

Structure-Aware Review Mining and Summarization

Fangtao Li¹, Chao Han¹, Minlie Huang¹, Xiaoyan Zhu¹,
Ying-Ju Xia², Shu Zhang² and Hao Yu²

¹State Key Laboratory of Intelligent Technology and Systems,

¹Tsinghua National Laboratory for Information Science and Technology,

¹Department of Computer Science and Technology, Tsinghua University

²Fujitsu Research and Development Center

fangtao06@gmail.com; zxy_dcs@tsinghua.edu.cn

Abstract

In this paper, we focus on object feature¹ based review summarization. Different from most of previous work with linguistic rules or statistical methods, we formulate the review mining task as a joint structure tagging problem. We propose a new machine learning framework based on Conditional Random Fields (CRFs). It can employ rich features to jointly extract positive opinions, negative opinions and object features for review sentences. The linguistic structure can be naturally integrated into model representation. Besides linear-chain structure, we also investigate conjunction structure and syntactic tree structure in this framework. Through extensive experiments on movie review and product review data sets, we show that structure-aware models outperform many state-of-the-art approaches to review mining.

1 Introduction

With the rapid expansion of e-commerce, people are more likely to express their opinions and hands-on experiences on products or services they have purchased. These reviews are important for both business organizations and personal costumers. Companies can decide on their strategies for marketing and products improvement. Customers can make a better decision when pur-

¹ Note that there are two meanings for word “feature”. We use “object feature” to represent the target entity, which the opinion expressed on, and use “feature” as the input for machine learning methods.

chasing products or services. Unfortunately, reading through all customer reviews is difficult, especially for popular items, the number of reviews can be up to hundreds or even thousands. Therefore, it is necessary to provide coherent and concise summaries for these reviews.

Gone With The Wind: Movie: <i>Positive:</i> great, good, amazing, ... , breathtaking <i>Negative:</i> bad, boring, waste time, ... , mistake Actor: <i>Positive:</i> charming , brilliant , great, ... , smart <i>Negative:</i> poor, fail, dirty, ... , lame Music: <i>Positive:</i> great, beautiful, very good, ... , top <i>Negative:</i> annoying, noise, too long, ... , unnecessary

Figure 1. Feature based Review Summarization

Inspired by previous work (Hu and Liu, 2004; Jin and Ho, 2009), we aim to provide object feature based review summarization. Figure 1 shows a summary example for movie “Gone with the wind”. The object (movie) features, such as “movie”, “actor”, with their corresponding positive opinions and negative opinions, are listed in a structured way. The opinions are ranked by their frequencies. This provides a concise view for reviews. To accomplish this goal, we need to do three tasks: 1), extract all the object features and opinions; 2), determine the sentiment polarities for opinions; 3), for each object feature, determine the relevant opinions, i.e. object feature-opinion pairs.

For the first two tasks, most previous studies employ linguistic rules or statistical methods (Hu and Liu, 2004; Popescu and Etzioni 2005). They mainly use unsupervised learning methods, which lack an effective way to address infrequent object features and opinions. They are also hard to incorporate rich overlapping features.

Actually, there are many useful features, which have not been fully exploited for review mining. Meanwhile, most of previous methods extract object features, opinions, and determine the polarities for opinions separately. In fact, the object features, positive opinions and negative opinions correlate with each other.

In this paper, we formulate the first two tasks, i.e. object feature, opinion extraction and opinion polarity detection, as a joint structure tagging problem, and propose a new machine learning framework based on Conditional Random Fields (CRFs). For each sentence in reviews, we employ CRFs to jointly extract object features, positive opinions and negative opinions, which appear in the review sentence. This framework can naturally encode the linguistic structure. Besides the neighbor context with linear-chain CRFs, we propose to use Skip-chain CRFs and Tree CRFs to utilize the conjunction structure and syntactic tree structure. We also propose a new unified model, Skip-Tree CRFs to integrate these structures. Here, “structure-aware” refers to the output structure, which model the relationship among output labels. This is significantly different from the previous input structure methods, which consider the linguistic structure as heuristic rules (Ding and Liu, 2007) or input features for classification (Wilson et al. 2009). Our proposed framework has the following advantages: First, it can employ rich features for review mining. We will analyze the effect of features for review mining in this framework. Second, the framework can utilize the relationship among object features, positive opinions and negative opinions. It jointly extracts these three types of expressions in a unified way. Third, the linguistic structure information can be naturally integrated into model representation, which provides more semantic dependency for output labels. Through extensive experiments on movie review and product review, we show our proposed framework is effective for review mining.

