Semantic Representation using Flexible Patterns

Roy Schwartz

The Hebrew University of Jerusalem, October 2013

Overview

- Lexico-syntactic Patterns
  - Patterns are useful for extracting semantic data
- Flexible Patterns
  - Lexico-syntactic patterns extracted in a **fully unsupervised** manner
- Also, (more) useful for extracting semantic data
  - Some interesting results from our lab
- Latest results
  - Authorship attribution of tweets using flexible patterns (EMNLP 2013)

Lexico-syntactic Patterns

Hearst, 1992

- Patterns of the form "***X*** *is a country*", "***X*** *such as* ***Y***", etc.

Lexico-syntactic Patterns

Hearst, 1992

- Patterns of the form "***X** is a country*", "***X** such as **Y***", etc.

- Patterns potentially capture the **context** in which a word participates

Lexico-syntactic Patterns

Hearst, 1992

- Patterns of the form "***X** is a country*", "***X** such as **Y***", etc.

- Patterns potentially capture the **context** in which a word participates

- For example:
  - A *dog* participates in patterns (contexts) such as:
  - "X barks", "X has a tail", "X and cats", …

Lexico-syntactic Patterns

- Hand crafted patterns have been used in many semantic tasks
- Acquiring the semantics of **single words**
  - Building semantic lexicons (Riloff and Shepherd, 1997; Roark and Charniak, 1998)
  - Semantic class learning (Kozareva et al., 2008)
- Acquiring the semantics of **relationships** between words
  - Discovering hyponymy (Hearst, 1992)
  - Discovering meronymy (Berland and Charniak, 1999)
  - Discovering Verb relations (Chklovski and Pantel, 2004)

Examples of Patterns

- Extracting country names
  - "X is a country"

Examples of Patterns

- Extracting country names
  - "X is a country"
  - ***Canada is a country*** *in north America*
  - *There's a sense in America that* ***France is a country*** *of culture*

Examples of Patterns

- Extracting country names
  - “X is a country”
  - ***Canada is a country*** *in north America*
  - *There's a sense in America that* ***France is a country*** *of culture*
- Extracting hyponymy relations
  - “X such as Y”

Examples of Patterns

- Extracting country names
  - “X is a country”
  - ***Canada is a country*** *in north America*
  - *There's a sense in America that* ***France is a country*** *of culture*
- Extracting hyponymy relations
  - “X such as Y”
  - *Cut the stems of boxed* ***flowers such as roses***
  - *I am responsible for preparing a range of* ***fruits such as apples***

Drawbacks of using Hand-Crafted Patterns

- Hand-crafted patterns are essentially **rule-based**

Drawbacks of using Hand-Crafted Patterns

- Hand-crafted patterns are essentially **rule-based**
- Require **human (experts) labor**

Drawbacks of using Hand-Crafted Patterns

- Hand-crafted patterns are essentially **rule-based**
- Require **human (experts) labor**
- Language-specific

Drawbacks of using Hand-Crafted Patterns

- Hand-crafted patterns are essentially **rule-based**
- Require **human (experts) labor**
- Language-specific
- Poor coverage

Flexible Patterns

- Patterns that are extracted **automatically**

Flexible Patterns

- Patterns that are extracted **automatically**
- Instead of defining a set of **fixed** patterns, we define **meta-patterns**
  - **Structures** of (potential) patterns
  - High frequency words (HFWs) are used instead of fixed words
  - E.g., "HFW_1 *X* HFW_2 *Y*"

Flexible Patterns

- Patterns that are extracted **automatically**
- Instead of defining a set of **fixed** patterns, we define **meta-patterns**
  - **Structures** of (potential) patterns
  - High frequency words (HFWs) are used instead of fixed words
  - E.g., "HFW_1 *X* HFW_2 *Y*"
- Frequent and informative patterns are selected

Extracted Flexible Patterns

"*HFW_1 X HFW_2 Y*"

- as X as Y
- the X the Y
- an X from Y
- from X to Y
- a X has Y
- to X big Y
- in X the Y
- an X do Y
- to X and Y
- …

Extracted Flexible Patterns

"HFW_1 X HFW_2 Y"

- as X as Y
- the X the Y
- an X from Y
- from X to Y
- a X has Y
- to X big Y
- in X the Y
- an X do Y
- to X and Y
- …

Benefits of using Flexible Patterns

- Flexible patterns are computed in a **fully unsupervised** manner
  - Do not require manual labor
  - Language and domain independent
  - Large coverage

- Flexible patterns have been shown to be useful in a range of NLP applications
  - Snow et al., 2005; Davidov and Rappoport, 2006; 2008a,b;2009; Davidov, Rappoport and Koppel 2007; Turney, 2008

Discovery of Semantic Noun Categories
Davidov and Rappoport, ACL 2006

- Cluster nouns into meaningful semantic groups

Discovery of Semantic Noun Categories
Davidov and Rappoport, ACL 2006

- Cluster nouns into meaningful semantic groups

- Use **symmetric** flexible patterns
  - *"X and Y", "X as well as Y", "neither X nor Y"*
  - Both "*cats and dogs*" and "*dogs and cats*" appear in the corpus

Discovery of Semantic Noun Categories
Davidov and Rappoport, ACL 2006

- Cluster nouns into meaningful semantic groups

- Use **symmetric** flexible patterns
  - *"X and Y", "X as well as Y", "neither X nor Y"*
  - Both "*cats and dogs*" and "*dogs and cats*" appear in the corpus

- Discovered categories include
  - Chemical elements, university names, languages, fruits, fishing baits…
  - Evaluation on English and Russian

Discovery of Concept-Specific Relationships
Davidov, Rappoport and Koppel, ACL 2007

- Given a concept *C*, find other concepts with some relation to it
  - (**Italy**) ➔ (**Rome**), (**Italian**), (**Tuscany**), …

Discovery of Concept-Specific Relationships
Davidov, Rappoport and Koppel, ACL 2007

- Given a concept *C*, find other concepts with some relation to it
  - (**Italy**) ➔ (**Rome**), (**Italian**), (**Tuscany**), …
- Find words that participate in flexible patterns along with *C*
  - "**Rome** is the capital of **Italy**", "**Tuscany** is a region in central **Italy**"

Discovery of Concept-Specific Relationships
Davidov, Rappoport and Koppel, ACL 2007

- Given a concept *C*, find other concepts with some relation to it
  - (**Italy**) → (**Rome**), (**Italian**), (**Tuscany**), …
- Find words that participate in flexible patterns along with *C*
  - "**Rome** is the capital of **Italy**", "**Tuscany** is a region in central **Italy**"
- Find other pairs of words for which the same relation exist
  - "**Paris** is the capital of **France**", "**Henan** is a region in central **China**"

Discovery of Concept-Specific Relationships
Davidov, Rappoport and Koppel, ACL 2007

- Given a concept *C*, find other concepts with some relation to it
  - (**Italy**) ➔ (**Rome**), (**Italian**), (**Tuscany**), …
- Find words that participate in flexible patterns along with *C*
  - "**Rome** is the capital of **Italy**", "**Tuscany** is a region in central **Italy**"
- Find other pairs of words for which the same relation exist
  - "**Paris** is the capital of **France**", "**Henan** is a region in central **China**"
- Merge groups of similar concept pairs into general relations
  - **capital-of(X,Y)**, **language-spoken-in(X,Y)**, **region-in(X,Y)**

Enhancement of Lexical Concepts
Davidov and Rappoport, EMNLP 2009

- Enhance the semantic specification of given a concept

Enhancement of Lexical Concepts
Davidov and Rappoport, EMNLP 2009

- Enhance the semantic specification of given a concept

- Take a concept and translate it to (45!) various languages
  - Disambiguate translations using web counts

