

Identifying Authorships of very Short Texts using Flexible Patterns

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

⁺The Hebrew University, ^{*}Bar Ilan University
ICRI-CI Retreat, May 2014



Agenda

- We have previously shown that ***flexible patterns*** are useful for extracting semantic information
- We apply this technology to a new task – identifying the author of a very short text

Flexible Patterns

as X as Y

X and Y

X is Y than Z

Flexible patterns encode semantic information

Flexible Patterns Examples

- “X and Y” indicates **semantic similarity** between X and Y:
 - apples and oranges
- “as X as Y” indicates that Y is X:
 - John is as clever as Mary
- “X can’t Y these Z. great!” indicates a **sarcastic** review
 - The Sony eBook can’t read these formats. Great!

Flexible Patterns

- Computed automatically from **plain text**
 - Require no **manual annotation** or **external knowledge**
- Useful in various NLP applications
 - Extracting semantic relations (Davidov, Rappoport and Koppel, ACL'07)
 - Detection of sarcasm (Tsur, Davidov and Rappoport, ICWSM'10)
 - Sentiment analysis (Davidov, Tsur and Rappoport, Coling'10)
 - ...

Authorship Attribution

Authorship Attribution of Micro-Messages,
Schwartz et al., EMNLP'13



• “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.”

• ...

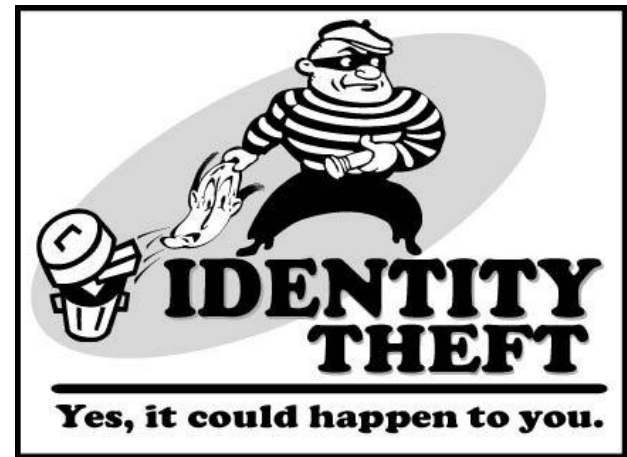
“Love all, trust a few, do wrong to none.”

History of Authorship Attribution

- Mendenhall, 1887
- Traditionally: long texts
- Recently: short texts
- Very recently: **very** short texts (≤ 140 chars)



Authorship Attribution Applications



Dataset

- A Twitter corpus that includes approximately 5×10^8 tweets from 2009-2010
- Randomly select 1000 users with 1000 tweets each

Text Classification

Feature Extraction

“Change will not come if we wait for some other person or some other time”

Feature Extraction

Character n-grams

“Change will not come if we wait for some other person or some other time”

“ome” => 3

“or s” => 2

“ange” => 1

...

Feature Extraction

Word N-grams

“Change will not come if we wait for some other
person or some other time”

“some other” => 2

“Change will not come” => 1

...

Feature Extraction

Flexible Patterns

“Change will not come if we wait for some other person or some other time”

