What a Nasty day: Exploring Mood-Weather Relationship from Twitter

Jiwei Li[1], Xun Wang[2] and Eduard Hovy[2]

[1]Department of Computer Science, Stanford University
[2]Language Technology Institute, Carnegie Mellon University

Oct 26, 2014

Motivation

Sentiment Analysis/ Extraction:

Motivation

Sentiment Analysis/ Extraction:

- Token

Motivation

Sentiment Analysis/ Extraction:

- Token
- Sentence

Motivation

Sentiment Analysis/ Extraction:

- Token
- Sentence
- Document

Motivation

Sentiment Analysis/ Extraction:

- Token
- Sentence
- Document

Feature Based Algorithms

Motivation

Sentiment Analysis/ Extraction:

- Token
- Sentence
- Document

Feature Based Algorithms

- SVM (Joachims., 1999)
- CRF (Lafferty e al., 2001)
- LDA, sLDA (Blei et al., 2003)
- Neural Network (Socher et al., 2013)

Motivation

Long-term Sentiment Analysis

Sentiment towards an entity over a long period of time.

Motivation

Long-term Sentiment Analysis

Sentiment towards an entity over a long period of time.

- food

Motivation

Long-term Sentiment Analysis

Sentiment towards an entity over a long period of time.

- food
- resort

Motivation

Long-term Sentiment Analysis

Sentiment towards an entity over a long period of time.

- food
- resort
- political issues/ figures

Motivation

Long-term Sentiment Analysis

Sentiment towards an entity over a long period of time.

- food
- resort
- political issues/ figures

News Articles,

Motivation

Long-term Sentiment Analysis

Sentiment towards an entity over a long period of time.

- food
- resort
- political issues/ figures

News Articles, Diplomatic Relations

Motivation

Long-term Sentiment Analysis

Sentiment towards an entity over a long period of time.

- food
- resort
- political issues/ figures

News Articles, Diplomatic Relations

Motivation

Long-term Sentiment Analysis

Sentiment towards an entity over a long period of time.

- food
- resort
- political issues/ figures

News Articles, Diplomatic Relations

Motivation

Long-term Sentiment Analysis

Motivation

Long-term Sentiment Analysis

The People's Daily : Governmental Newspaper in People's Republic of China

Motivation

Long-term Sentiment Analysis
The People's Daily : Governmental Newspaper in People's Republic of China

- 1960s: US Imperialism and its lackeys.

Motivation

Long-term Sentiment Analysis

The People's Daily : Governmental Newspaper in People's Republic of China

- 1960s: US Imperialism and its lackeys.
- 2000s: A healthy, steady and developmental relationship between China and US, conforms to the fundamental interests in both countries.

Motivation

Long-term Sentiment Analysis

The People's Daily : Governmental Newspaper in People's Republic of China

- 1960s: US Imperialism and its lackeys.
- 2000s: A healthy, steady and developmental relationship between China and US, conforms to the fundamental interests in both countries.

More technical, less political ! !

Outline

Outline

- Challenges
- Model
- Experiments
- Conclusion

Challenges

Challenge: Multiple entities, multiple sentiments.

Challenges

Challenge: Multiple entities, multiple sentiments.

US imperialism set up a puppet regime in Vietnam and sent expeditionary force. People of Vietnam prevailed over the modern-equipped US troops with a vengeance. The result of Johnson Government's intensifying invasion is that... The heroic Army of Vietnam, obtained great victory in the struggle against the US imperialism.

Negative

Challenges

Challenge: Multiple entities, multiple sentiments.

US imperialism set up a puppet regime in Vietnam and sent expeditionary force. People of Vietnam prevailed over the modern-equipped US troops with a vengeance. The result of Johnson Government's intensifying invasion is that... The heroic Army of Vietnam, obtained great victory in the struggle against the US imperialism.

Positive

Challenges

Challenge: Multiple entities, multiple sentiments.

US imperialism set up a puppet regime in Vietnam and sent expeditionary force. People of Vietnam prevailed over the modern-equipped US troops with a vengeance. The result of Johnson Government 's intensifying invasion is that... The heroic Army of Vietnam, obtained great victory in the struggle against the US imperialism.

Negative

Challenges

Challenge: Multiple entities, multiple sentiments.

