilar

- **POS Features:**
  - fine grained tags:
    - “a”, “be”, “do”, “have”,
    - “get”, “go”, “follow”, “please”

- **Topical Features:**
  - word significantly associated with each topic
  - e.g. “3” for basketball; “RIP” for death events
Multi-instance Learning Paraphrase Model

Manti bout to be the **next** Junior Seau

Teo is the little **new** Junior Seau

---

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji. “Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Instead of labels on each individual instance, the learner only observes labels on bags of instances.

Negative Bags

A bag is labeled negative, if all the examples in it are negative

Positive Bags

A bag is labeled positive, if there is at least one positive example

(Dietterich et al., 1997)
[Mini Tutorial] Multi-instance Learning

Latent Variable Model

A bag is labeled positive, if there is at least one positive example.

constraints

Positive Bag

Z₁

Z₂

Z₃

y

1

bag label (observed)

instance label (latent)

features
[Mini Tutorial] Multi-instance Learning

Latent Variable Model

- **Bag Label (observed)**
- **Instance Label (latent)**
- **Features**
- **Constraints**

A bag is labeled negative, if all the examples in it are negative.
[Mini Tutorial]  
Multi-instance Learning

Distantly Supervised Information Extraction

1. incomplete knowledge base problem  
2. distant supervision + human-labeled data  
3. IE + IR
[Recap] Multi-instance Learning Paraphrase Model

Manti bout to be the next Junior Seau

Teo is the little new Junior Seau

\[
\begin{align*}
Y_{\text{paraphrase}} & \quad \gamma_{\text{paraphrase}} \\
Y_{\text{non-paraphrase}} & \quad \gamma_{\text{non-paraphrase}}
\end{align*}
\]

sentence pair

word pair

features

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji. “Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Joint Word-Sentence Model

Model the assumption:

sentence-level paraphrase is anchored by at least one word pair.

Deterministic OR

\[
\sigma(z_i, y_i) = \begin{cases} 
1 & \text{if } y_i = \text{true} \land \exists j : z_j = 1 \\
1 & \text{if } y_i = \text{false} \land \forall j : z_j = 0 \\
0 & \text{otherwise}
\end{cases}
\]
Joint Word-Sentence Model

\[
P(z_i, y_i|w_i; \theta) = \prod_{j=1}^{m} \exp(\theta \cdot f(z_j, w_j)) \times \sigma(z_i, y_i)
\]

- \(i\)th sentence pair’s label (observed or to be predicated)
- \(j\)th word pair
- Parameters
- Features
- Deterministic OR

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji. “Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Learning Algorithm

Objective:
learn the parameters that maximize likelihood over the training corpus

\[ \theta^* = \arg \max_{\theta} P(y|w; \theta) = \arg \max_{\theta} \prod_{i} \sum_{z_i} P(z_i, y_i | w_i; \theta) \]

\( i \)th training sentence pair
all possible values
of the latent variables

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji.
“Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Learning Algorithm

**Perceptron-style Update:**

Viterbi approximation + online learning

\[ O(\# \text{ word pairs}) \]

\[
\frac{\partial \log P(y|w; \theta)}{\partial \theta} \approx \sum_i f(z_i^{*}, w_i) - \sum_i f(z'_i, w_i)
\]

- **reward correct** (conditioned on labels)
  \[ z^* = \arg \max_z P(z|w, y; \theta) \]

- **penalize wrong** (ignoring labels)
  \[ y', z' = \arg \max_{y, z} P(z, y|w; \theta) \]

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji. “Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Training Data
Twitter Trends

Pittsburgh Trends

#Odyssey
Promoted by Odyssey
#DukevsUNC
Tyus Jones
#EmpireFOX
Pittsburgh
Tar Heels
Steelers
Vanilla Ice
Xbox
Walmart
Annotation

Crowdsourcing

(Courtesy: The Sheep Market by Aaron Koblin)
Annotation

Crowdsourcing

Here Is The Question To You:
Original Sentence: *Borussia Dortmund advanced to the final*
Select ALL sentences that have similar meaning from below:

- Borussia Dortmund has clinched their Champions League final spot
- Real Madrid efforts are not enough as Cinderella Borussia Dortmund advances to the Champions League Final
- But it’s Borussia Dortmund whose heading to Wembley Park
- Congratulations Borussia Dortmund’s going to Wembley
A Problem

only 8% sentence pairs about the same topic have similar meaning

hurts both quantity and quality

non-experts lower their bars

Sentence Selection

SumBasic Algorithm

\[ \text{Salience}(s) = \sum_{w_i \in s} \frac{P(w_i)}{|w_i|} \]

8% → 16%

Topic Selection

Multi-Armed Bandits

\[
\max \sum_{\{i \mid r_i(t_0) > 0\}} \hat{\mu}_i(t_0) r_i(t_1)
\]

s.t. \( \sum_i c_i r_i(t_1) \leq (1 - \epsilon) B, \forall i : 0 \leq r_i(t_1) \leq l - r_i(t_0). \)

Twitter Paraphrase Dataset

18,762 sentence pairs labeled
cost only $200

1/3 paraphrase, 2/3 non-paraphrase (very balanced)

including a very broad range of paraphrases:
synonyms, misspellings, slang, acronyms and colloquialisms

Performance
Performance

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji.
“Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Performance

Precision  Recall


state-of-the-art of paraphrase identification

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji. “Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Performance

We see a graph comparing the performance of different models, with our model marked as "Our Model".

The bar chart shows the precision and recall for different models:

- **Our Model**
  - Precision: 66.4
  - Recall: 62.8

- **(Ji&Eisenstein,2013)**
  - Precision: 72.2
  - Recall: 72.6

- **Human Upperbound**
  - Precision: 75.2
  - Recall: 90.8

References:

- (Das&Smith, 2009)
- (Guo&Diab, 2012)
- (Ji&Eisenstein, 2013)

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji. “Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Innovations

That boy **Brook Lopez** with a deep 3
**brook lopez** hit a 3

Multi-instance Learning Paraphrase Model (MultiP)

- Twitter’s big data stream
- potential beyond Twitter and English
- joint sentence-word alignment
- extensible latent variable model

(a lot of space for future work)

Wei Xu, Alan Ritter, Chris Callison-Burch, Bill Dolan, Yangfeng Ji.
“Extracting Lexically Divergent Paraphrases from Twitter” In TACL (2014)
Impact

Shared Task: Paraphrase and Semantic Similarity in Twitter
19 research groups participated (100+ requested the data)
Thank you

Instructor: Wei Xu

www.cis.upenn.edu/~xwe/

Course Website: socialmedia-class.org

wawwww thankkkkkkkkkkkkkk you alotttttttttttttt!

thanks a lot

am grateful