The rest of this paper is organized as follows: In Section 2, we review related work. We describe our structure aware review mining methods in Section 3. Section 4 demonstrates the process of summary generation. In Section 5, we present and discuss the experiment results. Section 6 is the conclusion and future work.

2 Related Work

Object feature based review summary has been studied in several papers. Zhuang et al. (2006) summarized movie reviews by extracting object feature keywords and opinion keywords. Object feature-opinion pairs were identified by using a dependency grammar graph. However, it used a manually annotated list of keywords to recognize movie features and opinions, and thus the system capability is limited. Hu and Liu (2004) proposed a statistical approach to capture object features using association rules. They only considered adjective as opinions, and the polarities of opinions are recognized with WordNet expansion to manually selected opinion seeds. Popescu and Etzioni (2005) proposed a relaxation labeling approach to utilize linguistic rules for opinion polarity detection. However, most of these studies focus on unsupervised methods, which are hard to integrate various features. Some studies (Breck et al. 2007; Wilson et al, 2009; Kobayashi et al. 2007) have used classification based methods to integrate various features. But these methods separately extract object features and opinions, which ignore the correlation among output labels, i.e. object features and opinions. Qiu et al. (2009) exploit the relations of opinions and object features by adding some linguistic rules. However, they didn’t care the opinion polarity. Our framework can not only employ various features, but also exploit the correlations among the three types of expressions, i.e. object features, positive opinions, and negative opinions, in a unified framework. Recently, Jin and Ho (2009) propose to use Lexicalized HMM for review mining. Lexicalized HMM is a variant of HMM. It is a generative model, which is hard to integrate rich, overlapping features. It may encounter sparse data problem, especially when simultaneously integrating multiple features. Our framework is based on Conditional Random Fields (CRFs). CRFs is a discriminative model, which can easily integrate various features.

These are some studies on opinion mining with Conditional Random Fields. For example, with CRFs, Zhao et al (2008) and McDonald et al. (2007) performed sentiment classification in sentence and document level; Breck et al (2007) identified opinion expressions from newswire documents; Choi et al. (2005) determined opi-

nion holders to opinions also from newswire data. None of previous work focuses on jointly extracting object features, positive opinions and negative opinions simultaneously from review data. More importantly, we also show how to encode the linguistic structure, such as conjunction structure and syntactic tree structure, into model representation in our framework. This is significantly different from most of previous studies, which consider the structure information as heuristic rules (Hu and Liu, 2004) or input features (Wilson et al. 2009).

Recently, there are some studies on joint sentiment/topic extraction (Mei et al. 2007; Titov and McDonald, 2008; Snyder and Barzilay, 2007). These methods represent reviews as several coarse-grained topics, which can be considered as clusters of object features. They are hard to indentify the low-frequency object features and opinions. While in this paper, we will extract all the present object features and corresponding opinions with their polarities. Besides, the joint sentiment/topic methods are mainly based on review document for topic extraction. In our framework, we focus on sentence-level review extraction.

3 Structure Aware Review Mining

3.1 Problem Definition

To produce review summaries, we need to first finish two tasks: identifying object features, opinions, and determining the polarities for opinions. In this paper, we formulate these two tasks as a joint structure tagging problem. We first describe some related definitions:

Definition (Object Feature): is defined as whole target expression that the subjective expressions have been commented on. Object features can be products, services or their elements and properties, such as “character”, “movie”, “director” for movie review, and “battery”, “battery life”, “memory card” for product review.

Definition (Review Opinion): is defined as the whole subjective expression on object features. For example, in sentence “The camera is easy to use”, “easy to use” is a review opinion. “opinion” is used for short.

Definition (Opinion Polarity): is defined as the sentiment category for review opinion. In this paper, we consider two types of polarities: posi-

tive opinion and negative opinion. For example, “easy to use” belongs to positive opinion.