US imperialism set up a puppet regime in Vietnam and sent expeditionary force. People of Vietnam prevailed over the modern-equipped US troops with a vengeance. The result of Johnson Government's intensifying invasion is that... The heroic Army of Vietnam, obtained great victory in the struggle against the US imperialism.

Positive

Challenges

Challenge: Multiple entities, multiple sentiments.

US imperialism set up a puppet regime in Vietnam and sent expeditionary force. People of Vietnam prevailed over the modern-equipped US troops with a vengeance. The result of Johnson Government's intensifying invasion is that... The heroic Army of Vietnam, obtained great victory in the struggle against the US imperialism.

Positive

Positive Entities:
People of Vietnam,
the Workers' party
Ho Chi Minh
Vietnam People's Army

Negative Entities:
US imperialism
US troops
Johnson Government
Ngo Dinh Diem

Challenges

Challenge: Multiple entities, multiple sentiments.

US imperialism set up a puppet regime in Vietnam and sent expeditionary force. People of Vietnam prevailed over the modern-equipped US troops with a vengeance. The result of Johnson Government's intensifying invasion is that... The heroic Army of Vietnam, obtained great victory in the struggle against the US imperialism.

Positive

Johnson's Government

Positive Entities:
People of Vietnam,
the Workers' party
Ho Chi Minh
Vietnam People's Army

Negative Entities:
US imperialism
US troops
Johnson Government
Ngo Dinh Diem

Challenges

Challenge: Multiple entities, multiple sentiments.

US imperialism set up a puppet regime in Vietnam and sent expeditionary force. People of Vietnam prevailed over the modern-equipped US troops with a vengeance. The result of Johnson Government's intensifying invasion is that... The heroic Army of Vietnam, obtained great victory in the struggle against the US imperialism.

Positive

Johnson's Government

Not a coreference problem

Positive Entities:
People of Vietnam,
the Workers' party
Ho Chi Minh
Vietnam People's Army

Negative Entities:
US imperialism
US troops
Johnson Government
Ngo Dinh Diem

Challenges

Heavy use of linguistic phenomenon:

Challenges

Heavy use of linguistic phenomenon:

- rhetoric

Challenges

Heavy use of linguistic phenomenon:

- rhetoric
- metaphor (tiger made of paper)

Challenges

Heavy use of linguistic phenomenon:

- rhetoric
- metaphor (tiger made of paper)
- proverb

Challenges

Heavy use of linguistic phenomenon:

- rhetoric
- metaphor (tiger made of paper)
- proverb
- nicknames

Assumptions

Assumptions

- In a single news article, sentiment towards an entity is consistent.

Assumptions

- In a single news article, sentiment towards an entity is consistent.
  Negative (USA) $\rightarrow$ Positive entities irrelevant

Assumptions

- In a single news article, sentiment towards an entity is consistent.
  Negative (USA) $\rightarrow$ Positive entities irrelevant
- Over a certain period of time, sentiments towards an entity are inter-related.

Assumptions

- In a single news article, sentiment towards an entity is consistent.
  Negative (USA) $\rightarrow$ Positive entities irrelevant
- Over a certain period of time, sentiments towards an entity are inter-related.
  Negative (USA) $\rightarrow$ (Negative) tiger made of paper

- Model

Model

Notations:

Sentiment score ranging m (1-7)

- Antagonism
- Tension
- Disharmony
- Neutrality
- Goodness
- Friendliness
- Brotherhood

Model

Notations:

For specific entity e (e.g., USA, Vietnam)

Model

Notations:

For specific entity e (e.g., USA, Vietnam)

- Sentence s: m_s (sentiment toward e at current sentence)

Model

Notations:

For specific entity e (e.g., USA, Vietnam)

- Sentence s: m_s (sentiment toward e at current sentence)
- Document d: m_d (sentiment toward e at current document)

Model

Notations:

For specific entity e (e.g., USA, Vietnam)

- Sentence s: m_s (sentiment toward e at current sentence)
- Document d: m_d (sentiment toward e at current document)
- Time period t (3 months): m_t (sentiment toward e at current time).

