

Discourse Level Explanatory Relation Extraction from Product Reviews Using First-order Logic

Qi Zhang, Jin Qian, Huan Chen, Jihua Kang, Xuanjing Huang

School of Computer Science

Fudan University

Shanghai, P.R. China

{qz, 12110240030, 12210240054, 12210240059, xjhuang}@fudan.edu.cn

Abstract

Explanatory sentences are employed to clarify reasons, details, facts, and so on. High quality online product reviews usually include not only positive or negative opinions, but also a variety of explanations of why these opinions were given. These explanations can help readers get easily comprehensible information of the discussed products and aspects. Moreover, explanatory relations can also benefit sentiment analysis applications. In this work, we focus on the task of identifying subjective text segments and extracting their corresponding explanations from product reviews in discourse level. We propose a novel joint extraction method using first-order logic to model rich linguistic features and long distance constraints. Experimental results demonstrate the effectiveness of the proposed method.

1 Introduction

Through analyzing product reviews with high helpfulness ratings assigned by readers, we find that a large number of explanatory sentences are used to clarify the causes, details, or consequences of opinions. According to the statistic based on the dataset we crawled from a popular product review website, more than 56.1% opinion expressions are further explained by other sentences. Since most consumers are not experts, these explanations would bring lots of helpful and easy comprehension information for them. Suggestions about writing a product review also advise authors to include not only whether they like or dislike a product, but also why.¹

¹<http://www.reviewpips.com/>
[http://www.amazon.com/gp/community-help/customer-](http://www.amazon.com/gp/community-help/customer-reviews-guidelines)

For example, let us consider the following snippets extracted from online reviews:

Example 1: *TVs with lower refresh rates may suffer from motion blur. If you're watching a fast-paced football game, for example, you may notice a bit of blurring as the players run around the field.*

Example 2: *The LED screen is highly reflective. The reflection of my own face makes it very hard to see the subject I am trying to shoot.*

The first sentence of example 1 expresses negative opinion about refresh rate, which is one of the most important attributes of TV. The second sentence describes the consequence of it through an example. In example 2, detail descriptions are used to explain the reflection problem of the camera screen.

Although, explanations provide valuable information, to the best of our knowledge, there is no existing work that deals with explanation extraction for opinions in discourse level. We think that if explanatory relations can be automatically identified from reviews, sentiment analysis applications may benefit from it. Existing opinion mining approaches mainly focus on subjective text. They try to determine the subjectivity and polarity of fragments of documents (e.g. a paragraph, a sentence, a phrase and a word) (Pang et al., 2002; Riloff et al., 2003; Takamura et al., 2005; Mihalcea et al., 2007; Dasgupta and Ng, ; Hassan and Radev, 2010; Meng et al., 2012; Dragut et al., 2012). Fine-grained methods were also introduced to extract opinion holder, opinion expression, opinion target, and other opinion elements (Kobayashi et al., 2007; Wu et al., [reviews-guidelines](http://www.amazon.com/gp/community-help/customer-reviews-guidelines))

2011; Xu et al., 2013; Yang and Cardie, 2013). Major research directions and challenges of sentiment analysis can also be found in surveys (Pang and Lee, 2008; Liu, 2012).

In this work, we aim to identify subjective text segments and extract their corresponding explanations from product reviews in discourse level. We propose to use Markov Logic Networks (MLN) (Richardson and Domingos, 2006) to learn the joint model for subjective classification and explanatory relation extraction. MLN has been applied in several natural language processing tasks (Singla and Domingos, 2006; Poon and Domingos, 2008; Yoshikawa et al., 2009; Andrzejewski et al., 2011; Song et al., 2012) and demonstrated its advantages. It can easily incorporate rich linguistic features and global constraints by designing various logic formulas, which can also be viewed as templates or rules. Logic formulas are combined in a probabilistic framework to model soft constraints. Hence, the proposed approach can benefit a lot from this framework.

