Authorship Attribution of Micro-Messages

Roy Schwartz+, Oren Tsur+, Ari Rappoport+ and Moshe Koppel*

+The Hebrew University, *Bar Ilan University
In proceedings of EMNLP 2013
Overview

• Authorship attribution of tweets

• Users tend to adopt a unique style when writing short texts (k-signatures)

• A new feature for authorship attribution
  – Flexible patterns
  – Significant improvement over our baselines

• 6.1% improvement over state-of-the-art
Authorship Attribution

• “To be, or not to be: that is the question”
  • “Romeo, Romeo! wherefore art thou Romeo”
  • …

• “Taking a new step, uttering a new word, is what people fear most”
  • “If they drive God from the earth, we shall shelter Him underground.”
  • …

• “Before all masters, necessity is the one most listened to, and who teaches the best.”
  • “The Earth does not want new continents, but new men.”
  • …
Authorship Attribution

“Love all, trust a few, do wrong to none.”
History of Authorship Attribution

- Mendenhall, 1887
History of Authorship Attribution

• Mendenhall, 1887

• Traditionally: long texts
History of Authorship Attribution

- Mendenhall, 1887
- Traditionally: long texts
- Recently: short texts
History of Authorship Attribution

- Mendenhall, 1887

- Traditionally: long texts

- Recently: short texts

- Very recently: very short texts
History of Authorship Attribution

- Mendenhall, 1887
- Traditionally: long texts
- Recently: short texts
- Very recently: very short texts
Tweets as Candidates for Short Text

- Tweets are limited to 140 characters
Tweets as Candidates for Short Text

- Tweets are limited to 140 characters
- Tweets are (relatively) self contained
Tweets as Candidates for Short Text

• Tweets are limited to 140 characters

• Tweets are (relatively) self contained

• Compared to standard web data sentences
  – Tweets are shorter (14.2 words vs. 20.9)
  – Tweets have smaller sentence length variance (6.4 vs. 21.4)
Experimental Setup

• Methodology
  – SVM with linear kernel; character n-grams, word n-gram, flexible patterns features

• Experiments
  – Varying training set sizes, varying number of authors, recall-precision tradeoff

• Results
  – 6.1% improvement over current state-of-the-art
Experimental Setup

• Methodology
  – SVM with linear kernel, word n-gram, flexible patterns features

• Experiments
  – Varying training set sizes, number of authors, recall-precision tradeoff

Some Interesting Findings First

• Results
  – 6.1% improvement over current state-of-the-art
Interesting Finding

• Users tend to adopt a unique style when writing short texts
Interesting Finding

• Users tend to adopt a unique style when writing short texts

• K-signatures
  – A feature that is unique to a specific author A
  – Appears in at least k% of A’s training set, while not appearing in the training set of any other user
## K-signatures Examples

<table>
<thead>
<tr>
<th>Signature Type</th>
<th>10%-signature</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character n-grams</td>
<td>‘^_^’</td>
<td>REF oh ok ^_^ Glad you found it!</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hope everyone is having a good afternoon ^_^</td>
</tr>
<tr>
<td></td>
<td></td>
<td>REF Smirnoff lol keeping the goose in the freezer ^_^</td>
</tr>
<tr>
<td></td>
<td>‘yew’</td>
<td>gurl yew serving me tea nooch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>REF about wen yew and ronnie see each other</td>
</tr>
</tbody>
</table>
|                  |               | REF lol so yew goin to check out tini’s tonight huh???
K-signatures per User

100 authors, 180 training tweets per author
More about K-signatures

• Implicit?
More about K-signatures

• Implicit?

• Style or content?
More about K-signatures

• Implicit?

• Style or content?

• Useful classification features
## Structured Messages / Bots?

<table>
<thead>
<tr>
<th>User</th>
<th>20%-signature</th>
<th>Examples</th>
</tr>
</thead>
</table>
| 1    | I’m listening to: | **Intro:** http://www.last.fm/music/Sigur+R%C3%B3s http://bit.ly/3XJHyb  
I’m listening to: Midnight Oil? Under the Overpass: http://www.last.fm/music/Midnight+Oil http://bit.ly/7I4cg |
| 2    | news now (str) | #Hotel **News Now**(STR) 5 things to know: 27 May 2009: From the desks of the HotelNewsNow.com editor... http://bit.ly/aZTZOq #Tourism #Lodging  
#Hotel **News Now**(STR) Five sales renegotiating tactics: As bookings representatives press to reneg... http://bit.ly/bHPn2L  
#Hotel **News Now**(STR) Risk of hotel recession retreats: The Hotel Industry’s Pulse Index increases... http://bit.ly/a8EKrm #Tourism #Lodging |
| 3    | (NUM bids) end date: | NEW PINK NINTENDO DS LITE CONSOLE WITH 21 GIFTS + CASE: &amp;#163;66.50 **(13 Bids) End Date:** Tuesday Dec-08-2009 17:.. http://bit.ly/7uPt6V  
Methodology