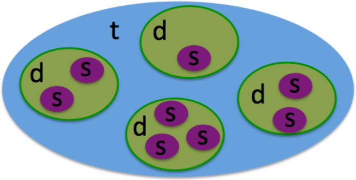
Model

Notations:

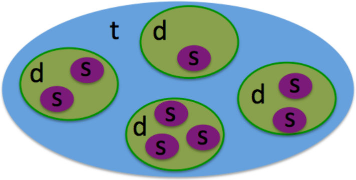
For specific entity e (e.g., USA, Vietnam)

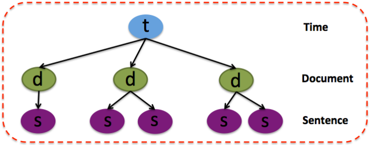
- Sentence s: m_s (sentiment toward e at current sentence)
- Document d: m_d (sentiment toward e at current document)
- Time period t (3 months): m_t (sentiment toward e at current time).

Model

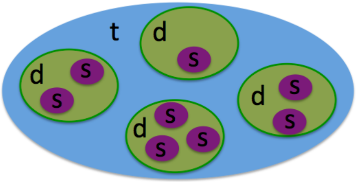


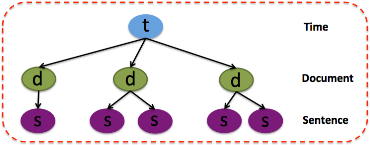
Model

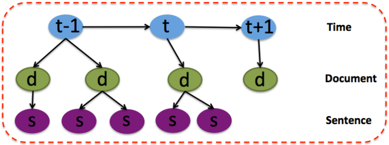




Model


t
d
s
d
s
s
d
s
s
s
d
s
s


t
d
d
d
s
s
s
s
s
Time
Document
Sentence


t-1
t
t+1
d
d
d
d
s
s
s
s
s
Time
Document
Sentence

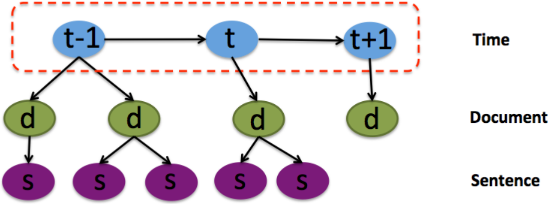
Model

Model

Markov Property at time level:

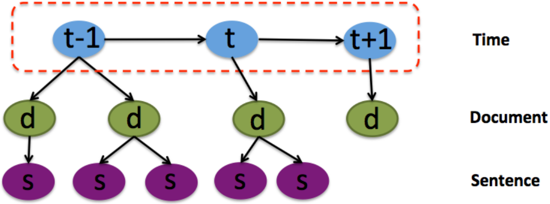
Model

Markov Property at time level:



Model

Markov Property at time level:



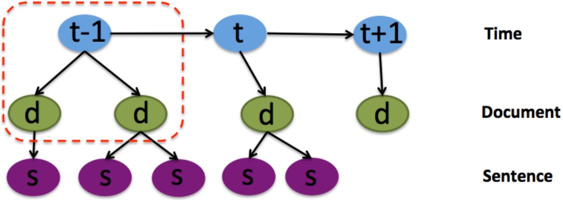
$$m_t \sim Poisson(m_{t-1})$$

Model

Sample document sentiment from time sentiment

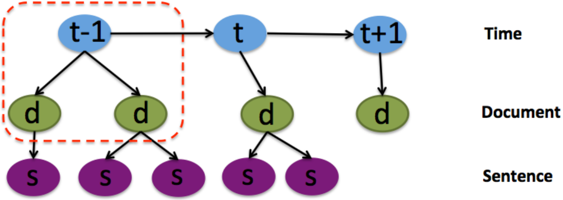
Model

Sample document sentiment from time sentiment



Model

Sample document sentiment from time sentiment



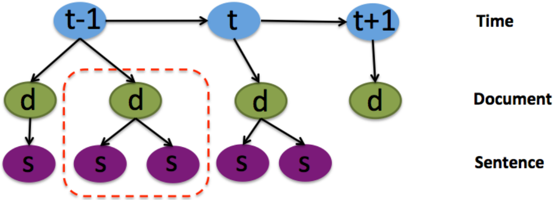
$$m_d \sim Poisson(m_t)$$

Model

Sample sentence sentiment from document sentiment

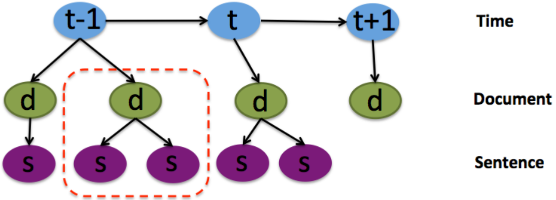
Model

Sample sentence sentiment from document sentiment



Model

Sample sentence sentiment from document sentiment



$$m_S \sim Poisson(m_d)$$

Model

Bootstrapping

- Gibbs sampling for sentiment score at time- document- level given sentence-level score.