To evaluate the proposed method, we crawled a large number of product reviews and constructed a labeled corpus through Amazon’s Mechanical Turk. Two tasks were deployed for labeling the corpus. We compared the proposed method with state-of-the-art methods on the dataset. Experimental results demonstrate that the proposed approach can achieve better performance than state-of-the-art methods.

The remaining part of this paper is organized as follows: In Section 2, we define the problem and give some examples to show the challenges of this task. Section 3 describes the proposed MLN based method. Dataset construction, experimental results and analyses are given in Section 4. In Section 5, we present the related work and Section 6 concludes the paper.

2 Problem Statement

Motivated by the argument structure of discourse relations used in Penn Discourse Treebank (Rashmi Prasad and Webber, 2008), in this work, we adopt the clause unit-based definition. It means that clauses are treated as the basic units of opinion expressions and explanations. Let $d = \{c_1, c_2, \dots, c_n\}$ be the clauses of document d . Directed graph

$G = (V, E)$ is used to represent the subjectivity of clauses and explanatory relationships between them. In the graph, vertices represent clauses, whose categories are specified by the vertex attributes. Directed edges describe the explanatory relationships between them, of which the heads are explanatory clauses. If clause c_a describes a set of facts which clarify the causes, context, situation, or consequences of another clause c_b , $c_a \rightarrow c_b$ is used to indicate that clause c_a explains c_b .

Adopting clause unit-based definition is based on the following reasons: 1) clause is normally considered as the smallest grammatical unit which can express a complete proposition (Kroeger, 2005); 2) from analyzing online reviews, we observe that a clause can express a complete opinion about one aspect in most of cases; 3) in Penn Discourse Treebank, the basic unit of discourse relations (with a few exceptions) is also taken to be a clause (Rashmi Prasad and Webber, 2008).

Figure 1(a) illustrates a sample document. Figure 1(b) is the corresponding output of the given document. In the graph, vertices whose color are black stand for subjective clauses. The other clauses are represented by white vertices. Edges describe the explanatory relationships between them, of which the heads are explanatory clauses.

Although the explanatory relation extraction task has been studied from the view of linguistic and discourse representation by existing works (Carston, 1993; Lascarides and Asher, 1993), the automatic extraction task is still an open question. Consider the following examples extracting from online reviews:

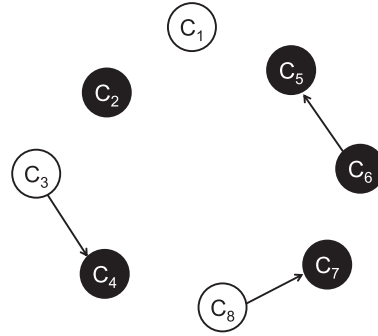
Example 3: *It takes great pictures. Color renditions, skin tones, exposure levels are all first rate.* From the example, we can observe that the second sentence explains the first one. However, the second sentence itself also expresses opinion on various opinion targets. In other words, both subjective and objective sentences can be used as explanations.

Example 4: *When we called their service center they made us wait for them the whole day and no one turned up. This level of service is simply not acceptable.* The first sentence in example 4 explains the second one. Hence, the feature of relative location between two sentences does not always work well in all cases.

Example 5: *This backpack is great! its very big*

(c1) I have both the Panasonic LX3 and the Canon S90. (c2) Both cameras are quite different but truly excellent. (c3) The S90 is a true pocket camera. (c4) It is very compact. (c5) The build quality is also top notch. (c6) It feels solid and it is easy to grip. (c7) It is so small and convenient, (c8) you will find that you will always carry it with you.

(a) Example Review



(b) Directed Graph Representation

Figure 1: Directed graph representation of a sample document.

and fits more than enough stuff. Many sentences, which express explanatory relation, do not contain any connectives (e.g. “because”, “the reason is”, and so on). Lin et al.(2009) generalized four challenges (include ambiguity, inference, context, and world knowledge) to automated implicit discourse relation recognition. In this task, we also need to address those challenges.

From the these examples, we can observe that extracting explanatory relations from product reviews is a challenging task. Both linguistic and global constraints should be carefully studied.