Enhancement of Lexical Concepts
Davidov and Rappoport, EMNLP 2009

- Enhance the semantic specification of given a concept
- Take a concept and translate it to (45!) various languages
  - Disambiguate translations using web counts
- Apply mono-lingual concept acquisition on translated concepts

Enhancement of Lexical Concepts
Davidov and Rappoport, EMNLP 2009

- Enhance the semantic specification of given a concept
- Take a concept and translate it to (45!) various languages
  - Disambiguate translations using web counts
- Apply mono-lingual concept acquisition on translated concepts
- Re-translate new specifications
  - Merge results from different languages and
  - Enhance original specification

Enhancement of Lexical Concepts
Davidov and Rappoport, EMNLP 2009

- Enhance the semantic specification of given a concept
- Take a concept and translate it to (45!) various languages
  - Disambiguate translations using web counts
- Apply mono-lingual concept acquisition on translated concepts
- Re-translate new specifications
  - Merge results from different languages and
  - Enhance original specification
- Human Evaluation on English, Hebrew and Russian

Sentence-Level Semantics

- Flexible patterns can also be used as sentence-level features
  – Sentences that use the same flexible patterns share a semantic property

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

- Identify the content or “style” expressed in the sentence

Sarcasm Detection

Tsur, Davidov and Rappoport, ICWSM 2010

- Automatically detect sarcastic product reviews
  - *“Where am I?” (GPS device)*
  - *“Great for insomniacs” (book)*
  - *“Defective by design” (ipod)*

Sarcasm Detection

Tsur, Davidov and Rappoport, ICWSM 2010

- Automatically detect sarcastic product reviews
  - *“Where am I?” (GPS device)*
  - *“Great for insomniacs” (book)*
  - *“Defective by design” (ipod)*

- Use a semi-supervised classification algorithm
  - Use both syntactic and **flexible pattern** classification features
  - Flexible patterns are the most valuable features

Sarcasm Detection

Tsur, Davidov and Rappoport, ICWSM 2010

- Automatically detect sarcastic product reviews
  - *“Where am I?” (GPS device)*
  - *“Great for insomniacs” (book)*
  - *“Defective by design” (ipod)*

- Use a semi-supervised classification algorithm
  - Use both syntactic and **flexible pattern** classification features
  - Flexible patterns are the most valuable features

- “W can’t X Y Z. Great!”
  - Kindle can’t read protected formats. Great!
  - The new Ipod can’t play mp3 files. Great!

Sentiment Analysis

Davidov, Tsur and Rappoport, Coling 2010

- Detect the sentiment of tweets

Sentiment Analysis

Davidov, Tsur and Rappoport, Coling 2010

- Detect the sentiment of tweets

- Use #hashtags and emoticons as sentiment labels
  - *Everyone needs to hear the new BANE song* #awesome
  - *first batch of wild starter dough failed* #sad

Sentiment Analysis

Davidov, Tsur and Rappoport, Coling 2010

- Detect the sentiment of tweets

- Use #hashtags and emoticons as sentiment labels
  - *Everyone needs to hear the new BANE song* #awesome
  - *first batch of wild starter dough failed* #sad

- Classify tweets using both syntactic and **flexible pattern** features
  - Once again, flexible patterns provide the largest added value

So Far

- Flexible patterns are a great tool for modeling semantics
  - Words, word relations, sentences
  - Fully unsupervised and language independent

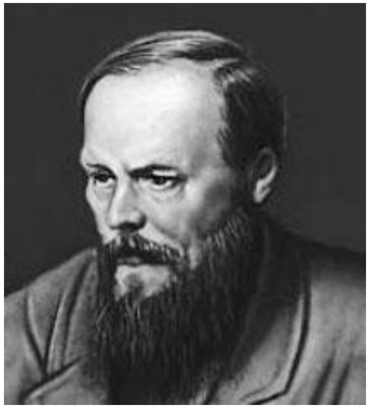
Authorship Attribution of Micro-Messages

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

[+]The Hebrew University, [*]Bar Ilan University

In proceedings of EMNLP 2013

Authorship Attribution

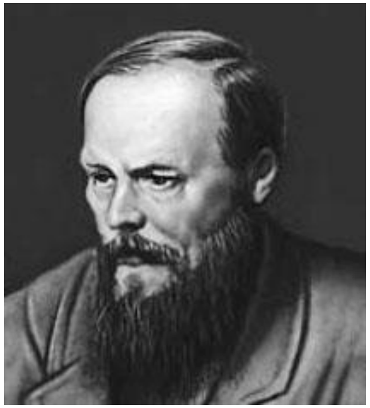


- “To be, or not to be: that is the question”
- “Romeo, Romeo! wherefore art thou Romeo”
- …

- “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 t

- 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 kern  , word n-gram, **flexible patterns** features
- Experiments
  - Varying training set sizes, ber 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

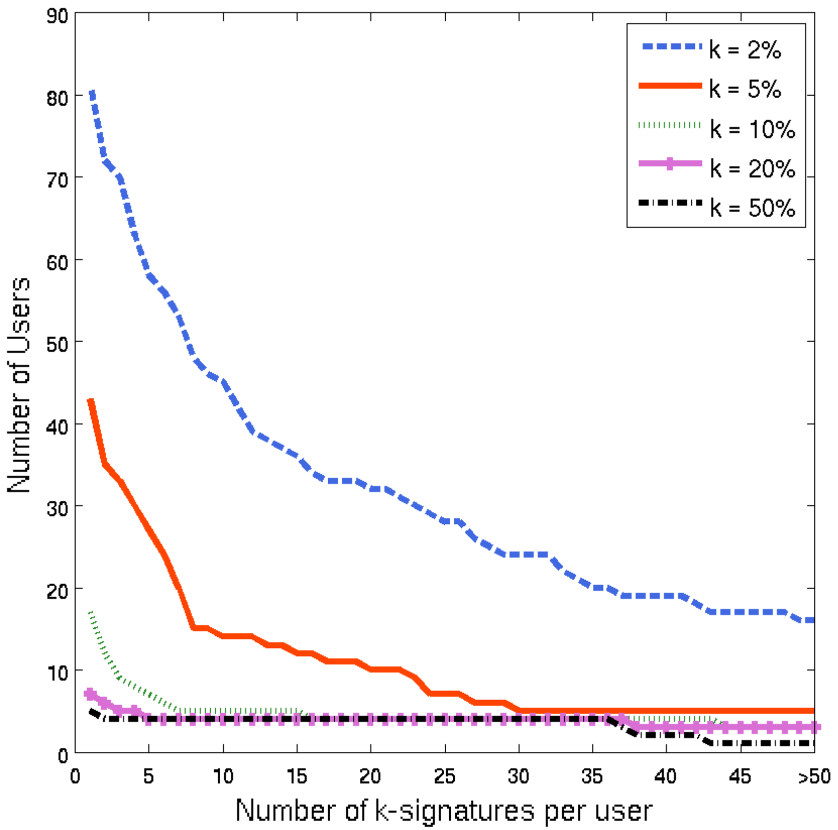
- K-signatures
  - A feature that is unique to a specific author *A*
  - Appears in at least k% of *A*'s training set, while not appearing in the training set of any other user

K-signatures Examples

Signature Type	10%-signature	Examples
Character n-grams	‘ ^_^’	REF oh ok_^_^ Glad you found it!
		Hope everyone is having a good afternoon_^_^
		REF Smirnoff lol keeping the goose in the freezer_^_^
	‘yew ’	gurl **yew**_serving me tea nooch
		REF about wen **yew**_and ronnie see each other
		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?
- Not appearing in the training set of **any other user**?