Model

Bootstrapping

- Gibbs sampling for sentiment score at time- document- level given sentence-level score.
- Given document-level score, expand entity cluster, subjective lexicon.

Model

Bootstrapping

- Gibbs sampling for sentiment score at time- document- level given sentence-level score.
- Given document-level score, expand entity cluster, subjective lexicon.

Model

Input:

Model

Input:

- Entity: Vietnam

Model

Input:

- Entity: Vietnam
- Sentiment score at document level m_d

Model

Input:

- Entity: Vietnam
- Sentiment score at document level m_d
- Small collection of subjective lexicon: $\{great, 7\}$, $\{evil, 1\}$

Model

Input:

- Entity: Vietnam
- Sentiment score at document level m_d
- Small collection of subjective lexicon: $\{great, 7\}$, $\{evil, 1\}$
- $T = \{Vietnam\}$: List of entities.

Model

Model

(1) Vietnam is a great country.

(2) People's Army are great.

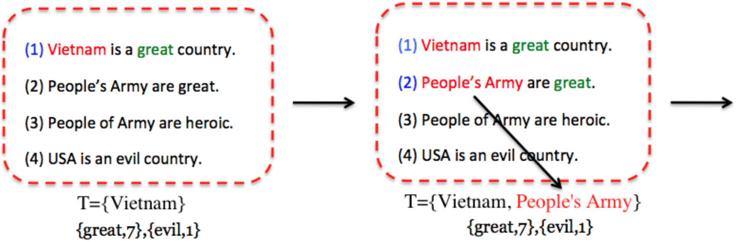
(3) People of Army are heroic.

(4) USA is an evil country.

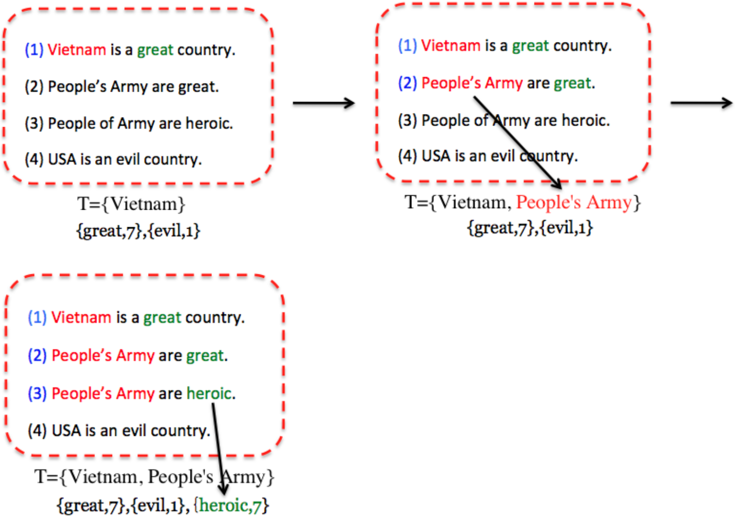
T={Vietnam}

{great,7},{evil,1}

Model


(1) Vietnam is a great country.
(2) People's Army are great.
(3) People of Army are heroic.
(4) USA is an evil country.
T={Vietnam}
{great,7},{evil,1}
(1) Vietnam is a great country.
(2) People's Army are great.
(3) People of Army are heroic.
(4) USA is an evil country.
T={Vietnam, People's Army}
{great,7},{evil,1}

Model


(1) Vietnam is a great country.
(2) People's Army are great.
(3) People of Army are heroic.
(4) USA is an evil country.
T={Vietnam}
{great,7},{evil,1}
(1) Vietnam is a great country.
(2) People's Army are great.
(3) People of Army are heroic.
(4) USA is an evil country.
T={Vietnam, People's Army}
{great,7},{evil,1}
(1) Vietnam is a great country.
(2) People's Army are great.
(3) People's Army are heroic.
(4) USA is an evil country.
T={Vietnam, People's Army}
{great,7},{evil,1}, {heroic,7}

Model

(1) Vietnam is a great country.

