Sentiment Analysis on the People's Daily

Jiwei Li[1] and Eduard Hovy[2]

[1]Computer Science Department, Stanford University, Stanford, CA 94305, USA
[2]Language Technology Institute, Carnegie Mellon University, PA 15213, USA
jiweil@stanford.edu   ehovy@andrew.cmu.edu

Abstract

We propose a semi-supervised bootstrapping algorithm for analyzing China's foreign relations from the People's Daily. Our approach addresses sentiment target clustering, subjective lexicons extraction and sentiment prediction in a unified framework. Different from existing algorithms in the literature, time information is considered in our algorithm through a hierarchical bayesian model to guide the bootstrapping approach. We are hopeful that our approach can facilitate quantitative political analysis conducted by social scientists and politicians.

1 Introduction

"We have no permanent allies, no permanent friends, but only permanent interests."

-Lord Palmerston

Newspapers, especially those owned by official governments, e.g., *Pravda* from Soviet Union, or *People's Daily* from P.R. China, usually provide direct information about policies and viewpoints of government. As national policies change over time, the tone that newspapers adopt, especially sentiment, changes along with the policies. For example, there is a stark contrast between the American newspapers' attitudes towards Afghanistan before and after 911. Similarly, consider the following examples extracted from the People's Daily[1]:

- People' Daily, Aug 29th, 1963
  All those who are being oppressed and exploited, Unite !! Beat US Imperialism and its lackeys.

- People's Daily, Oct, 20th, 2002
  A healthy, steady and developmental relationship between China and US, conforms to the fundamental interests of people in both countries, and the trend of historical development.

Automatic opinion extraction from newspapers such as people's daily can facilitate sociologists 'or political scientists' research or help political pundits in their decision making process. While our approach applies to any newspaper in principle, we focus here on the People's Daily[2] (Renmin Ribao), a daily official newspaper in the People's Republic of China.

While massive number of works have been introduced in sentiment analysis or opinion target extraction literature (for details, see Section 2), a few challenges limit previous efforts in this specific task: First, the heavy use of linguistic phenomenon in the People's Daily including rhetoric, metaphor, proverb, or even nicknames, makes existing approaches less effective for sentiment inference as identifying these expressions is a hard NLP problem in nature.

Second, as we are more interested in the degree of sentiment rather than binary classification (i.e., positive versus negative) towards an entity (e.g. country or individual) in the news article, straightforward algorithms to apply would be document-level sentiment analysis approaches such as vector machine/regression (Pang et al., 2002) or supervised LDA (Blei and McAuliffe, 2010). A single news article, usually contains different attitudes towards multiple countries or individuals simultaneously (say praising "friends" and criticizing "enemies"), as shown in the following example from the People's Daily of Mar. 17th, 1966:

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

---

[1]Due to the space constraints, we only show the translated version in most of this paper.

vengeance... The result of Johnson Government's intensifying invasion is that.... There will be the day, when people from all over the world execute the heinous US imperialism by hanging on a gibbet.... The heroic people of Vietnam, obtained great victory in the struggle against the USA imperialism...

The switching of praising of Vietnam and criticizing of the USA would make aforementioned document-level machine learning algorithms based on bags of words significantly less effective if not separating attitudes towards Vietnam from toward the USA in the first place. Meanwhile, the separating task is by no means trivial in news articles. While *US imperialism, US troops, Johnson Government, invaders, Ngo Dinh Diem*[3] all point to the USA or its allies, *People of Vietnam, the Workers' party*[4], *Ho Chi Minh*[5], *Vietnam People's Army* point to North Vietnam side. Clustering entities according to sentiment, especially in Chinese, is fundamentally a difficult task. And our goal, trying to identify entities towards whom an article holds the same attitudes, is different from standard coreference resolution, since for us the co-referent group may include several distinct entities.

To address the aforementioned problems, in this paper, we propose a sentiment analysis approach based on the following assumptions:

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

2. *Over a certain period of time, sentiments towards an entity are inter-related.*

The assumptions will facilitate opinion analysis: (1) if we can identify the attitude towards an entity (e.g., Vietnam) in a news article as positive, then negative attitudes expressed in the article are about other entities. (2) The assumption enables sentiment inference for unseen words in a bootstrapping way without having to employ sophisticated NLP algorithms. For example, from 1950s to 1960s, USA is usually referred to as "a tiger made of paper" in translated version. It is a metaphor indicating things that appear powerful (tiger) but weak in na-

---

[3]Leader of South Vietnam

[4]Ruling political party of Vietnam.