More about K-signatures

- Implicit?
- Style or content?
- Not appearing in the training set of **any other user**?
- **Useful classification features**

Structured Messages / Bots?

User	20%-signature	Examples
1	**I'm listening to :**	**I'm listening to:** Sigur R?s ? Intro: http://www.last.fm/music/Sigur+R%C3%B3s http://bit.ly/3XJHyb
		I'm listening to: Tina Arena ? In Command: http://www.last.fm/music/Tina+Arena http://bit.ly/7q9E25
		I'm listening to: Midnight Oil ? Under the Overpass: http://www.last.fm/music/Midnight+Oil http://bit.ly/7IH4cg
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: £66.50 **(13 Bids) End Date:** Tuesday Dec-08-2009 17:.. http://bit.ly/7uPt6V
		Microsoft Xbox 360 Game System - Console Only - Working: US $51.99 **(25 Bids) End Date:** Saturday Dec-12-2009 13:.. http://bit.ly/8VgdTv
		Microsoft Sony Playstation 3 (80 GB) Console 6 Months Old: £190.00 **(25 Bids) End Date:** Sunday Dec-13-2009 21:21:39 G.. http://bit.ly/7kwtDS

Methodology

- Features
  - Character n-grams, word n-grams, **flexible patterns**
  - First authorship attribution to use flexible patterns
- Model
  - Multiclass SVM with a linear kernel
- Ten-fold cross validation

Experiments

- Varying training set sizes
  - 10 groups of 50 authors each, 50-1000 training tweets pet author

Experiments

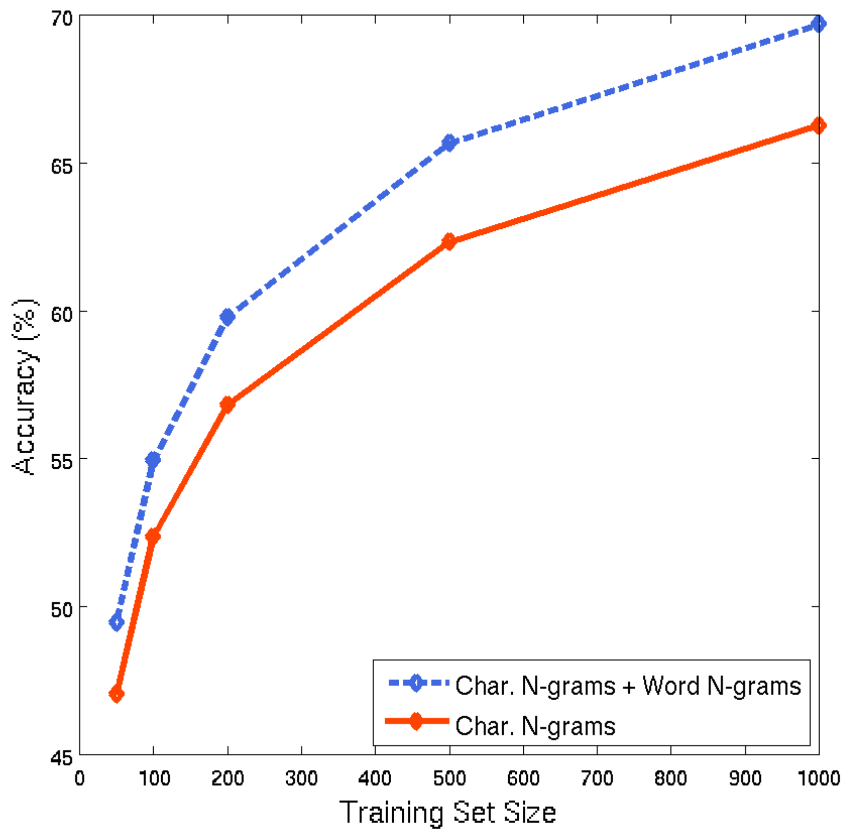
- Varying training set sizes
  - 10 groups of 50 authors each, 50-1000 training tweets pet author
- Varying numbers of authors
  - 50-1000 authors, 200 training tweets per author

Experiments

- Varying training set sizes
  - 10 groups of 50 authors each, 50-1000 training tweets pet author
- Varying numbers of authors
  - 50-1000 authors, 200 training tweets per author
- Recall-precision tradeoff
  - “don’t know” option