For our review mining task, we need to represent three types of expressions: object features, positive opinions, and negative opinions. These expressions may be words, or whole phrases. We use BIO encoding for tag representation, where the non-opinion and neutral opinion words are represented as “O”. With Negation (N), which is only one word, such as “not”, “don’t”, as an independent tag, there are totally 8 tags, as shown in Table 1. The following is an example to denote the tags:

The/O camera/FB comes/O with/O a/O pitiful/CB 32mb/FB compact/FI flash/FI card/FI ./O

FB	Feature Beginning	CB	Negative Beginning
FI	Feature Inside	CI	Negative Inside
PB	Positive Beginning	N	Negation Word
PI	Positive Inside	O	Other

Table 1. Basic Tag Set for Review Mining

3.2 Structure Aware Model

In this section, we describe how to encode different linguistic structure into model representation based on our CRFs framework.

3.2.1 Using Linear CRFs.

For each sentence in a review, our task is to extract all the object features, positive opinions and negative opinions. This task can be modeled as a classification problem. Traditional classification tools, e.g. Maximum Entropy model (Berger et al, 1996), can be employed, where each word or phrase will be treated as an instance. However, they independently consider each word or phrase, and ignore the dependency relationship among them.

Actually, the context information plays an important role for review mining. For example, given two continuous words with same part of speech, if the previous word is a positive opinion, the next word is more likely a positive opinion. Another example is that if the previous word is an adjective, and it is an opinion, the next noun word is more likely an object feature.

To this end, we formulate the review mining task as a joint structure tagging problem, and propose a general framework based on Conditional Random Fields (CRFs) (Lafferty et al., 2001) which are able to model the dependencies

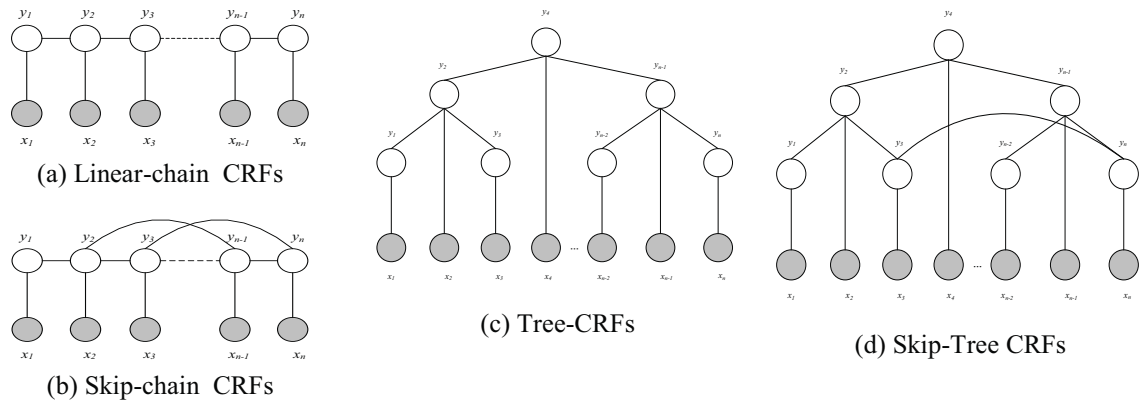


Figure 2 CRFs models

between nodes. (See Section 3.2.5 for more about CRFs)

In this section, we propose to use linear-chain CRFs to model the sequential dependencies between continuous words, as discussed above. It views each word in the sentence as a node, and adjacent nodes are connected by an edge. The graphical representation is shown in Figure 2(a). Linear CRFs can make use of dependency relationship among adjacent words.

3.2.2 Leveraging Conjunction Structure

We observe that the conjunctions play important roles on review mining: If the words or phrases are connected by conjunction “and”, they mostly belong to the same opinion polarity. If the words or phrases are connected by conjunction “but”, they mostly belong to different opinion polarity, as reported in (Hatzivassiloglou and McKeown, 1997; Ding and Liu, 2007). For example, “This phone has a very cool and useful feature – the speakerphone”, if we only detect “cool”, it is hard to determine its opinion polarity. But if we see “cool” is connected with “useful” by conjunction “and”, we can easily acquire the polarity of “cool” as positive. This conjunction structure not only helps to determine the opinions, but also helps to recognize object features. For example, “I like the special effects and music in this movie”, with word “music” and conjunction “and”, we can easily detect that “special effects” as an object feature.