[5]One of Founders of Democratic Republic of Vietnam (North Vietnam) and Vietnam Workers' party.

ture (made of paper). If it is first identified that during the designated time period, China held a pretty negative attitude towards the USA based on clues such as common negative expressions (e.g., "evil" or "reactionary"), we can easily induce that "a tiger made of paper", is a negative word.

Based on aforementioned two assumptions, we formulate our approach as a semi-supervised model, which simultaneously bootstrap sentiment target lists, extracts subjective vocabularies and performs sentiment analysis. Time information is considered through a hierarchical bayesian model to guide time-, document-, sentence- and term- level sentiment inference. A small seed set of subjective words constitutes our only source of supervision.

The main contributions of this paper can be summarized as follows:

1. We propose a semi-supervised bootstrapping algorithm tailored for sentiment analysis in the People's daily where time information is incorporated. We are hopeful that sentiment cues can shed insights on other NLP tasks such as coreference or metaphor recognition.

2. In Analytical Political Science, the quantitative evaluation of diplomatic relations is usually a manual task (Robinson and Shambaugh, 1995). We are hopeful that our algorithm can enable automated political analysis and facilitate political scientists' and historians' work.

2 Related Works

Significant research efforts have been invested into sentiment analysis and opinion extraction. In one direction, researchers look into predicting overall sentiment polarity at document-level (Pang and Lee, 2008), aspect-level (Wang et al., 2010; Jo and Oh, 2011), sentence-level (Yang and Cardie, 2014) or tweet-level (Agarwal et al., 2011; Go et al., 2009), which can be treated as a classification/regression problem by employing standard machine-learning techniques, such as Naive Bayesian, SVM (Pang et al., 2002) or supervised-LDA (Blei and McAuliffe, 2010) with different types of features (i.e., unigram, bigram, POS).

Other efforts are focused on targeted sentiment extraction (Choi et al., 2006; Kim and Hovy, 2006; Jin et al., 2009; Kim and Hovy, 2006). Usually, sequence labeling models such as CRF (Lafferty et al.,

2001) or HMM (LIU et al., 2004) are employed for identifying opinion holders (Choi et al., 2005), topics of opinions (Stoyanov and Cardie, 2008) or opinion expressions (e.g. (Breck et al., 2007; Johansson and Moschitti, 2010; Yang and Cardie, 2012)). Kim and Hovy (2004; 2006) identified opinion holders and targets by exploring their semantics rules related to the opinion words. Choi et al. (2006) jointly extracted opinion expressions, holders and their *is-from* relations using an ILP approach. Yang and Cardie (2013) introduced a sequence tagging model based on CRF to jointly identify opinion holders, opinion targets, and expressions.

Methods that relate to our approach include semi-supervised approaches such as pipeline or propagation algorithms (Qiu et al., 2011; Qiu et al., 2009; Zhang et al., 2010; Duyu et al., 2013). Concretely, Qiu et al. (2011) proposed a rule-based semi-supervised framework called *double propagation* for jointly extracting opinion words and targets. Compared to existing bootstrapping approaches, our framework is more general one with less restrictions[6]. In addition, our approach harness global information (e.g. document-level, time-level) to guide the bootstrapping algorithm. Another related work is the approach introduced by O'Connor et al. (O'Connor et al., 2013) that extracts international relations from political contexts.

3 the People's Daily

The People's Daily[7] (Renmin Ribao), established on 15 June 1946, is a daily official newspaper in the People's Republic of China, with a approximate circulation of 2.5 million worldwide. It is widely recognized as the mouthpiece of the Central Committee of the Communist Party of China (CPC) (Wu, 1994). Editorials and commentaries are usually regarded both by foreign observers and Chinese readers as authoritative statements of government policy[8]. According to incomplete statistics, there have benn at least 13 major redesigns (face-liftings) for the Peo-

ple's Daily in history, the most recent in 2013.

4 Model

In this section, we present our model in detail.