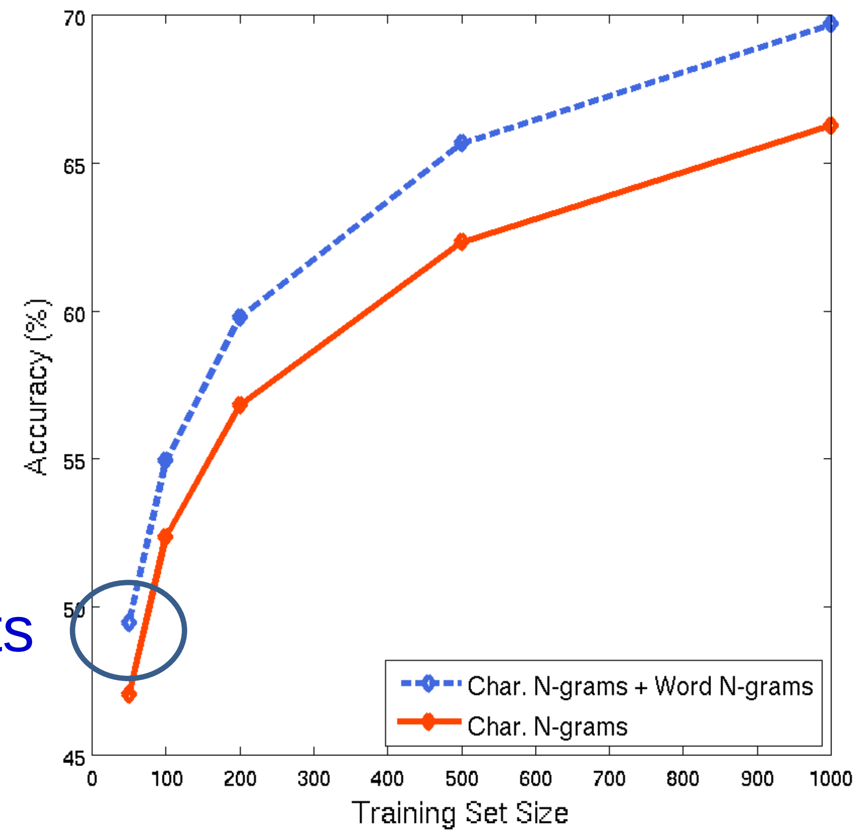
Varying Training Set Sizes

50 Authors (2% Random Baseline)



Varying Training Set Sizes

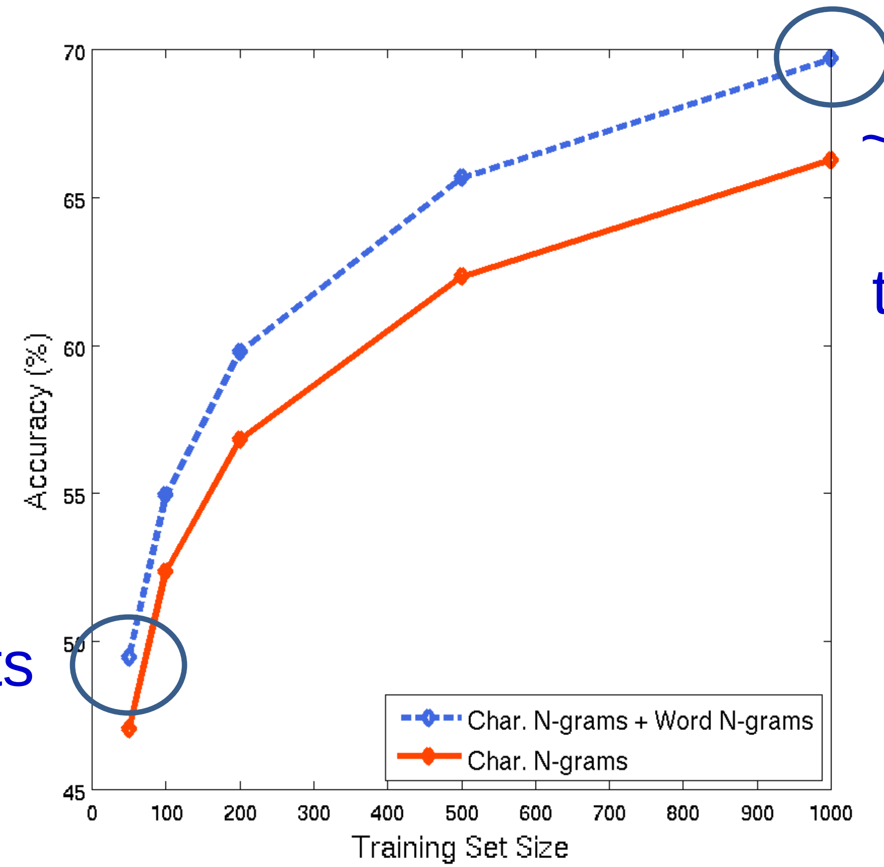
50 Authors (2% Random Baseline)



~50% accuracy
(50 training tweets
per author)

Varying Training Set Sizes

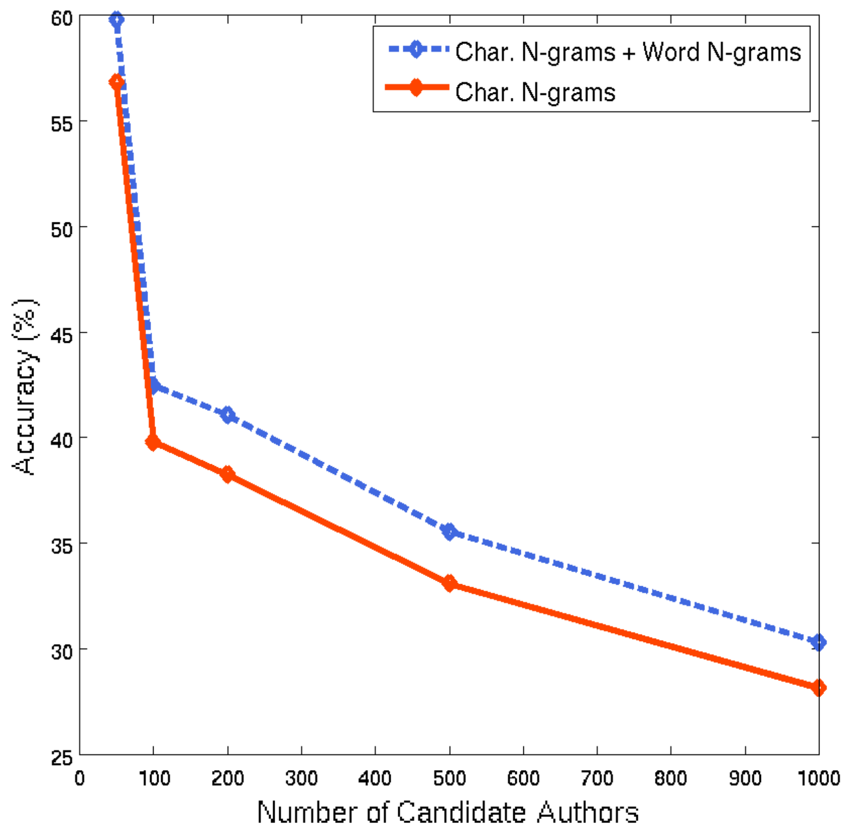
50 Authors (2% Random Baseline)



~70% accuracy (1000 training tweets per author)

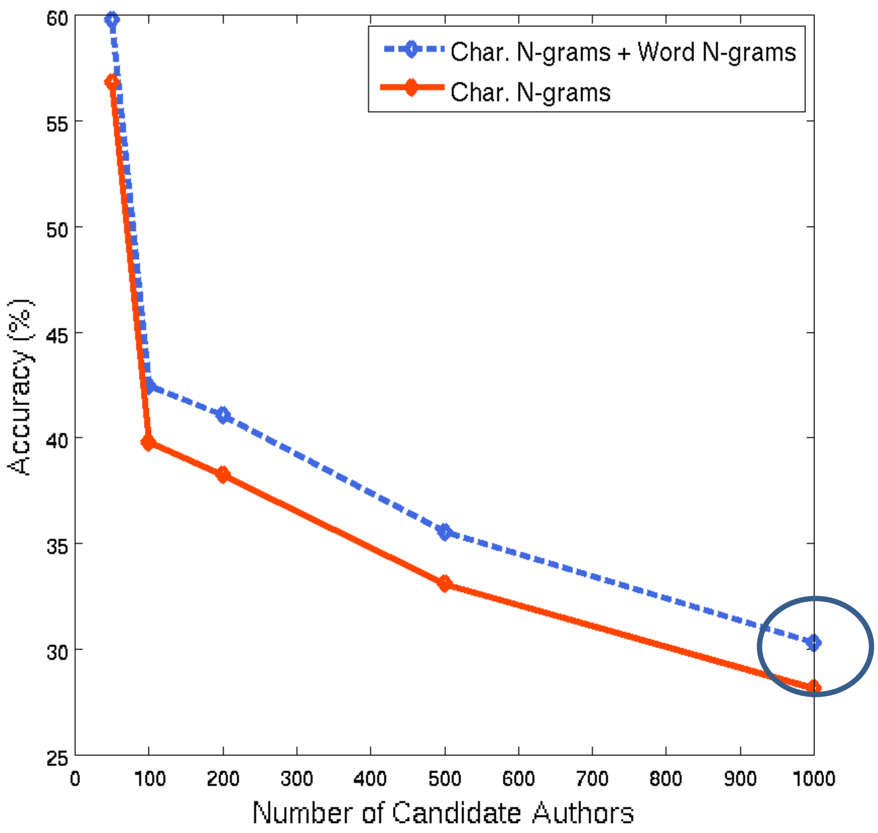
~50% accuracy (50 training tweets per author)