To model the long distance dependency with conjunctions, we use Skip-chain CRFs model to detect object features and opinions. The graphical representation of a Skip-chain CRFs, given in Figure 2(b), consists of two types of edges: li-

near-edge (y_{t-1} to y_t) and skip-edge (y_i to y_j). The linear-edge is described as linear CRFs. The skip-edge is imported as follows:

We first identify the conjunctions in the review sentence, with a collected conjunction set, including “and”, “but”, “or”, “however”, “although” etc. For each conjunction, we extract its connected two text sequences. The nearest two words with same part of speech from the two text sequences are connected with the skip-edge. Here, we just consider the noun, adjective, and adverb. For example, in “good pictures and beautiful music”, there are two skip-edges: one connects two adjective words “good” and “beautiful”; the other connects two nouns “pictures” and “music”. We also employ the general sentiment lexicons, SentiWordNet (Esuli and Sebastiani, 2006), to connect opinions. Two nearest opinion words, detected by sentiment lexicon, from two sequences, will also be connected by skip-edge. If the nearest distance exceeds the threshold, this skip edge will be discarded. Here, we consider the threshold as nine.

Skip-chain CRFs improve the performance of review mining, because it naturally encodes the conjunction structure into model representation with skip-edges.

3.2.3 Leveraging Syntactic Tree Structure

Besides the conjunction structure, the syntactic tree structure also helps for review mining. The tree denotes the syntactic relationship among words. In a syntactic dependency representation, each node is a surface word. For example, the corresponding dependency tree (Klein and Manning, 2003) for the sentence, “I really like this long movie”, is shown in Figure 3.

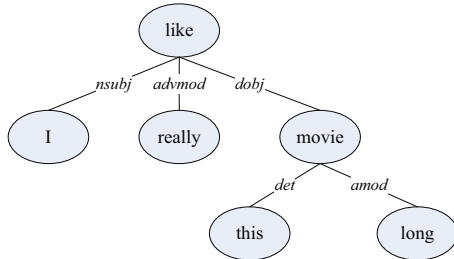


Figure 3. Syntactic Dependency Tree Representation

In linear-chain structure and skip-chain structure, “like” and “movie” have no direct edge, but in syntactic tree, “movie” is directly connected with “like”, and their relationship “dobj” is also included, which shows “movie” is an objective of “like”. It can provide deeper syntactic dependencies for object features, positive opinions and negative opinions. Therefore, it is important to consider the syntactic structure in the review mining task.

In this section, we propose to use Tree CRFs to model the syntactic tree structure for review mining. The representation of a Tree CRFs is shown in Figure 2(c). The syntactic tree structure is encoded into our model representation. Each node is corresponding to a word in the dependency tree. The edge is corresponding to dependency tree edge. Tree CRFs can make use of dependency relationship in syntactic tree structure to boost the performance.

3.2.4 Integrating Conjunction Structure and Syntactic Tree Structure

Conjunction structure provides the semantic relations correlated with conjunctions. Syntactic tree structure provides dependency relation in the syntactic tree. They represent different semantic dependencies. It is interesting to consider these two dependencies in a unified model. We propose Skip-Tree CRFs, to combine these two structure information. The graphical representation of a Skip-Tree CRFs, given in Figure 2(d), consists of two types of edges: tree edges and conjunction skip-edges. We hope to simultaneously model the dependency in conjunction structure and syntactic tree structure.

We also notice that there is a relationship “conj” in syntactic dependency tree. However, we find that it only connects two head words for a few coordinating conjunction, such as “and”, “or”, “but”. Our designed conjunction skip-edge provides more information for joint structure tagging. We analyze more conjunctions to con-

nect not only two head words, but also the words with same part of speech. We also connect the words with sentiment lexicon. We will show that the skip-tree CRFs, which combine the two structures, is effective in the experiment section.

3.2.5 Conditional Random Fields

A CRFs is an undirected graphical model G of the conditional distribution $P(Y|X)$. Y are the random variables over the labels of the nodes that are globally conditioned on X , which are the random variables of the observations. The conditional probability is defined as:

$$P(Y|X) = \frac{1}{Z(X)} \exp \left(\sum_{e \in E, i} \gamma_i t_i(e, Y|e, X) + \sum_{v \in V, i} \mu_i s_i(v, Y|v, X) \right)$$

where $Z(x)$ is the normalization factor, s_i is the state function on node, t_i is the transition functions on edge, and γ_i and μ_i are parameters to estimate (Sutton and McCallum, 2006).