4.1 Target and Expression extraction

We first extract *expressions* (attitude or sentiment related terms or phrases) and *target* (entities toward whom the opinion holder (e.g., the People's Daily) holds an attitude). See the following examples:

1. *[Albania Workers' party][T] is the [glorious][E] [party][T] of [Marxism and Leninism][E].*

2. *The [heroic][E] [people of Vietnam][T] obtained [great][E] [victory][E] against [the U.S. imperialism][T,E].*

3. *We strongly [warn][E] Soviet Revisionism][E,T].*

While the majority of subjective sentences omit the opinion holder, as in Examples 1 and 2, there are still a few circumstances where opinion holders (e.g., "we", "Chinese people", "Chinese government") are retained (Example 3). Some words (i.e. U.S. imperialism) can be both target and expression, and there can be multiple targets (Example 2) within one sentence.

We use a semi-Markov Conditional Random Fields (semi-CRFs) (Sarawagi and Cohen, 2004; Okanohara et al., 2006) algorithm for target and expression extraction. Semi-CRF are CRFs that relax the Markovian assumptions and allow for sequence labeling at the segment level. It has been demonstrated more powerful that CRFs in multiple sequence labeling applications including NER (Okanohara et al., 2006), Chinese word segmentation (Andrew, 2006) and opinion expression identification (Yang and Cardie, 2012). Our approach is an extension of Yang and Cardie (2012)'s system[9]. Features we adopted included:

- word, part of speech tag, word length.
- left and right context words within a window of 2 and the correspondent POS tags.
- NER feature.

---

[6]Qiu et al.'s rule base approach makes strong assumptions that consider opinion word to adjectives and targets to be nouns/noun, thus only capable of capturing sentences with simple patterns.

[7]`paper.people.com.cn/rmrb/`

[8]`http://en.wikipedia.org/wiki/People's_Daily`

---

[9]Yang and Cardie's system focuses on expression extraction (not target) and identifies *direct subjective expression (DSE)* and *expressive subjective expression (ESE)*.

- subjectivity lexicon features from dictionary[10]. The lexicon consists of a set of Chinese words that can act as strong or weak cues to subjectivity.
- segment-level syntactic features defined in (Yang and Cardie, 2012).

Most existing NER systems can barely recognize entities such as [Vietnamese People's Army] as a unified name entity in that Chinese parser usually divides them into a series of separate words, namely [Vietnamese/People's Army]. To handle this problem, we first employ the Stanford NER engine[11] and then iteratively 'chunk' consecutive words, at least one of which is labeled as a name entity by the NER engine, before checking whether the chunked entity matches a bag of words contained in Chinese encyclopedia, e.g., Baidu Encyclopedia[12] and Chinese Wikipedia[13].

4.2 Notation

Here we describe the key variables in our model. Let C_i denote the name entity of country i, G_i denote its corresponding collection of news articles. G_i is divided into 60*4=240 time spans (one for each quarter of the year, 60 years in total), $G_i = \{G_{i,t}\}$. $G_{i,t}$ is composed of a series of documents $\{d\}$, and d is composed of a series of sentences $\{S\}$, which is represented as a tuple $S = \{E_S, T_S\}$, where E_S is the expression and T_S is the target of current sentence.

Sentiment Score m: As we are interested in the degree of positiveness or negativeness, we divided international relations into 7 categories: *Antagonism* (score 1), *Tension* (score 2), *Disharmony* (score 3), *Neutrality* (score 4), *Goodness* (score 5), *Friendliness* (score 6), *Brotherhood (Comradeship)* (score 7) based on researches in political science literature[14]. Each of $G_{i,t}$, document d, sentence S and expression term w is associated with a sentiment score $m_{i,t}$, m_d, m_S and m_w, respectively. M denotes the list of subjective terms, $M = \{w, m_w\}$
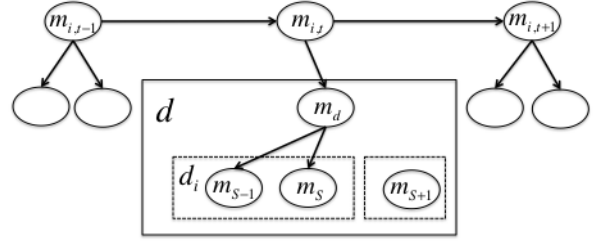
Figure 1: Hierarchical Bayesian Model for Inference

Document Target List T_i^d: We use T_i^d to denote the collection of entity targets in document $d \in G_i$ which the People's daily holds similar attribute towards. For example, suppose document d belongs to Vietnam article collection ($C_i = Vietnam$), T_i^d can be {Vietnam, Workers' party, People's Army, Ho Chi Minh}. While U.S., U.S. troops and Lyndon Johnson are also entity targets found in d, they are not supposed to be included in T_i^d since the author holds opposite attributes.