• Features
  – Character n-grams, word n-grams

• Model
  – Multiclass SVM with a linear kernel
Experiments

• Varying training set sizes
  – 10 groups of 50 authors each, 50-1000 training tweets per author
Experiments

• Varying training set sizes
  – 10 groups of 50 authors each, 50-1000 training tweets per author

• Varying numbers of authors
  – 50-1000 authors, 200 training tweets per author
Experiments

• Varying training set sizes
  – 10 groups of 50 authors each, 50-1000 training tweets per author

• Varying numbers of authors
  – 50-1000 authors, 200 training tweets per author

• Recall-precision tradeoff
  – “don’t know” option
Varying Training Set Sizes
50 Authors (2% Random Baseline)

Authorship Attribution of Micro-Messages @ Schwartz et al., EMNLP 2013
Varying Training Set Sizes
50 Authors (2% Random Baseline)

~50% accuracy (50 training tweets per author)
Varying Training Set Sizes

50 Authors (2% Random Baseline)

~70% accuracy (1000 training tweets per author)

~50% accuracy (50 training tweets per author)

Authorship Attribution of Micro-Messages @ Schwartz et al., EMNLP 2013
Varying Numbers of Authors
200 Training Tweets per Author

Authorship Attribution of Micro-Messages @
Schwartz et al., EMNLP 2013
Varying Numbers of Authors

200 Training Tweets per Author

~30% accuracy (1000 authors, 0.1% baseline)
Recall-Precision Tradeoff

Authorship Attribution of Micro-Messages @ Schwartz et al., EMNLP 2013
Recall-Precision Tradeoff

~90% precision, >~60% recall

Authorship Attribution of Micro-Messages @ Schwartz et al., EMNLP 2013
Recall-Precision Tradeoff

~90% precision, >~60% recall

~70% precision, ~30% recall
Flexible Patterns

• A generalization of word n-grams
  – Capture potentially unseen word n-grams

• Computed automatically from plain text
  – Language and domain independent
Flexible Patterns Examples

• the X of the
  – Go to the house of the rising sun
  – Can you hear the sound of the wind?

• as X as Y.
  – John is as clever as Mary.
  – Dogs run as fast as 30mph.
Flexible Patterns

• Shown to be useful in various NLP applications
  – Extraction of semantic relationships (Davidov, Rappoport and Koppel, ACL 2007)
  – Enhancing lexical concepts (Davidov and Rappoport, EMNLP 2009)
  – Detection of sarcasm (Tsur, Davidov and Rappoport, ICWSM 2010)
  – Sentiment analysis (Davidov, Tsur and Rappoport, Coling 2010)
  – …

• First work to apply flexible patterns on authorship attribution
Flexible Patterns Features

- Examples of tweets written by the same author
  - "the way I treated her"
  - "half of the things I’ve seen"
  - "the friends I have had for years"
  - "in the neighborhood I grew up in"
Flexible Patterns Features

- Examples of tweets written by the same author
  - “the way I treated her”
  - “half of the things I’ve seen”
  - “the friends I have had for years”
  - “in the neighborhood I grew up in”

- No word n-gram feature is able to capture this author’s style
Flexible Patterns Features

• Examples of tweets written by the same author
  – “the way I treated her”
  – “half of the things I’ve seen”
  – “the friends I have had for years”
  – “in the neighborhood I grew up in”

• No word n-gram feature is able to capture this author’s style

• Author’s character n-grams (“the”, “ I ”) are unindicative
Flexible Patterns Features

• Examples of tweets written by the same author
  – “the way I treated her”
  – “half of the things I’ve seen”
  – “the friends I have had for years”
  – “in the neighborhood I grew up in”

• No word n-gram feature is able to capture this author’s style

• Author’s character n-grams (“the”, “I”) are unindicative
Some more Results

• Flexible patterns obtains a statistically significant improvement over our baselines
  – 2.9% improvement over character n-grams
  – 1.5% improvement over character n-grams + word n-grams
Some more Results

• Flexible patterns obtains a statistically significant improvement over our baselines
  – 2.9% improvement over character n-grams
  – 1.5% improvement over character n-grams + word n-grams

• Our system obtains a 6.1% improvement over current state-of-the-art (Layton et al., 2010)
  – Using the same dataset

• We thank Robert Layton for providing us with his dataset
Summary

• Accurate authorship attribution of very short texts
  – 6.1% improvement over current state-of-the-art

• Many authors use k-signatures in their writing of short texts
  – A partial explanation for our high-quality results

• Flexible patterns are useful authorship attribution features
  – Statistically significant improvement
Authorship Attribution

“Love all, trust a few, do wrong to none.”
“Love all, trust a few, do wrong to none.”