“will not X if we Y” $\Rightarrow 1$

“some other X or some other Y” $\Rightarrow 1$

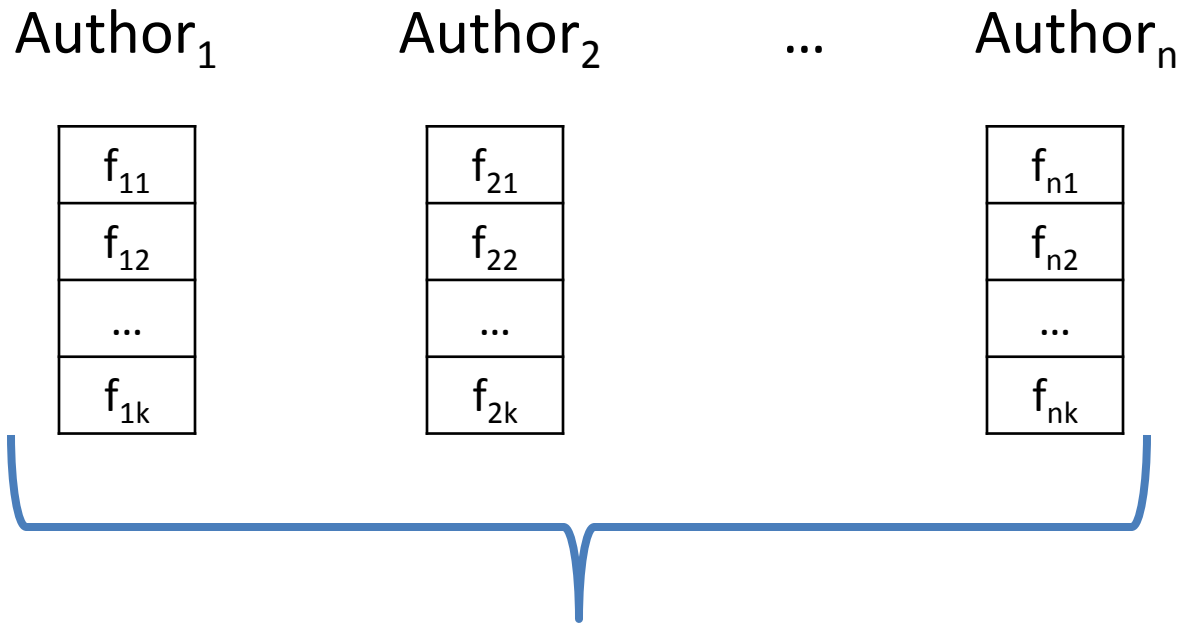
...

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

“*the X I*”

Text Classification Model



Multiclass SVM with a Linear Kernel

Results

- 71.2% classification accuracy
 - 50 authors, 1000 training tweets per author
 - 6.1% improvement over the current state of the art
- Flexible patterns obtain a statistically significant improvement over other features

K-Signatures

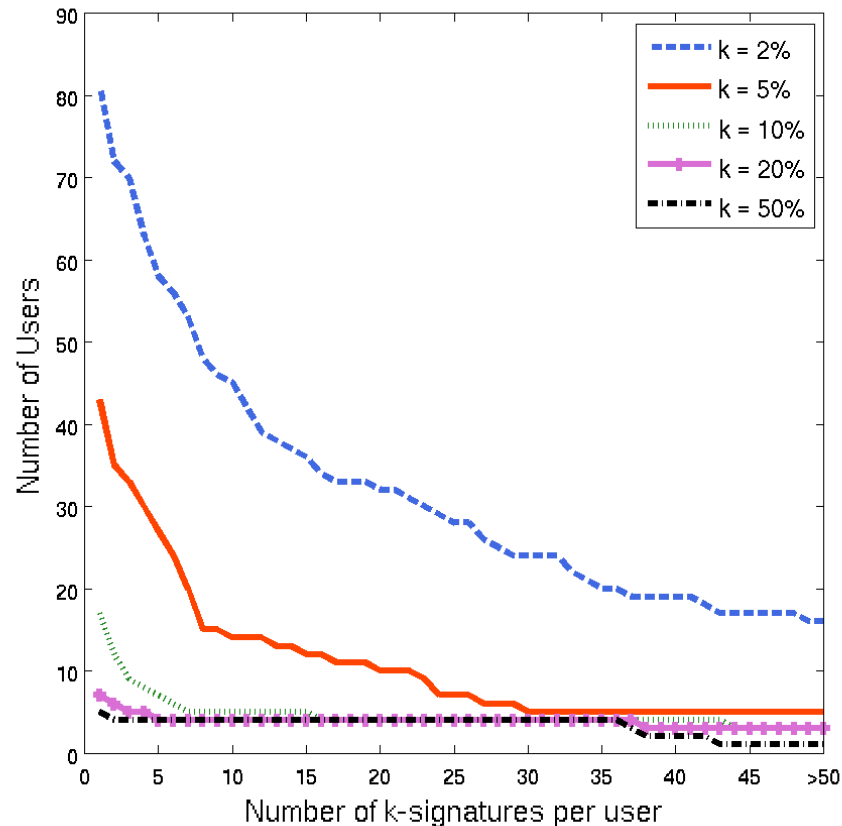
- Users tend to adopt a **unique style** when writing short texts
- K-signatures
 - A feature that is unique to a specific author A
 - Appears in at least $k\%$ of A 's training set, while not appearing in the more than 0.5% of the training set of any other user