Sentence List d_i: We further use d_i denotes the subset of sentences in d talking about entities from target list T_i^d. Similarly, in a Vietnam related article, sentences talking about the U.S. are not supposed to be included in d_i.

4.3 Hierarchical Bayesian Markov Model

In our approach, time information is incorporated through a hierarchical Bayesian Markov framework where $m_{i,t}$ is modeled as a first-order Poisson Process given the coherence assumption in time-dependent political news streams.

$$m_{i,t} \sim Poisson(m_{i,t}, m_{i,t-1}) \quad (1)$$

For each document $d \in G_{i,t}$, m_d is sampled from a Poisson distribution with mean value of $m_{i,t}$.

$$m_d \sim Poisson(m_d, m_{i,t}) \quad (2)$$

For sentence $S \in d_i$, m_S is sampled from m_d from a Poisson distribution based on m_d.

$$m_S \sim Poisson(m_S, m_d) \quad (3)$$

4.4 Intialization

Given a labeled subjective list M, for article $d \in G_i$, we initialize T_i^d with the name of entity C_i, d_i with sentences satisfying $T_S = C_i$ and $E_S \in M$. m_S

$m_d = 7$
brotherhood

(1) A is great
(2) B is great
(3) B is heroic
(4) C is evil

$M = \{(great,7),(evil,1)\}$
$T_i^d = \{A\}$
$d_i = \{(1)\}$

(a)

$m_d = 7$
brotherhood

(1) A is great
(2) B is great
(3) B is heroic
(4) C is evil

$M = \{(great,7),(evil,1)\}$
$T_i^d = \{A,B\}$
$d_i = \{(1),(2)\}$

(b)

$m_d = 7$
brotherhood

(1) A is great
(2) B is great
(3) B is heroic
(4) C is evil

$M = \{(great,7),(evil,1),(heroic,7)\}$
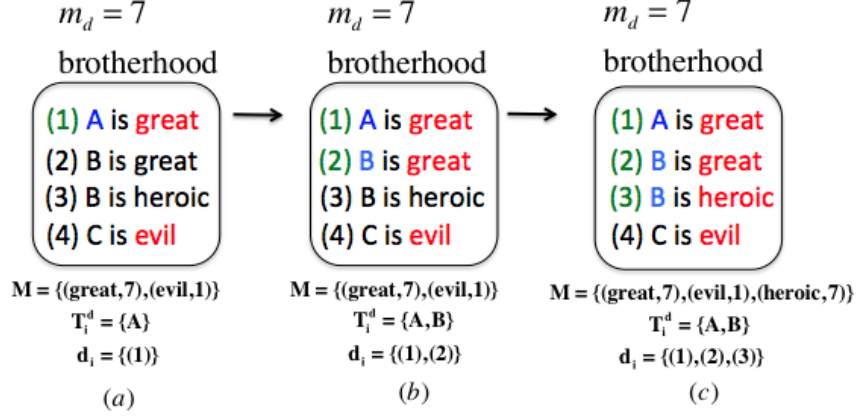$T_i^d = \{A,B\}$
$d_i = \{(1),(2),(3)\}$

(c)

Figure 2: A brief demonstration of the adopted semi-supervised algorithm. (a)→(b): Sentence (2) is added to d_i due to the presence of already known subjective term "great" . Target B is added to target list T_i^d. (b)→(c): term "heroic" is added to subjective word list M with score 7 since it modifies target B.

for $S \in d_i$, is initialized as the average score of its containing expression E_s based on M. Then the MCMC algorithm is applied by iteratively updating m_d and $m_{i,t}$ according to the posterior distribution. Let $P(m|\cdot)$ denotes the probability of parameter m given all other parameters and the posterior distributions are given by:

$$P(m_d = \lambda|\cdot) \propto Poisson(\lambda, m_{i,t}) \prod_{S \in d_i} Poisson(\lambda, m_S)$$

$$P(m_{i,t} = \lambda|\cdot) \propto Poisson(\lambda, m_{i,t-1})$$
$$\times Poisson(m_{i,t+1}, \lambda) \cdot \times \prod_{d \in G_{i,t}} Poisson(m_d, \lambda)$$