Inference and Parameter Estimation. For Linear CRFs, dynamic programming is used to compute the maximum a posteriori (MAP) of Y given X . For more complicated graphs with cycles, we employ Tree Re-Parameterization (TRP) algorithm (Wainwright et al. 2001) for approximate inference.

Given the training Data $D = \{x^{(i)}, y^{(i)}\}_{i=1}^n$, the parameter estimation is to determine the parameters based on maximizing the log-likelihood $L_\theta = \sum_{i=1}^n \log p(y^{(i)}|x^{(i)})$. In Linear CRFs model, dynamic programming and L-BFGS algorithm can be used to optimize objective function L_θ , while for complicated CRFs, TRP is used instead to calculate the marginal probability.

3.3 Feature Space

In this section, we describe the features used in the learning methods. All the features are listed in Figure 4. Word features include the word’s token, lemma, and part of speech. The adjacent words’ information is considered. We detect whether the negation words appear in the previous four words as a binary feature. We also detect whether this word is the superlative form, such as “best”, and comparative form, such as “better”, as binary features. Two types of dictionaries are employed. We use WordNet to acquire the synonyms and antonyms for each word. SentiWordNet (Esuli and Sebastiani, 2006) is used to acquire the prior polarity for each word. We use the words with positive or negative score

Word Feature:	
	Word token
	Word lemma
	Word part of speech
	Previous word token, lemma, part of speech
	Next word token, lemma, part of speech
	Negation word appears in previous 4 words
	Is superlative degree
	Is comparative degree
Dictionary Feature	
	WordNet Synonym
	WordNet Antonym
	SentiWordNet Prior Polarity
Sentence Feature	
	Num of positive words in SentiWordNet
	Num of negative words in SentiWordNet
	Num of Negation word
Syntactic Features:	
	Parent word
	Parent SentiWordnet Prior Polarity
	In subject
	In copular
	In object
Edge Feature	
	Conjunction word
	Syntactic relationship

Figure 4. Features for learning Methods

above a threshold (0.6). Sentence Feature provides sentence level information. It includes the count of positive words and negative words, which are detected by SentiWordNet. We also incorporate the count of negation words as a feature. There are some syntactic features from dependency tree. Parent word and its polarity are considered. We also detect if the word is subject, object or copular. For edge features, the conjunction words are incorporated as corresponding skip-edge features. The syntactic relationship is considered as a feature for corresponding tree-edge. For classification and linear CRFs models, we just add this edge features as general features.

4 Review Summary Generation

After extracting the object features and opinions, we need to extract the relevant opinions for each feature. In this paper, we identify the nearest opinion word/phrase for each object feature as object feature-opinion pair, which is widely used in previous work (Hu and Liu, 2004; Jin and Ho, 2009). The review summary is generated as a list of structured object feature-opinion pairs, as shown in Figure 1.

5 Experiment

5.1 Experiment setup

Data Set: For our structure tagging task, we need to know the labels for all the words in reviews. In this paper, we manually annotate two types of these review data sets. One is movie review, which contains five movies with totally 500 reviews. The other is product review, which contains four products with totally 601 reviews. We need to label all object features, positive opinions, negative opinions, and the object feature-opinion pairs for all sentences. Each sentence is labeled by two annotators. The conflict is checked by the third person. Finally, we acquire 2207 sentences for movie review and 2533 sentences for product review. For each type, including movie and product, the data set is divided into five parts. We select four parts as training data, and the fifth part as testing data.

Evaluation Metric:

Precision, Recall and F measure are used to test our results, as Jin and Ho (2009).

5.2 Baselines

First word	Second Word	Third Word
JJ	NN or NNS	Anything
RB, RBR or RBS	JJ	NN or NNS
JJ	JJ	NN or NNS
NN or NNS	JJ	Not NN or NNS

Table 2. Rules in rule based method

Rule based Method:

The rule based method is used in Jin and Ho (2009), which is motivated by (Hu and Liu, 2004; Turney, 2002). The employed rules are shown in Table 2. The matching adjective is identified as opinion, and matching nouns are extracted as object features. To determine the polarities of the opinions, 25 positive adjectives and 25 negative adjectives are used as seeds, and then expanded by searching synonyms and antonyms in WordNet. The polarity of a word is detected by checking the collected lists.