(2) People's Army are great.

(3) People of Army are heroic.

(4) USA is an evil country.

T={Vietnam}

{great,7},{evil,1}

(1) Vietnam is a great country.

(2) People's Army are great.

(3) People of Army are heroic.

(4) USA is an evil country.

T={Vietnam, People's Army}

{great,7},{evil,1}

(1) Vietnam is a great country.

(2) People's Army are great.

(3) People's Army are heroic.

(4) USA is an evil country.

T={Vietnam, People's Army}

{great,7},{evil,1}, {heroic,7}

(1) Vietnam is a great country.

(2) People's Army are great.

(3) People's Army are heroic.

(4) USA is an evil country.

T={Vietnam, People's Army}

{great,7},{evil,1}, {heroic,7}

Model

A is great

Model

A is great
B is heroic

Model

A is great
B is heroic
C is evil.

Model

A is great
B is heroic
C is evil.

Albania Workers' party is the glorious party of Marxism and Leninism

Model

A is great
B is heroic
C is evil.

Albania Workers' party is the glorious party of Marxism and Leninism
We strongly warn Soviet Revisionism.

Target/Expression Extraction

A is Great .

B is evil.

Red: Target

Green: Expression

Target/Expression Extraction

A is Great .

B is evil.

Red: Target

Green: Expression

Albania Workers' party is the glorious party of Marxism and Leninism

The heroic people of Vietnam obtained great victory

Target/Expression Extraction

Segment based Sequence Model:
Semi-CRF (Sarawagi and Cohen, 2004; Yang and Cardie, 2013)

Target/Expression Extraction

Segment based Sequence Model:
Semi-CRF (Sarawagi and Cohen, 2004; Yang and Cardie, 2013)

600 manually labeled sentences

Features

- word, part of speech tag, word length, NER

Target/Expression Extraction

Segment based Sequence Model:
Semi-CRF (Sarawagi and Cohen, 2004; Yang and Cardie, 2013)

600 manually labeled sentences

Features

- word, part of speech tag, word length, NER
- window context (token, NER, POS ...)

Target/Expression Extraction

Segment based Sequence Model:
Semi-CRF (Sarawagi and Cohen, 2004; Yang and Cardie, 2013)

600 manually labeled sentences

Features

- word, part of speech tag, word length, NER
- window context (token, NER, POS ...)
- Subjectivity lexicon features

Target/Expression Extraction

Segment based Sequence Model:
Semi-CRF (Sarawagi and Cohen, 2004; Yang and Cardie, 2013)

600 manually labeled sentences

Features

- word, part of speech tag, word length, NER
- window context (token, NER, POS ...)
- Subjectivity lexicon features
- Syntactic features, e.g., VPcluster, VPpred, VParg (Yang and Cardie, 2013)

model

Summary of Model:

model

Summary of Model:

- Identify Target/Expression for sentences.

model

Summary of Model:

- Identify Target/Expression for sentences.
- Obtain score for a few sentences whose expressions are found in the subjective list.

model

Summary of Model:

- Identify Target/Expression for sentences.
- Obtain score for a few sentences whose expressions are found in the subjective list.
- Gibbs sampling to estimate document- time- level sentiment score.

model

Summary of Model:

- Identify Target/Expression for sentences.
- Obtain score for a few sentences whose expressions are found in the subjective list.
- Gibbs sampling to estimate document- time- level sentiment score.
- Bootstrapping: expand subjective list and entity cluster.

Some details

- Strict error prevention in bootstrapping.

Some details

- Strict error prevention in bootstrapping.
- Ignore sentences with multiple targets or with negations. Ignore long or short sentences.