Varying Numbers of Authors
200 Training Tweets per Author



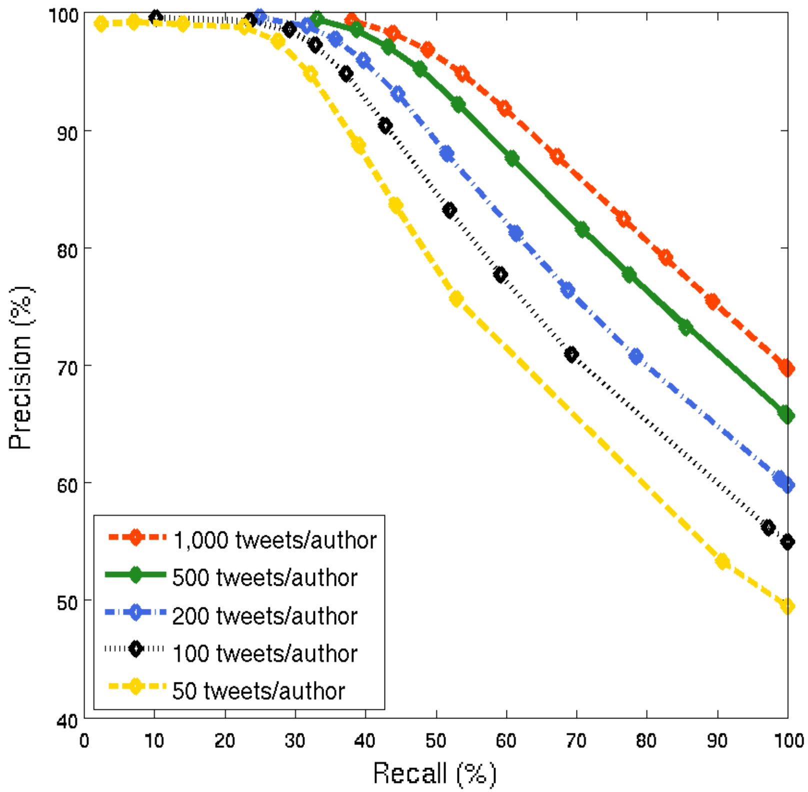
Varying Numbers of Authors

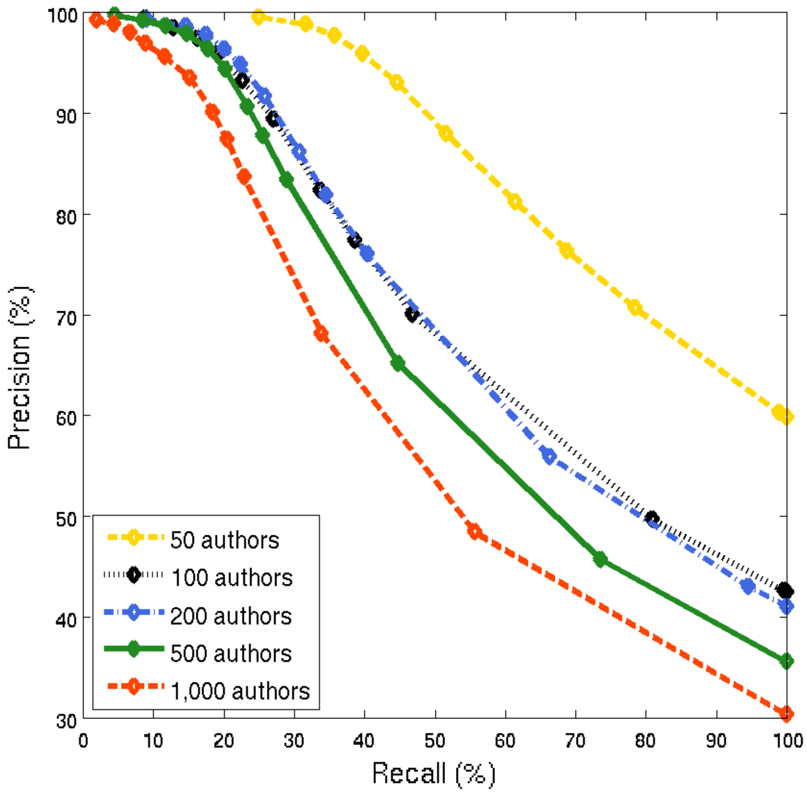
200 Training Tweets per Author



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

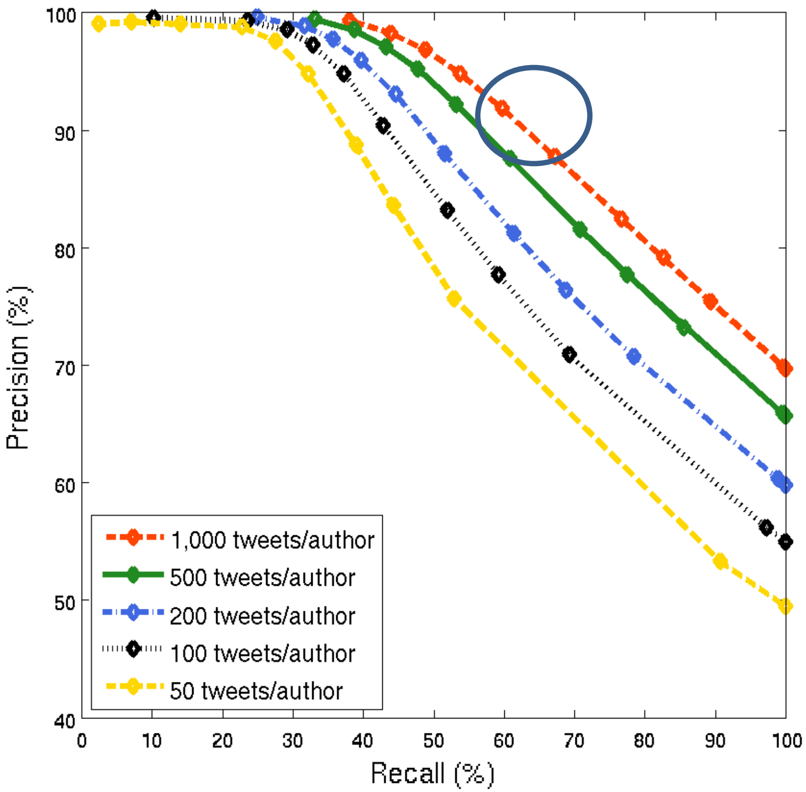
Recall-Precision Tradeoff

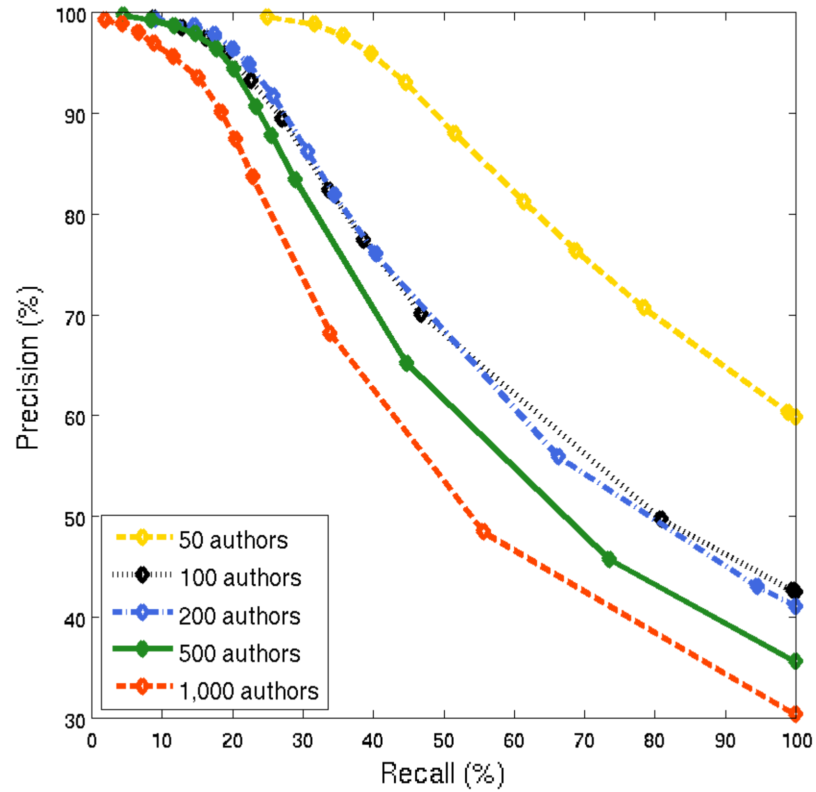