3 The Proposed Approach

In this section, we present our method for jointly classifying the subjectivity of text segments and extracting explanatory relations. Firstly, we briefly describe the framework of Markov Logic Networks. Then, we introduce the clause extraction method based on the definition described in the Section 2. Finally, we present the first-order logic formulas including local formulas and global formulas used for joint modeling in this work.

3.1 Markov Logic Networks

A MLN consists of a set of logic formulas that describe first-order knowledge base. Each formula consists of a set of first-order predicates, logical connectors and variables. Different with first-order logic, these hard logic formulas are softened and can be violated with some penalty (the weight of

formula) in MLN.

We use \mathcal{M} to represent a MLN and $\{(\phi_i, w_i)\}$ to represent formula ϕ_i and its weight w_i . These weighted formulas define a probability distribution over sets of possible worlds. Let y denote a possible world, the $p(y)$ is defined as follows (Richardson and Domingos, 2006):

$$p(y) = \frac{1}{Z} \exp \left(\sum_{(\phi_i, w_i) \in \mathcal{M}} w_i \sum_{c \in C^{n_{\phi_i}}} f_c^{\phi_i}(y) \right),$$

where each c is a binding of free variable in ϕ_i to constraints; $f_c^{\phi_i}(y)$ is a binary feature function that returns 1 if the true value is obtained in the ground formula we get by replacing the free variables in ϕ_i with the constants in c under the given possible world y , and 0 otherwise; $C^{n_{\phi_i}}$ is all possible bindings of variables to constants, and Z is a normalization constant.

Many methods have been proposed to learn the weights of MLN using both generative and discriminative approaches (Richardson and Domingos, 2006; Singla and Domingos, 2006). There are also several MLN learning packages available online such as thebeast², Tuffy³, PyMLNs⁴, Alchemy⁵, and so on.

²<http://code.google.com/p/thebeast>

³<http://hazy.cs.wisc.edu/hazy/tuffy/>

⁴<http://www9-old.in.tum.de/people/jain/mlns/>

⁵<http://alchemy.cs.washington.edu/>

Describing the attributes of words	
$subjLexicon(w)$	The word w belongs to the subjective lexicon (Baccianella et al., 2010).
$relationLexicon(w)$	The word w belongs to the lexicon of explanation relation connectives (Pitler and Nenkova, 2009).
Describing the attributes of the clause c_i	
$word(i, w)$	The clause c_i has word w .
$firstWord(i, w)$	The first word of clause c_i is word w .
$pos(i, w, t)$	The POS tag of word w is t in clause c_i .
$dep(i, h, m)$	Word m and h are governor and dependent of a dependency relation in clause c_i .
Describing the attributes of relations between clause c_i and clause c_j	
$clauseDistance(i, j, m)$	Distance between clause c_i and clause c_j in clauses is m .
$sentenceDistance(i, j, n)$	Distance between clause c_i and clause c_j in sentences is n .

Table 1: Descriptions of observed predicates.

3.2 Clause Identification

We model the clause boundary identification problem through sequence labeling and use Conditional Random Fields (CRFs) to identify clause boundaries. Words and part-of-speech (POS) tags are used as feature sets. Since we do not allow embedded segments, the performance of our method is promising, which achieves the F1 score of 92.8%. The result is comparable with the best results obtained during the CoNLL-2001 campaign (Tjong et al., 2001).

3.3 Formulas

In this work, we propose to use predicate $subj(i)$ to indicate that the i th clause is subjective and $explain(i, j)$ to indicate that the j th clause explains the i th clause. Both $subj$ and $explain$ are hidden predicates and jointly modeled by MLN. We use local and global formulas to model rich linguistic features and long distance constraints.

3.3.1 Local Formulas

The local formulas relate one or more observed predicates to exactly one hidden predicate. In this work, we define a list of observed predicates to describe the properties of individual clauses and attributes of relations between two clauses. The observed predicates and descriptions are shown in

Table 1. The observed predicates can be categorized into 3 groups: words, clauses, and relations between clauses. We use two lexicons to capture background knowledge of words. Lexical, part-of-speech tag, and dependency relation are used to describe a single clause. We also propose two predicates to model distance between clauses.