Some details

- Strict error prevention in bootstrapping.
- Ignore sentences with multiple targets or with negations. Ignore long or short sentences.
- ICTCLAS for Chinese segmentation (Zhang et al., 2003)

Some details

- Strict error prevention in bootstrapping.
- Ignore sentences with multiple targets or with negations.
  Ignore long or short sentences.
- ICTCLAS for Chinese segmentation (Zhang et al., 2003)
- Chunk consecutive words.
  Example: [People's / Army].
  Chinese encyclopedia (e.g., Baidu encyclopedia and Chinese Wikipedia)

Some details

- Strict error prevention in bootstrapping.
- Ignore sentences with multiple targets or with negations. Ignore long or short sentences.
- ICTCLAS for Chinese segmentation (Zhang et al., 2003)
- Chunk consecutive words.
  Example: [People's / Army].
  Chinese encyclopedia (e.g., Baidu encyclopedia and Chinese Wikipedia)
- Compound sentences are segmented into clauses based on parse trees.

- Experiments

Experiments

Target/Expression Extraction

Single: Sentences with only one target.

Experiments

Target/Expression Extraction
Single: Sentences with only one target.

Experiments

Target/Expression Extraction
Single: Sentences with only one target.
Comparing Semi-CRF with CRF

	P	R	F
	Total		
semi-CRF	**0.74**	**0.78**	**0.76**
CRF	0.73	0.66	0.68
	Single		
semi-CRF	**0.87**	**0.92**	**0.90**
CRF	0.80	0.87	0.83

Experiments

Subjective Lexicon

antagonism (m=1)	残暴(extremely cruel), 敌人(enemy)
tension (m=2)	愤慨(indignation), 侵犯(offend)
disharmony (m=3)	失望(disappointed), 遗憾(regret)
neutrality (m=4)	关切, 关注(concern)
goodness (m=5)	发展的(developmental), 尊重(respect)
friendship (m=6)	友谊(friendship), 朋友(friend)
brotherhood (m=7)	伟大(firmly), 兄弟(brother)

Experiments

Foreign Relation Evaluation

Gold Standards: Qualitative Political Analysis

Experiments

Foreign Relation Evaluation

Gold Standards: Qualitative Political Analysis

Institute of Modern International Relations, Tsinghua University,

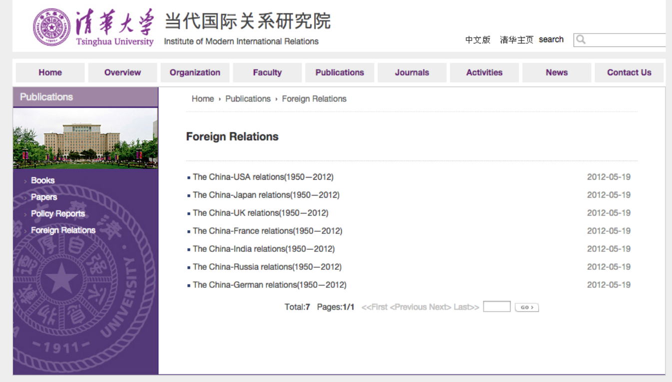
Experiments

Foreign Relation Evaluation

Gold Standards: Qualitative Political Analysis

Institute of Modern International Relations, Tsinghua University,

Experiments



Experiments

清華大学 Tsinghua University 当代国际关系研究院 Institute of Modern International Relations

中文版 清华主页 search

Home | Overview | Organization | Faculty | Publications | Journals | Activities | News | Contact Us

Home › Publications › Foreign Relations › Content

The China-USA relations(1950—2012)

	1	2	3	4	5	6	7
1950	-7.5	-7.5	-7.5	-7.5	-7.5	-7.6	-7.6
1951	-8.2	-8.2	-8.2	-8.2	-8.3	-8.2	-7.8
1952	-7.9	-7.9	-7.9	-7.7	-7.8	-7.8	-7.8
1953	-8.0	-8.0	-7.8	-7.4	-7.3	-7.2	-6.1
1954	-6.1	-6.0	-6.0	-5.9	-5.9	-5.8	-5.8
1955	-6.9	-7.0	-7.0	-6.7	-6.7	-6.7	-6.3
1956	-5.9	-5.9	-5.9	-5.9	-5.9	-5.9	-5.9
1957	-5.8	-5.8	-5.8	-5.8	-5.8	-5.9	-5.9
1958	-5.7	-5.7	-5.7	-5.7	-5.7	-5.7	-5.7
1959	-5.7	-5.7	-5.7	-5.7	-5.7	-5.7	-5.7