Recall-Precision Tradeoff

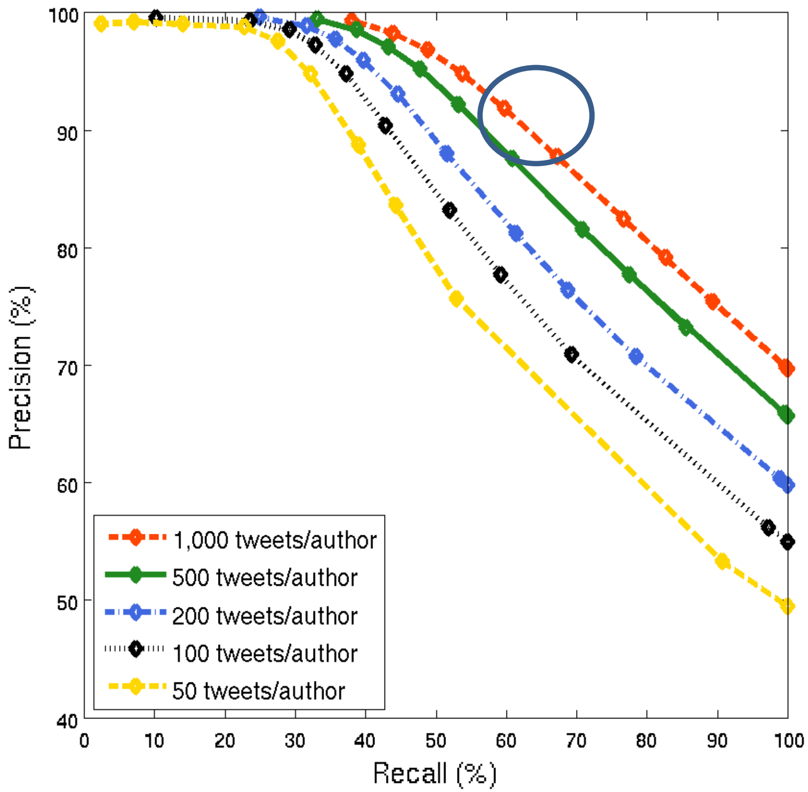
~90% precision,
>~60% recall



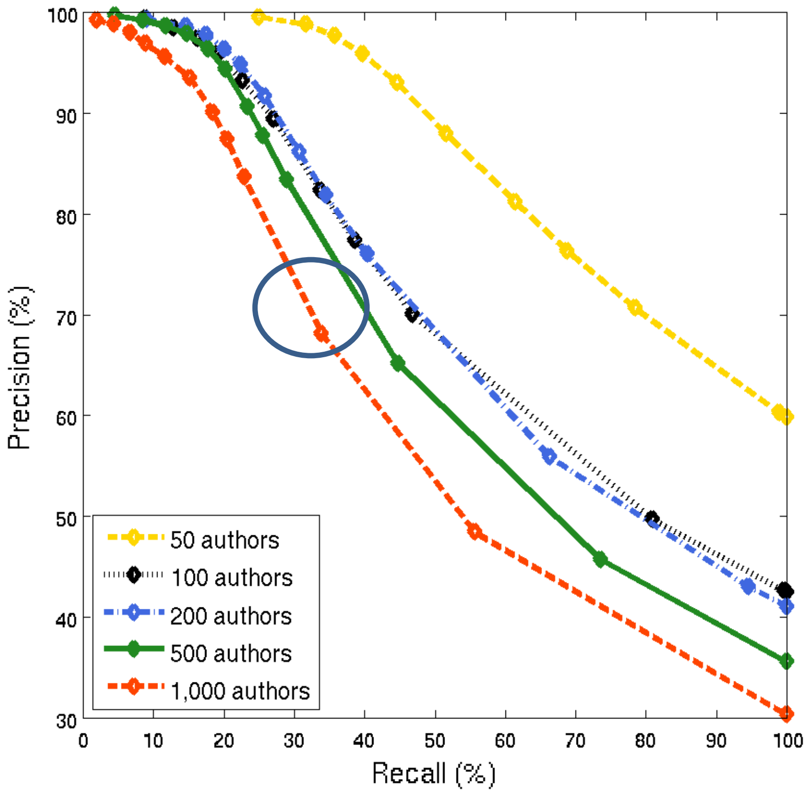


Recall-Precision Tradeoff

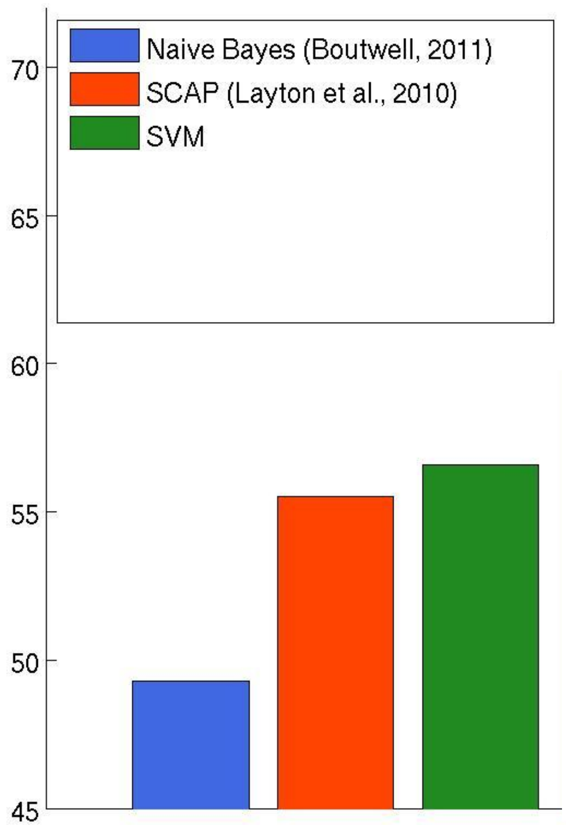
~90% precision,
>~60% recall



~70% precision,
~30% recall

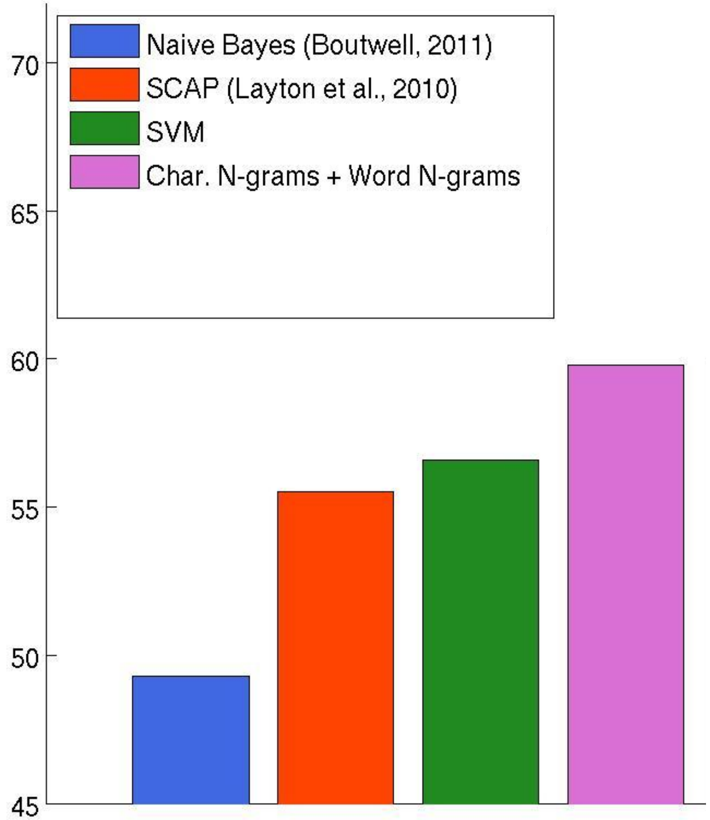


Comparison to Previous Work