Lexicon based Method:

The object features and opinions extraction is same as rule based method. The general sentiment lexicon SentiWordNet is employed to detect the polarity for each word.

Lexicalized HMM:

The object features and opinions are identified by Lexicalized HMM (L-HMM), as Jin and Ho (2009). L-HMM is a variant of HMM. It has two observations. The current tag is not only related

	Methods	Object Features			Positive Opinions			Negative Opinions			Overall		
		P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
Movie Review	Rule	41.2	32.3	36.2	82.9	31.1	45.3	23.5	13.7	17.3	49.2	25.7	33.8
	Lexicon	41.2	32.3	36.2	64.0	38.1	47.8	19.6	6.8	10.2	41.6	25.8	31.8
	L-HMM	88.0	52.6	65.9	82.1	49.6	61.9	65.9	41.1	50.6	78.7	47.8	59.5
	MaxEnt	83.4	75.1	79.1	82.2	65.0	72.6	74.1	29.5	42.2	79.9	56.5	66.2
	Linear CRFs	81.8	78.4	80.1	79.1	63.9	70.7	75.8	32.2	45.2	79.0	58.2	67.0
Product Review	Rule	53.5	35.6	42.8	74.4	22.5	34.6	17.1	8.9	11.7	48.3	22.3	30.6
	Lexicon	53.5	35.6	42.8	48.9	29.7	40.0	14.7	3.7	5.9	39.1	23.0	29.0
	L-HMM	83.9	48.7	61.6	90.3	56.8	69.8	47.2	25.2	32.9	73.8	43.6	54.8
	MaxEnt	83.4	55.1	66.4	82.2	65.0	72.6	64.1	30.0	40.4	76.6	49.9	60.4
	Linear CRFs	91.1	56.3	69.6	88.7	70.4	78.5	67.7	32.6	44.0	82.5	53.1	64.6

Table 3. Comparison Results with Baselines

(the learning methods only employ word token and part of speech as features).

	Methods	Object Features			Positive Opinions			Negative Opinions			Overall		
		P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
Movie Review	MaxEnt	82.8	76.6	79.6	80.3	67.8	73.5	82.8	36.3	50.5	81.9	60.2	69.4
	Linear CRFs	83.5	75.4	79.2	77.8	71.4	74.5	70.9	53.4	60.9	77.4	66.8	71.7
	Skip CRFs	83.9	78.7	81.2	81.8	73.4	77.4	75.2	62.3	68.2	80.3	71.5	75.7
	Tree CRFs	84.1	79.0	81.5	82.7	75.4	78.9	76.7	61.0	67.9	81.2	72.2	76.2
	SkipTreeCRFs	85.5	82.0	83.7	82.3	80.0	81.1	80.2	66.4	72.7	82.6	76.2	79.3
Product Review	MaxEnt	80.0	70.8	75.1	85.6	65.7	74.3	65.1	37.8	47.8	76.9	58.1	66.2
	Linear CRFs	84.0	72.9	78.1	86.7	72.0	78.6	60.4	49.6	54.5	77.0	64.8	70.4
	Skip CRFs	84.8	73.5	78.7	87.8	74.5	80.6	73.1	50.4	59.6	81.2	66.1	73.2
	Tree CRFs	83.0	72.7	77.5	86.6	73.4	79.4	64.3	54.8	59.2	78.0	67.0	72.1
	SkipTreeCRFs	87.1	74.1	80.1	91.8	76.7	83.6	81.1	57.0	67.0	86.6	69.3	77.0

Table 4. Comparative experiments with all features

with the previous tag, but also correlates with previous observations. They use word token and part of speech as two features.

Classification based Method:

We also formulate the review mining as a classification task. Each word is considered as an instance. Maximum Entropy (MaxEnt) is used in this paper.