Experiments

Evaluation Matrics: Pearson Correlation

Baselines

- Bootstrapping+NoTime

Experiments

Evaluation Matrics: Pearson Correlation

Baselines

- Bootstrapping+NoTime
- Document-level SVR (Joachims, 1999; Pang and Lee, 2008), bag of words

Experiments

Evaluation Matrics: Pearson Correlation

Baselines

- Bootstrapping+NoTime
- Document-level SVR (Joachims, 1999; Pang and Lee, 2008), bag of words
- Supervised-LDA

Experiments

Model	Pearson Correlation
SVR	0.482
sLDA	0.527
No-Time	0.808
Ours	0.884

Experiments

Model	Pearson Correlation
SVR	0.482
sLDA	0.527
No-Time	0.808
Ours	0.884

Experiments

Analysis

Experiments

Analysis

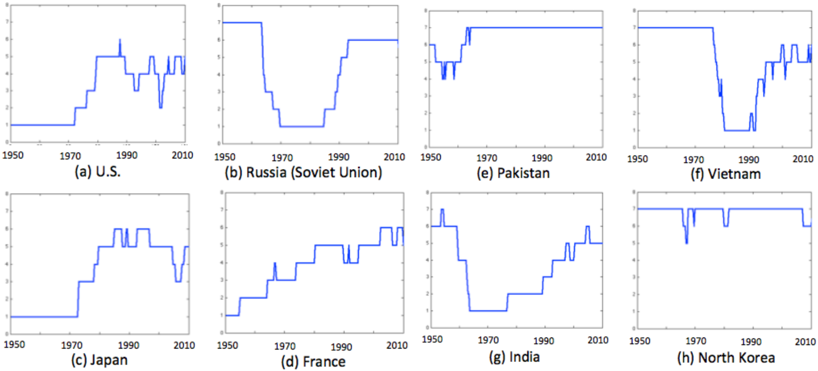
We have no permanent allies, no permanent friends, but only permanent interests -Lord Palmerston

Experiments

Analysis

We have no permanent allies, no permanent friends, but only permanent interests -Lord Palmerston

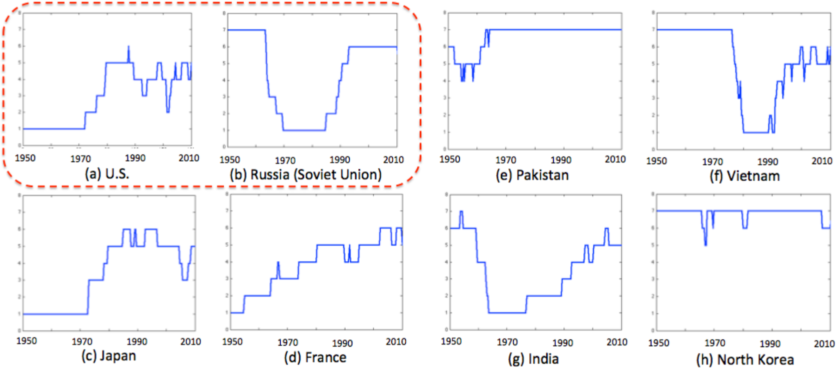
Experiments



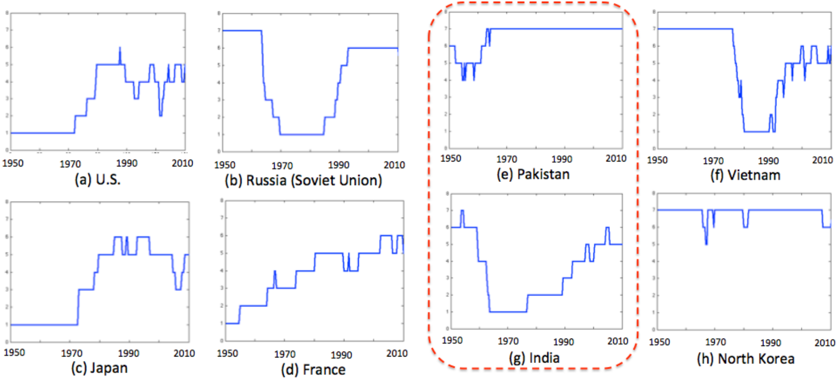
(a) U.S. (b) Russia (Soviet Union) (e) Pakistan (f) Vietnam