(4)

4.5 Semi-supervised Bootstrapping

Our semi-supervised learning algorithm updates M, T_i^d, d_i, S_d and S_i^d iteratively. A brief interpretation is shown in Figure 2 and the details are shown in Figure 4. Concretely, for each sentence $S \in d - d_i$, **step 1** means, if its expression E_S exists in subjective list M, we added its target T_S to T_i^d and S to d_i. **step 2** means if the target T_S exists in T_i^d, its expression, E_s, is added to subjective list M with score m_d. As M and T_i^d change in the iteration, in **step 3**, we again go over all unconsidered sentences with new M and T_i^d. m_d and $m_{i,t}$ are then updated based on new m_S using MCMC in Equ. 4. Note that sentences with pronoun target are not involved in the bootstrapping procedure.

Input: Entity C_i, G_i, subjective term list M
- for each entity i, each document d
  $T_i^d = \{C_i\}$, $d_i = \{S|S \in d, C_i = T_S, E_s \in M\}$
  for each sentence $S \in d_i$:
  . $m_s = \frac{1}{|E_S \in M|} \sum m_{E_s}$
- Iteratively update $m_{i,t}$, m_d using MCMC based on posterior probability shown in Equ.4.

Output: $\{d_i\}$, $\{T_i^d\}$, $\{m_{i,t}\}$ and $\{m_d\}$

Figure 3: Initialization Algorithm.

4.6 Error Prevention in Bootstrapping

Error propagation is highly influential and damaging in bootstrapping algorithms, especially when extending very limited data to huge corpora. To avoid the collapse of the algorithm, we select candidates for opinion analysis in a extremely strict manner, at the sacrifice of many subjective sentences[15]. Concretely, we only consider sentences with exactly one target and at least one expression. Sentences with multiple targets (e.g., Example 2 in Section 4.1) or no expressions, or no targets are discarded.

In addition to the strict sentence selection approach, we adopt the following methods for self-correction in the boot-strapping procedure:

1. For $T_1, T_2 \in T_i^d$, $(E_1, T_1) \in S_1$, $(E_2, T_2) \in S_2$, $E_1, E_2 \in M$, if $|m_{E_1} - m_{E_2}| > 1$: Expel E_1 and E_2 from M, expel T_1 and T_2 from T_i^d,

---

[15]Negative effect of strict sentence selection can be partly compensated by the consideration of time-level information

Input: Entity $\{C_i\}$, Articles Collections $\{G_i\}$, subjective term list M, sentiment score $\{m_d\}$, $\{m_{i,t}\}$, target list for each document $\{T_i^d\}$

Algorithm:

while not convergence:

● for each entity C_i, document d:

   for each sentence $S \in d - d_i$

      1. if $E_S \in M, T_s \notin T_i^d$

         $T_i^d = T_i^d \bigcup T_s, d_i = d_i \bigcup S, m_S = m_d$

      2. if $T_s \in T_i^t, E_s \notin M$

         $M = M \bigcup (E_s, S_d), d_i = d_i \bigcup S, m_s = m_d$

      3. if $E_S \in M, T_S \in T_i^d$

         $d_i = d_i \bigcup S, m_S = m_{E_s}$

● Iteratively update $m_{i,t}$, m_d using MCMC based on posterior probability shown in Equ.4 .

end while:

Output: subjective term list M, score $\{m_{i,t}\}$

Figure 4: Semi-supervised learning algorithm.

with the exception of original labeled data.

Explanation: If sentiment scores for two expressions, whose correspondent targets both belong to the target list T_i^d, diverge enough, we discard both expressions and targets based according to Assumption 1: sentiments towards one entity (or its allies) in an article should be consistent.

2. $\exists S \in d, T_S \in T_i^d, |m_{E_S} - m_d| > 1, T_S$ is expelled from T_i^d.

Explanation: If target T_S for sentence S belongs to T_i^d, but its corresponding expression E_s is not consistent with article-level sentiment m_d, T_S is expelled from T_i^d.

5 Experiment

5.1 Data and Preprocessing

Our data set is composed of the People's daily from 1950 to 2010, across a 60-year time span. News articles are first segmented using ICTCLAS Chinese segmentation word system[16] (Zhang et al., 2003). Articles with fewer than 200 Chinese words are discarded. News articles are clustered by the presence of a country's name more than 2 times based on a country name list from Wikipedia[17]. Articles men-

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)

Table 1: Illustration of subjective list M

tioning more than 5 different countries are discarded since they usually talk about international conferences. Note that one article can appear in different collections (example in Section 1 will appear in both Vietnam and the U.S. collection).