5.3 Experiment results

Since Lexicalized HMM employ word token and part of speech as features (Jin and Ho, 2009), we first conduct comparative experiments with these two features for learning methods. Table 3 shows the results. The rule based method is a little better than lexicon based method. Senti-WordNet is designed for general opinion mining, which may be not suitable for domain specific review mining task. For rule based method, the seeds are selected in the review domain, which is more suitable for domain specific task. However, both methods achieve low performance. This because that they only employ simple linguistic rules to extract object features and opinions, which is not effective for infrequent cases and phrase cases. Lexicalized HMM is an extension

of HMM. It uses word token and part of speech as two observations. The current tag is not only related with the previous tag, but also correlates with previous two observations. Lexicalized HMM can employ dependency relationship among adjacent words. However, it doesn't achieve the expected result. This is because that Lexicalized HMM is a generative model, which is hard to incorporate rich overlapping features. Even Lexicalized HMM uses linear interpolation smoothing technique. The data sparsity problem seriously hurt the performance. There are many sentences with zero probability. MaxEnt classifier is a discriminative model, which can incorporate various features. However, it independently classifies each word, and ignores the dependency among successive words. The linear CRFs model achieves best performances for movie review, and product review in overall F-score. This is because that, in our joint structure tagging framework, linear CRFs can employ the global structure to make use of the adjacent dependency relation, and easily incorporate various features to boost the performance.

We also conduct the comparative experiments with all features. From Table 4, we can see that linear CRFs, which consider the chain structure,

	Object Features			Positive Opinions			Negative Opinions			Overall		
	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
Basic	83.8	79.2	81.4	79.5	71.0	75.0	76.1	37.0	49.8	79.8	62.4	70.0
Basic +Word Feature	84.0	81.4	82.7	79.2	75.6	77.4	78.9	48.6	60.2	80.7	68.6	74.1
Basic +Dictionary	80.5	76.6	78.5	82.7	76.3	79.4	76.5	60.3	67.4	80.0	71.0	75.2
Basic +Sentence	82.5	75.6	78.9	80.4	75.4	77.8	84.0	46.7	60.0	82.3	65.9	73.2
Basic +Syntactic	84.5	70.8	77.0	79.6	73.9	76.7	79.5	47.9	59.8	81.2	64.2	71.7
Basic + Edge	84.1	80.1	82.1	79.5	75.4	77.4	82.4	47.9	60.6	82.0	67.8	74.2
All Features	85.5	82.0	83.7	82.3	80.0	81.1	80.2	66.4	72.7	82.6	76.2	79.3

Table 5. Feature Evaluations with Skip Tree CRFs (movie)

still achieve better results than MaxEnt classifier method. Skip-chain CRFs model the conjunction structure in the sentence. We can see that the Skip-chain CRFs achieve better results than linear CRFs. This shows that conjunction structure is really important for review mining. For example “although this camera takes great pictures, it is extremely fragile.”, “fragile” is not correctly classified by MaxEnt and Linear CRFs. But the Skip-chain CRFs can correctly classify “fragile” as negative opinion, with conjunction “although”, and the skip edge between “great” and “fragile”. Tree CRFs encode the syntactic tree structure into model representation. Compared with linear-CRFs, the performances are improved for most of expression identification tasks, except for a little decline for product object feature, which may be because that the tags “FB” and “FI” are out of order when transferring to tree structure. There are no significant difference between Skip-Chain CRFs and Tree CRFs. Conjunction structure and syntactic structure represent the semantic dependency from different views. When integrating these two types of dependencies, the Skip-Tree CRFs achieve better overall results than both Skip-Chain CRFs and Tree CRFs.

Table 5 shows the movie review result for Skip Tree model for different types of features. The basic feature only employs word token as feature set. Other features are defined as shown in Figure 4. By adding different features, we find that they all achieve overall improvements than basic feature. The dictionary features are the most important features, especially for positive opinion and negative opinion identification, which shows the importance of prior word’s sentiment. Word features also play important roles: Part of speech is reported useful in several papers (such as Jin and Ho, 2009); the superlative and comparative forms are good indicators for opinion words. Syntactic features acquire limited

improvement in this experiment. They may overlap with CRF based structure model. We also find that sentence level features contribute to the review mining task. Edge feature is also important. It makes the skip edge and tree edge with the semantic representation. When combining all the features, the result is significantly improved compared with any single feature set, which shows that it is crucial to integrate various features for review mining.