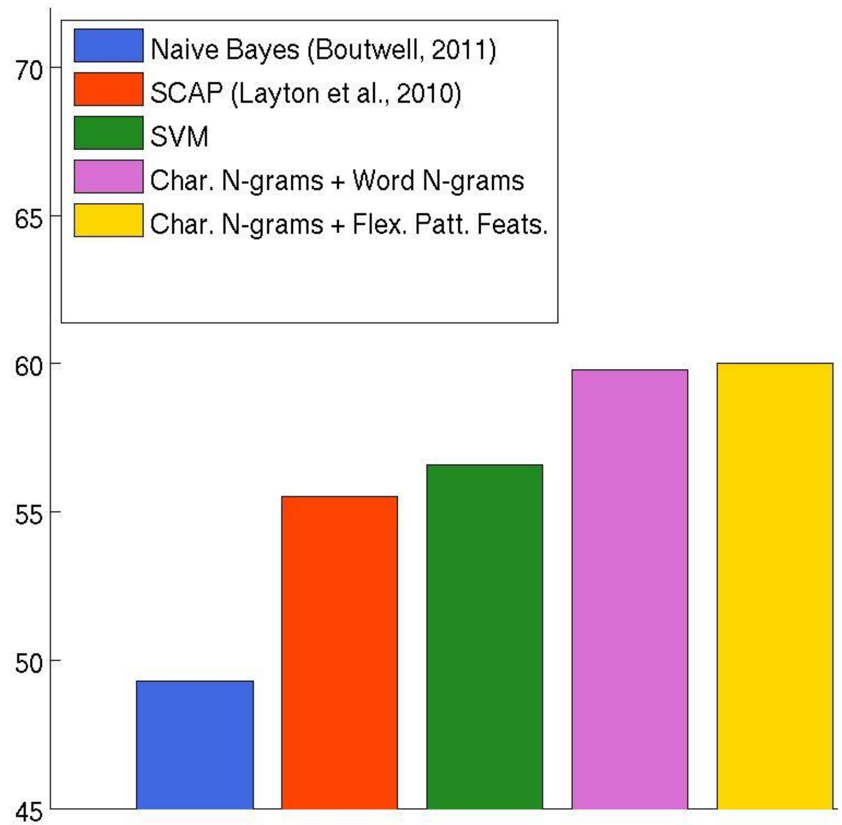
- We thank Robert Layton for providing us with his dataset

Comparison to Previous Work



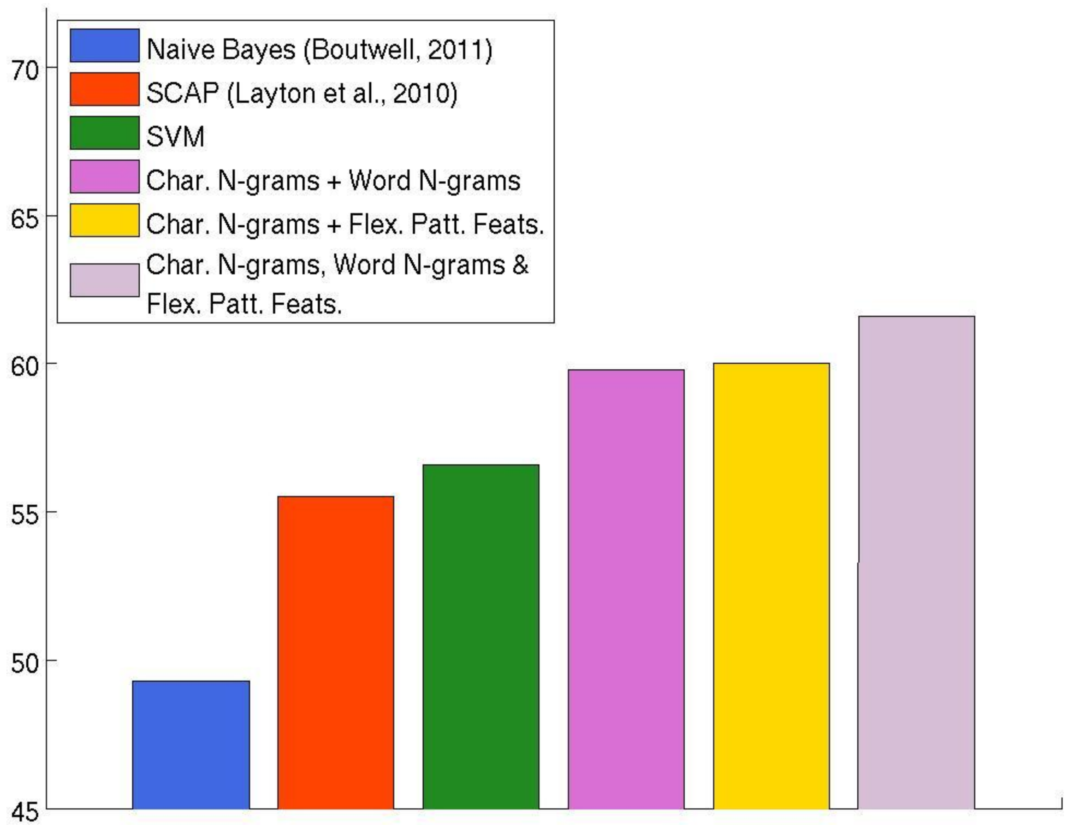
- We thank Robert Layton for providing us with his dataset

Comparison to Previous Work



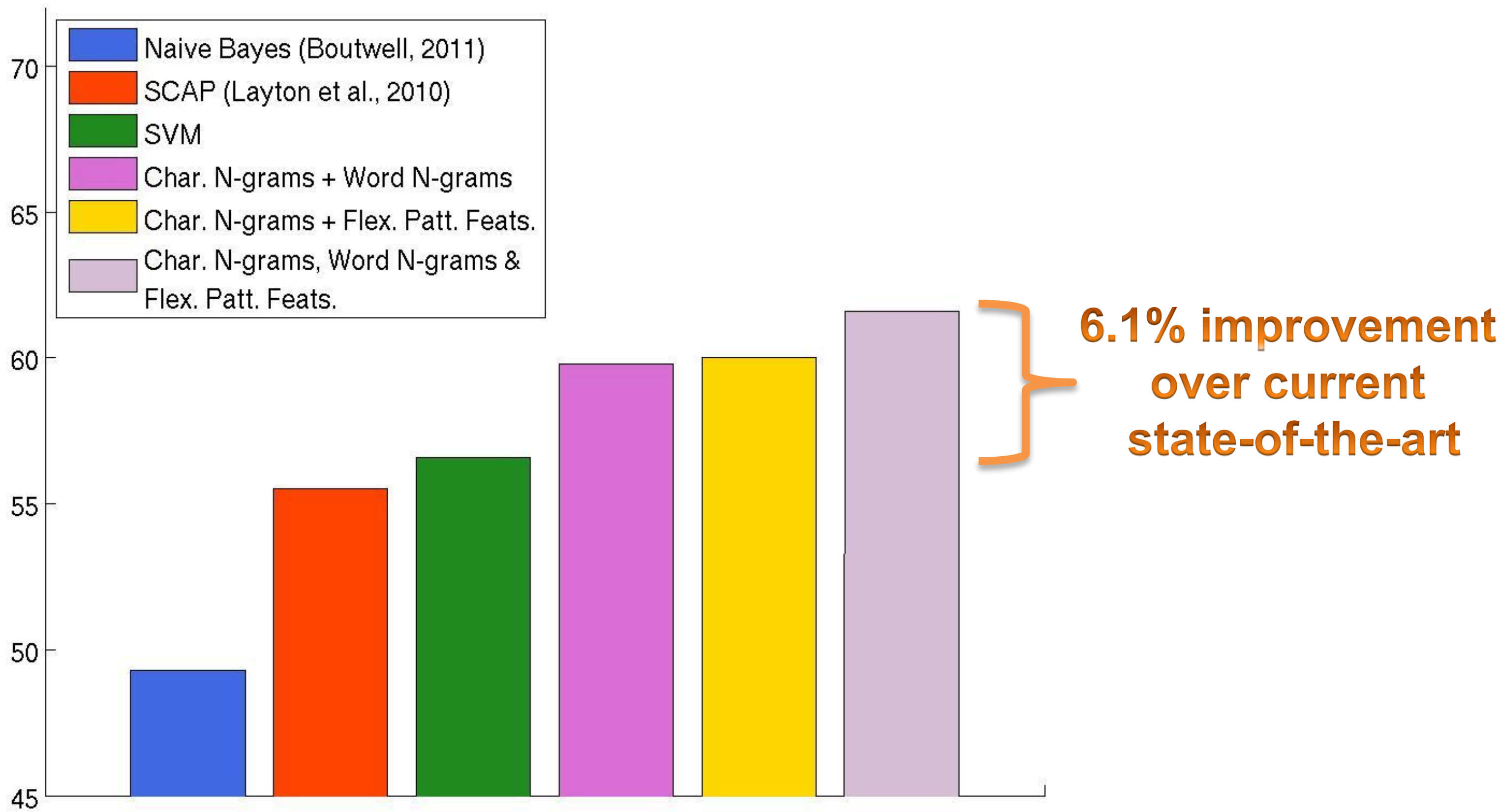
- We thank Robert Layton for providing us with his dataset