K-signatures Examples

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

K-signatures per User

100 authors, 180 training tweets per author

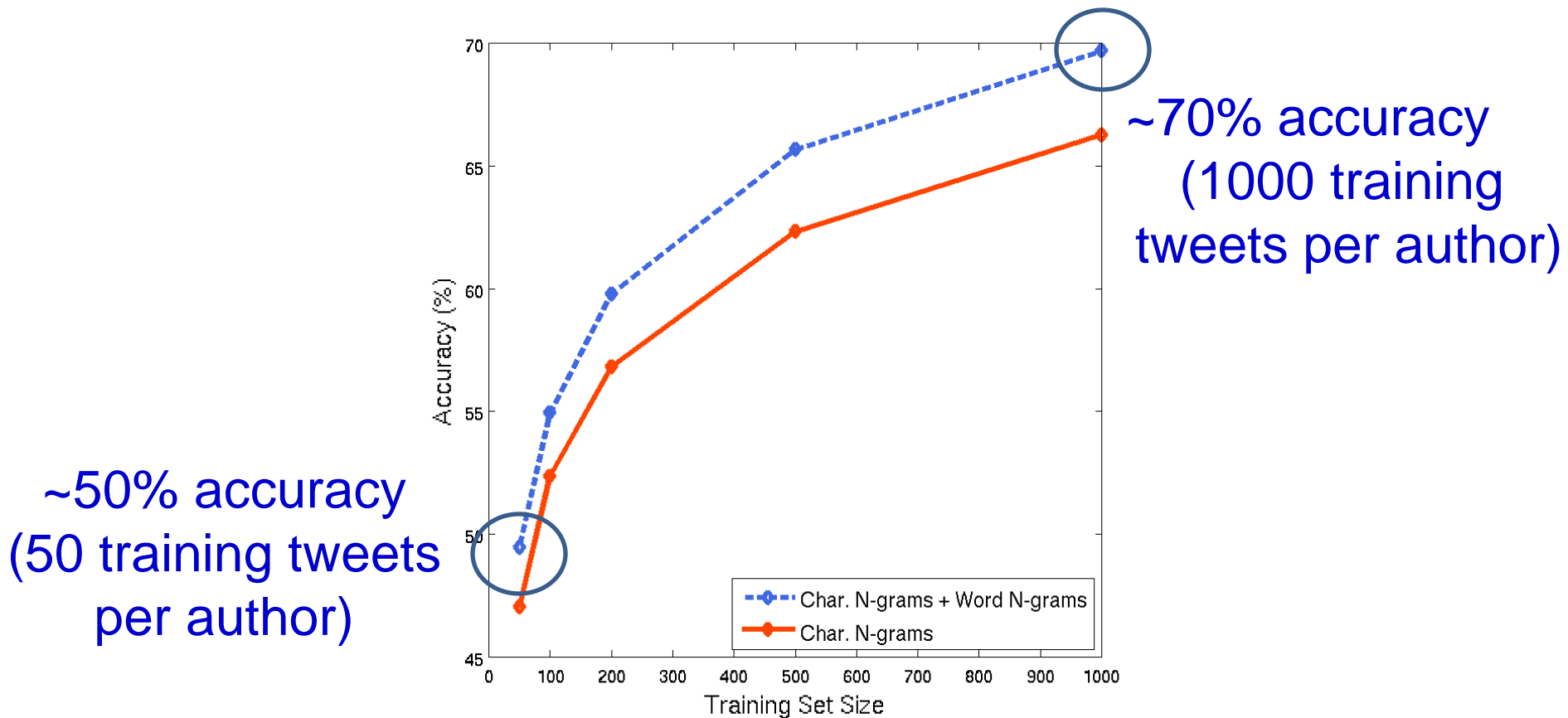


Structured Messages / Bots?