A review summary example, generated by our methods, is shown in Figure 1.

6 Conclusion

In this paper, we formulate the review mining task as a joint structure tagging problem. A new framework based on Conditional Random Fields is proposed. The framework can employ rich features to simultaneously extract object features, positive opinions and negative opinions. With this framework, we investigate the chain structure, conjunction structure and syntactic tree structure for review mining. A new unified model, called skip tree CRFs, is proposed for review mining. Through extensive experiments, we show that our proposed framework is effective. It outperforms many state-of-the-art methods.

In future work, we will improve the object feature-opinion pair detection with other learning methods. We also want to cluster the related object features to provide more concise review summary.

Acknowledgement

This work was partly supported by Chinese NSF grant No.60973104 and No. 60803075, a grant from Fujitsu Research Center, and a grant from the International Development Research Center, Ottawa, Canada. We also thank the anonymous reviewers, Qiang Yang, Lei Zhang and Qiu Long for their valuable comments.

References

- A. Berger and Vincent Della Pietra and Stephen A. Della Pietra. 1996. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*.
- E. Breck, Y. Choi, and C. Cardie. 2007. Identifying expressions of opinion in context. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.
- Y. Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In *Proceedings of HLT-EMNLP*.
- X. Ding and Bing Liu. 2007. The Utility of Linguistic Rules in Opinion Mining. In *Proceedings of SIGIR*.
- A. Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of LREC*.
- V. Hatzivassiloglou and K. McKeown. 1997. Predicting the semantic orientation of adjectives. *Proceedings of the Joint ACL/EACL Conference*.
- M. Hu and B. Liu. 2004. Mining and Summarizing Customer Reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*.
- W. Jin, H.H. Ho. 2009. A novel lexicalized HMM-based learning framework for web opinion mining. *Proceedings of the 26th Annual International Conference on Machine Learning (ICML 2009)*.
- J. Lafferty, A. McCallum, F. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proc. 18th International Conf. on Machine Learning (ICML)*.
- D. Klein and Christopher D. Manning. 2003. Fast Exact Inference with a Factored Model for Natural Language Parsing. In *Advances in Neural Information Processing Systems 15 (NIPS 2002)*,
- N. Kobayashi, K. Inui, and Y. Matsumoto. 2007. Opinion Mining from Web documents: Extraction and Structurization. *Journal of Japanese society for artificial intelligence*.
- R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. *Proceedings of the Association for Computational Linguistics (ACL)*.
- Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web*.
- A. Popescu and O. Etzioni. 2005. Extracting Product Features and Opinions from Reviews. *Proceedings of 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP'05)*, 339-346.
- G. Qiu, B. Liu, J. Bu and C. Chen. 2009. Expanding Domain Sentiment Lexicon through Double Propagation, *International Joint Conference on Artificial Intelligence (IJCAI-09)*.
- B. Snyder and R. Barzilay. 2007. Multiple Aspect Ranking using the Good Grief Algorithm", In *Proc. of NAACL*
- C. Sutton, A. McCallum. 2006. An Introduction to Conditional Random Fields for Relational Learning. In "Introduction to Statistical Relational Learning". Edited by Lise Getoor and Ben Taskar. MIT Press.
- I. Titov and R. McDonald. 2008. A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In *Proceeding of the Association for Computational Linguistics (ACL)*.
- P. D. Turney. 2002. Thumbs up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of Association for Computational Linguistics (ACL'02)*.
- M. Wainwright, T. Jaakkola, and A. Willsky. 2001. Tree-based reparameterization for approximate estimation on graphs with cycles. In *Proceedings of Advances in Neural Information Processing Systems (NIPS'2001)*. pp. 1001-1008.
- T. Wilson, Janyce Wiebe, and Paul Hoffmann 2009. Recognizing Contextual Polarity: an exploration of features for phrase-level sentiment analysis. *Computational Linguistics* 35(3).
- J. Zhao, Kang Liu, Gen Wang. 2008. Adding Redundant Features for CRFs-based Sentence Sentiment Classification. In *Proceedings of EMNLP*.
- L. Zhuang, Feng Jing and Xiaoyan Zhu. 2006. Movie Review Mining and Summarization. In *Proceedings of CIKM*.