Comparison to Previous Work



- We thank Robert Layton for providing us with his dataset

Comparison to Previous Work



- We thank Robert Layton for providing us with his dataset

Flexible Patterns

- 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

- 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

- 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

- 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”*

“the X I”

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

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

Flexible Patterns

- 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"*

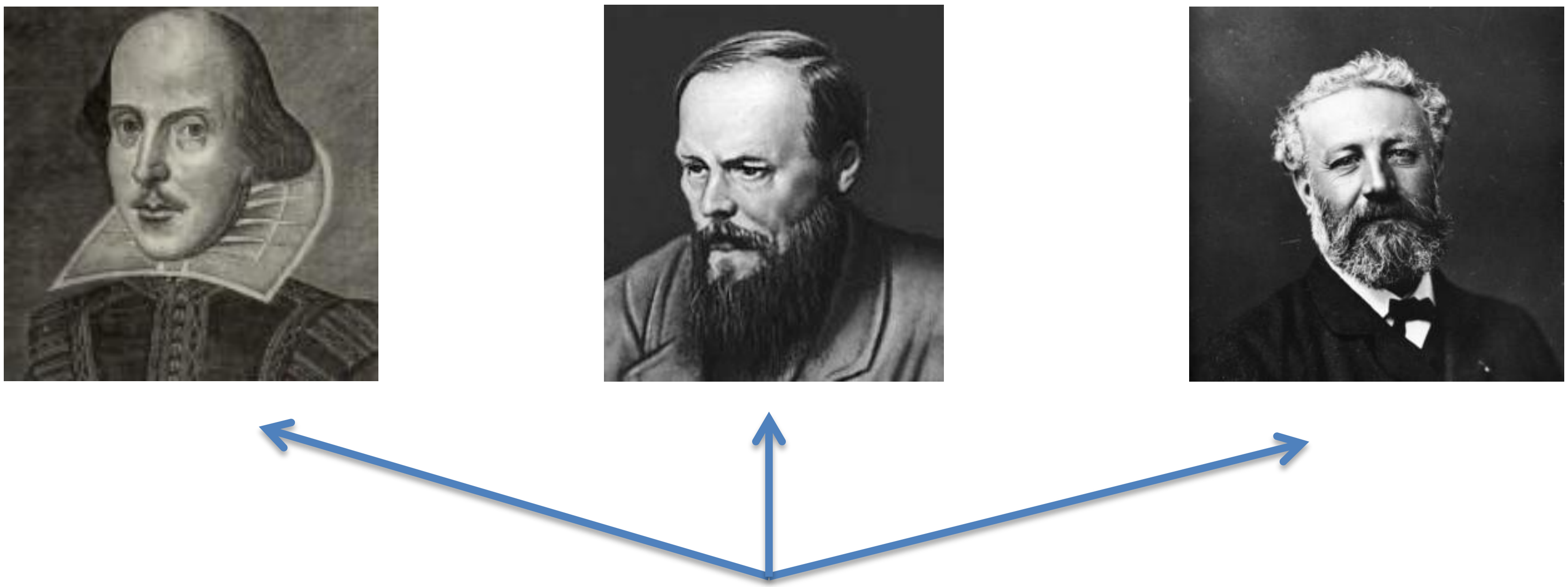
"the X I"

- No word n-gram feature is able to capture this author's style
- Author's character n-grams ("the", " I ") are unindicative
- Flexible patterns obtain a statistically significant improvement over our baselines

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.“

Authorship Attribution

"Love all, trust a few, do wrong to none."

Flexible Patterns and Syntax

- Can flexible patterns represent syntax? Selectional preferences?

Flexible Patterns and Syntax

- Can flexible patterns represent syntax? Selectional preferences?

X will Y

X did not Y

X is Y

X gave Y to Z

Flexible Patterns and Syntax

- Can flexible patterns represent syntax? Selectional preferences?

X will Y

X ha sido Y

X did not Y

X הולך Y

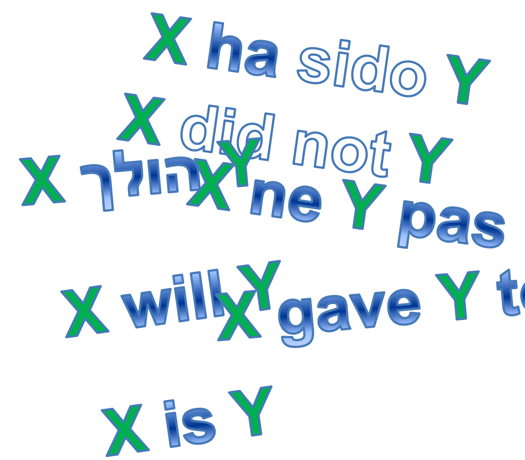
X is Y

X ne Y pas

X gave Y to Z

Flexible Patterns and Syntax

- Can flexible patterns represent syntax? Selectional preferences?



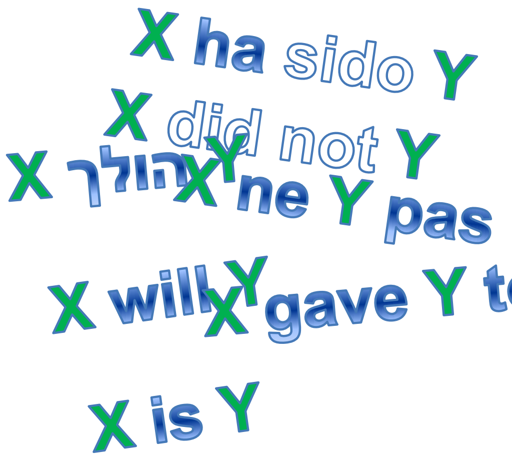
- Use POS information?
  - N did not V

Flexible Patterns and Syntax

- Can flexible patterns represent syntax? Selectional preferences?

- Use POS information?
  - N did not V

- Use morphology?
  - X is Ying



Summary

- Flexible patterns are a great tool for modeling semantics
  - Words, word relations, sentences
  - Fully unsupervised and language independent
- Still a long way to go
  - Model semantics using semantic features (represented by flexible patterns)

roys02@cs.huji.ac.il

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