User	20%-signature	Examples
1	I'm listening to :	I'm listening to: Sigur Rós ? Intro: http://...
		I'm listening to: Tina Arena ? In Command: http://...
		I'm listening to: Midnight Oil ? Under the Overpass: http://
2	news now (str)	#Hotel News Now(STR) 5 things to know: 27May ...
		#Hotel News Now(STR) Five sales renegotiating ...
		#Hotel News Now(STR) Risk of hotel recession ...
3	(NUM bids) end date :	NEW PINK NINTENDO DS LITE (13 Bids) End Date: ...
		Microsoft Xbox 360 Game System (25 Bids) End Date: ...
		Microsoft Sony Playstation 3 ... (25 Bids) End Date: ...

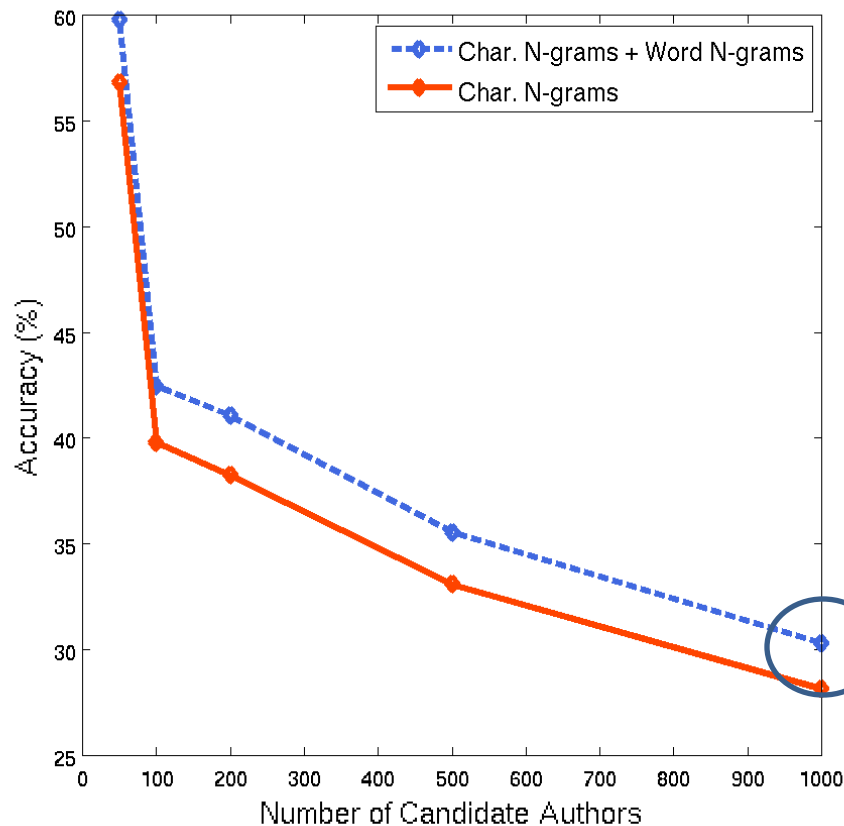
Varying Training Set Sizes

50 Authors (2% Random Baseline)



Varying Numbers of Authors

200 Training Tweets per Author



~30% accuracy
(1000 authors,
0.1% baseline)

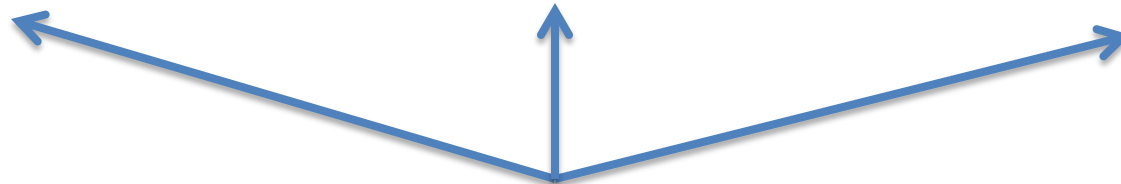
Summary

- Accurate authorship attribution of very short texts
- Many authors use k-signatures in their writing of short texts
- Flexible patterns are useful authorship attribution features

What's Next?

- Use flexible patterns to classify words into semantic categories with minimal supervision
- Automatically obtain a complete semantic description of a concept
 - A **dog** is an *animal*, which *barks*, has a *tail*, is *faithful*, is related to *cats*, etc.

Authorship Attribution



?

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



roys02@cs.huji.ac.il

<http://www.cs.huji.ac.il/~roys02/>