Compound sentences are segmented into clauses based on dependency parse tree. Then those containing more than 50 characters or less than 4 characters are discarded. To avoid complicated inference, sentences with negation indicators are discarded.

5.2 Obtaining Subjectivity Word List

Since there are few Chinese subjectivity lexicons (with degrees) available and those exist may not serve our specific purpose, we manually label a small number of Chinese subjective terms as seed corpus. We divided the labeling process into 2 steps rather than directly labeling vocabularies[18]. We first selected 100 news articles and assigned each of them (as well as the appropriate country entity C_i) to 2 students majoring in *International Studies*, asking them to give a label sentiment score (1 to 7) according to the rules described in Section 4.2. 20 students participated in the procedure. Since annotators have plenty of background knowledge, they agreed on 98 out of 100. Second, we selected out subjectivity lexicons by matching to a comprehensive subjectivity lexicons list[19]. and ask 2 students select the candidates that signal the document-level label from the first step. According to whether a word a selected or not, the value of $Cohen's \kappa$ is 0.78, showing substantial agreement. For the small amount of labels on which the judges disagree, we recruited an extra judge and to serve as a tie breaker. Table 1 shows some labeled examples.

	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

Table 2: Results for Expressions/Targets extraction.

5.3 Targets and Expressions Extraction

As the good performance of *semi-CRF* in opinion extraction has been demonstrated in previous work (Yang and Cardie, 2012), we briefly go over model evaluation in this subsection for brevity. We manually labeled 600 sentences and performed 5-fold cross validation for evaluation. We compare *semi-CRF* to *Standard CRF*. We report performances on two settings in Table 2. The first setting, *Total*, corresponds to performance on the whole dataset, while second one *Single*, denotes the performance on the set of sentences with only one target, which we are more interested in because multiple-target sentences are discarded in our algorithm. It turned out that *semi-CRF* significantly outputs standard CRF, approaching 0.90 F-1 score on *Single* setting.

5.4 Foreign Relation Evaluation

Gold-standard foreign relations are taken from Political Science research at the Institute of Modern International Relations, Tsinghua University, extracted from monthly quantitative China foreign relations reports with 7 countries (U.S., Japan, Russia/Soviet, England, France, India, and Germany) from 1950 to 2012[20].

We consider several baselines. For fair comparison, we use identical processing techniques for each approach. Some baselines make article-level predictions, for which we obtain time-period level relation prediction by averaging the documents.

Coreference+Bootstrap (CB): We first implemented Ngai and Wang's Chinese coreference system (2007). We then bootstrap sentiment terms and score based on entity coreference.

No-time: A simplified version of our approach

Model	Ours	CB	No-time
Pearson	0.895	0.753	0.808
Model	SVR-d	SLDA	SVR-S
Pearson	0.482	0.427	0.688

Table 3: Pearson Correlation with Gold Standard.

where each article is considered as an independent unit and no time-level information is considered. m_d is obtained by averaging its containing sentences and used for later bootstrapping.

SVR-d: Uses SVM^{light}(Joachims, 1999) to train a linear SVR (Pang and Lee, 2008) for document-level sentiment prediction using the unigram feature. The 100 labeled documents are used as training data.

SLDA: supervised-LDA (Blei and McAuliffe, 2010) for document-level label prediction. Topic number is set to 10, 20, 50, 100 respectively and we report the best result.

SVR-S: Sentence-level SVR to sentences with presence of entity C_i[21]. We obtain document-level prediction by averaging its containing sentences and then time-period level prediction by averaging its containing documents.

We report the Pearson Correlation with gold standards in table 3. As we can observe, simple document-level regression models, i.e., **SVR** and **SLDA** do not fit this task. The reason is simple: one article d can appear in different collections. Recall the Vietnam example in Section 1, it appears in both $G_{Vietnam}$ and $G_{the\ U.S.}$. Sentiment prediction for d should be totally opposite in the two document collections: very positive in $G_{Vietnam}$ and very negative in G_{USA}. But document level prediction would treat them equally. Our approach outperforms **No-Time**, illustrating the meaningfulness of exploiting time-level information in our task. Our system approaches around 0.9 correlation with the gold standards. The reason why **No-Time** is better than **CB** is also simple: **CB** includes only coreferent entities in the target list (e.g., America for the USA article collection), and therefore overlooks rich information provided by non-coreferent entities (e.g., President Nixon or Nixon Government). **No-Time**

---

[20]Details found here http://www.imir.tsinghua.edu.cn/publish/iisen/7523/index.html.