(c) Japan (d) France (g) India (h) North Korea

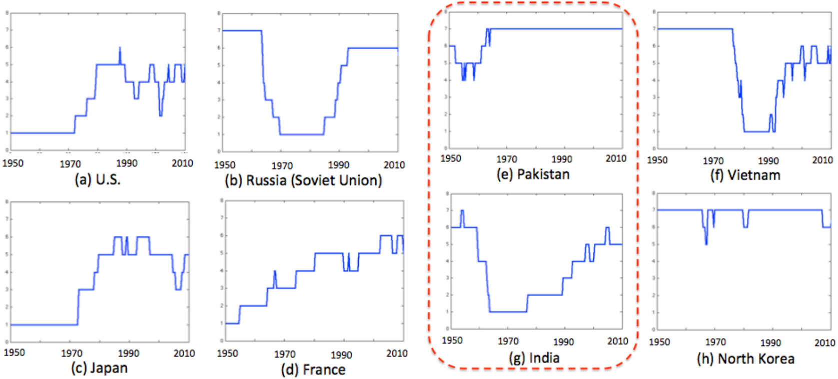
Experiments



Experiments

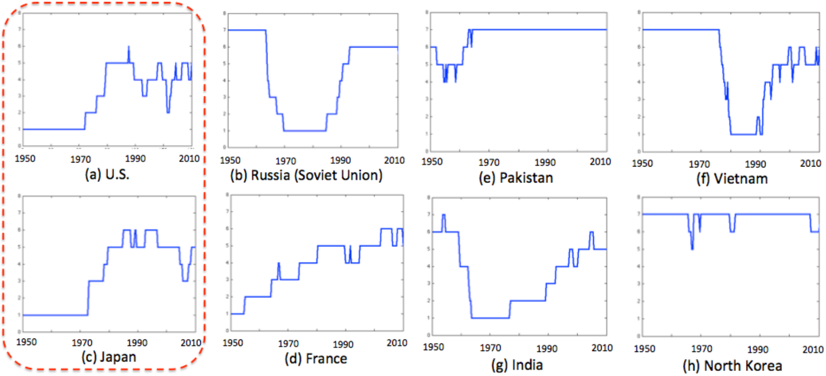


Experiments

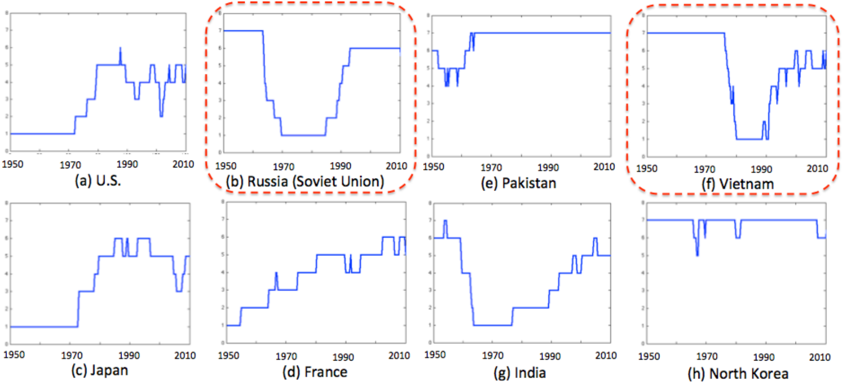


The enemy of my enemy is my friend -Arabic proverb

Experiments



Experiments



Conclusion

Conclusion

Conclusion

Conclusion

- We propose a semi-supervised algorithm that harnesses higher level information (i.e., document-level, time-level).

Conclusion

Conclusion

- We propose a semi-supervised algorithm that harnesses higher level information (i.e., document-level, time-level).
- Quantitative evaluation of diplomatic relations that may facilitate political scientists' work.

Target/ Expression dataset:
`http://web.stanford.edu/~jiweil/dataset`

Target/ Expression dataset:
`http://web.stanford.edu/~jiweil/dataset`

Thank you !

Target/ Expression dataset:
`http://web.stanford.edu/~jiweil/dataset`

Thank you !

Questions, Suggestions