[21]Features we explore include word entities in current sentence, POS, a window of $k \in \{1, 2\}$ words from the target and the expression and corresponding POS, and the dependency path between target and expression.

instead groups entities according to attitude, thereby enabling more information to be harnessed. For **SVR-S**, as the regression model trained from limited labeled data can hardly cover unseen terms during testing, the performance is just OK. **SVR-S** also suffers from overlooking rich sources of information since it only considers sentences with exact mention of the name entity of the corresponding country.
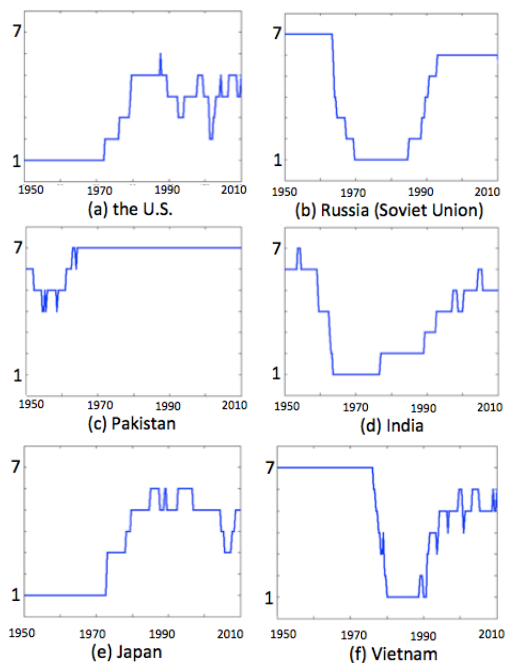

Figure 5: Examples of China's Foreign Relations.

6 Diplomatic Relations

"The enemy of my enemy is my friend"
—Arabic proverb

A central characteristic of post-World War Second international system with which China had to deal would be overwhelming preeminence of the USA and USSR as each of the superpowers stood at the center of a broad alliance system who was engaged in an intense and protracted global conflict with the other. We choose 6 countries and report results in Figure 5. One of interesting things we can observe from Figure 5 is that foreign attitudes are usually divergent towards two opposing forces: Sino-American relation (see Figure 5(a)) began to improve when the Sino-Soviet relation (see Figure 5(b)) reached its bottom at the beginning of 1970s. Similar patterns appear for Sino-Pakistan

(see Figure 5(c)), Sino-India relations (see Figure 5(d)) in early 1960s[22], and Sino-Vietnam 5(f)), Sino-American relations in late 1970s. On the contrast, attitudes are usually consistent toward allied forces: Sino-Japan relations with Sino-USA relations before 1990s, and Sino-Vietnam relations with Sino-Soviet relations in late 1970s and 1980s.

Figure 6 presents top clustering target (T_i^d) in the USA and Soviet Union/Russia article collection. As some of vocabulary terms can be both target and expression, we use blue to label terms with positive sentiment, red to label negative ones. As we can see from Figure 6, targets(T) extracted by our model show a very clear pattern where allies and co-referent entities are grouped. Another interesting thing is, the subjectivity of target words from different times is generally in accord with the relation curves shown in Figure 5.

7 Conclusion and Discussion

In this paper, we propose a sentiment analysis algorithm to track China's foreign relations from the People's Daily. Our semi-supervised algorithm harnesses higher level information (i.e., document-level, time-level) by incorporating a hierarchical Bayesian approach into the framework, to resolve sentiment target clustering, create subjective lexicons, and perform sentiment prediction simultaneously. While we focus here on the People's Daily for diplomatic relation extraction, the idea of our approach is general and can be extended broadly. Another contribution of this work is the creation a comprehensive Chinese subjective lexicon list. We are hopeful that our approach can not only facilitate quantitative research by political scientists, but also shed light on NLP applications such as coreference and metaphor, where sentiment clues can be helpful.

It is worth noting that, while harnessing time-level information can indeed facilitate opinion analysis, especially when labeled data is limited in our specific task, it is not a permanent-perfect assumption, especially considering the diversity and treacherous currents at the international political stage.

At algorithm-level, to avoid error propagation due to limitations of current sentiment analysis tools

---

[22] A fan of history can trace the crucial influence of the USSR in Sino-India relation in 1960s

Figure 6: Top coreference terms Towards USA and Soviet Union/Russia versus time. Blue denotes words that are both Target and positive words in M. Red denotes words that are both Target and negative words in M

(even though semi-CRF produces state-of-art performance in target and expression extraction task, a performance of 0.8 F-value, when applied to the whole corpus, can by no means satisfy our requirements), we discard a great number of sentences, among which is contained much useful information. How to resolve these problems and improve opinion extraction performance is our long-term goal in sentiment analysis/opinion extraction literature.

References

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*.

Galen Andrew. 2006. A hybrid markov/semi-markov conditional random field for sequence segmentation. In *EMNLP*.

David M Blei and Jon D McAuliffe. 2010. Supervised topic models. *arXiv preprint arXiv:1003.0783*.

Eric Breck, Yejin Choi, and Claire Cardie. 2007. Identifying expressions of opinion in context. In *IJCAI*.

Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Identifying sources of opinions with conditional random fields and extraction patterns. In *EMNLP*.

Yejin Choi, Eric Breck, and Claire Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *EMNLP*.

Tang Duyu, Qin Bing, Zhou LanJun, Wong KamFai, Zhao Yanyan, and Liu Ting. 2013. Domain-specific sentiment word extraction by seed expansion and pattern generation. *arXiv preprint arXiv:1309.6722*.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*.

Wei Jin, Hung Hay Ho, and Rohini K Srihari. 2009. A novel lexicalized hmm-based learning framework for web opinion mining. In *ICML*.

Yohan Jo and Alice H Oh. 2011. Aspect and sentiment unification model for online review analysis. In *ICWSM*.

Thorsten Joachims. 1999. Making large scale svm learning practical.

Richard Johansson and Alessandro Moschitti. 2010. Syntactic and semantic structure for opinion expression detection. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics.

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*.

John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Yun-zhong LIU, Ya-ping LIN, and Zhi-ping CHEN. 2004. Text information extraction based on hidden markov model [j]. *Acta Simulata Systematica Sinica*.

Tie-Yan Liu. 2009. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*.

Grace Ngai and Chi-Shing Wang. 2007. A knowledge-based approach for unsupervised chinese coreference

resolution. *Computational Linguistics and Chinese Language Processing*, 12(4):459–484.

Daisuke Okanohara, Yusuke Miyao, Yoshimasa Tsuruoka, and Jun'ichi Tsujii. 2006. Improving the scalability of semi-markov conditional random fields for named entity recognition. In *ACL*.

Brendan O'Connor, Brandon M Stewart, and Noah A Smith. 2013. Learning to extract international relations from political context. In *ACL*.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP*.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009. Expanding domain sentiment lexicon through double propagation. In *IJCAI*.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*.

Thomas W Robinson and David L Shambaugh. 1995. *Chinese foreign policy: theory and practice*. Oxford University Press.

Sunita Sarawagi and William W Cohen. 2004. Semi-markov conditional random fields for information extraction. In *NIPS*.

Veselin Stoyanov and Claire Cardie. 2008. Topic identification for fine-grained opinion analysis. In *Proceedings of the 22nd International Conference on Computational Linguistics*.

Hongning Wang, Yue Lu, and Chengxiang Zhai. 2010. Latent aspect rating analysis on review text data: a rating regression approach. In *SIGKDD*.

Guoguang Wu. 1994. Command communication: The politics of editorial formulation in the people's daily. *China Quarterly*, 137:194–211.

Bishan Yang and Claire Cardie. 2012. Extracting opinion expressions with semi-markov conditional random fields. In *EMNLP*.

Bishan Yang and Claire Cardie. 2014. Context-aware learning for sentence-level sentiment analysis with posterior regularization. *ACL*.

Hua-Ping Zhang, Hong-Kui Yu, De-Yi Xiong, and Qun Liu. 2003. Hhmm-based chinese lexical analyzer ictclas. In *Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17*.

Lei Zhang, Bing Liu, Suk Hwan Lim, and Eamonn O'Brien-Strain. 2010. Extracting and ranking product features in opinion documents. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*.