Table 2 lists the local formulas used in this work. The “+” notation in the formulas indicates that each constant of the logic variable should be weighted separately. For subjective classification and relation extraction, we construct a number of formulas respectively.

For subjective classification, the first two formulas model the influence of lexical and POS tag. It is similar as the bag-of-words model, which is a simplifying representation and has been successfully used for various natural language processing tasks. Since words which provide positive or negative opinions may provide important information for subjectivity classification, we combine predicates of words and lexicon of opinion words. Bigrams are also proved to be useful for textual classification in several NLP tasks. Hence, we also combine predicates about individual word and POS tag to capture this kind of information. Word-level relations are explicitly presented at the dependency trees, we

Formulas for subjective classification

$\text{word}(i, w_+) \Rightarrow \text{subj}(i)$
 $\text{pos}(i, w_+, t_+) \Rightarrow \text{subj}(i)$
 $\text{word}(i, w_+) \wedge \text{subjLexicon}(w) \Rightarrow \text{subj}(i)$
 $\text{pos}(i, w_+, t_+) \wedge \text{subjLexicon}(w) \Rightarrow \text{subj}(i)$
 $\text{word}(i, w_1+) \wedge \text{word}(i, w_2+) \Rightarrow \text{subj}(i)$
 $\text{pos}(i, w_1+, t_+) \wedge \text{pos}(i, w_2+, t_+) \Rightarrow \text{subj}(i)$
 $\text{word}(i, w_1+) \wedge \text{word}(i, w_2+) \wedge \text{subjLexicon}(w_1) \Rightarrow \text{subj}(i)$
 $\text{word}(i, w_1+) \wedge \text{word}(i, w_2+) \wedge \text{subjLexicon}(w_2) \Rightarrow \text{subj}(i)$
 $\text{dep}(i, w_1+, w_2+) \Rightarrow \text{subj}(i)$
 $\text{dep}(i, w_1+, w_2+) \wedge \text{subjLexicon}(w_1) \Rightarrow \text{subj}(i)$
 $\text{dep}(i, w_1+, w_2+) \wedge \text{subjLexicon}(w_2) \Rightarrow \text{subj}(i)$

Formulas for explanatory relation extraction

$\text{word}(i, w_1+) \wedge \text{word}(j, w_2+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{pos}(i, w_1+, t_+) \wedge \text{pos}(j, w_2+, t_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{dep}(i, h_1+, m_1+) \wedge \text{dep}(j, h_2+, m_2+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{word}(i, w_1+) \wedge \text{word}(j, w_2+) \wedge \text{clauseDistance}(i, j, m_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{pos}(i, w_1+, t_+) \wedge \text{pos}(j, w_2+, t_+) \wedge \text{clauseDistance}(i, j, m_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{dep}(i, h_1+, m_1+) \wedge \text{dep}(j, h_2+, m_2+) \wedge \text{clauseDistance}(i, j, m_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{word}(i, w_1+) \wedge \text{word}(j, w_2+) \wedge \text{sentenceDistance}(i, j, n_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{pos}(i, w_1+, t_+) \wedge \text{pos}(j, w_2+, t_+) \wedge \text{sentenceDistance}(i, j, n_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{dep}(i, h_1+, m_1+) \wedge \text{dep}(j, h_2+, m_2+) \wedge \text{sentenceDistance}(i, j, n_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{word}(i, w_1+) \wedge \text{word}(j, w_2+) \wedge \text{firstWord}(j, w_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{pos}(i, w_1+, t_+) \wedge \text{pos}(j, w_2+, t_+) \wedge \text{firstWord}(j, w_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{dep}(i, h_1+, m_1+) \wedge \text{dep}(j, h_2+, m_2+) \wedge \text{firstWord}(j, w_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{word}(i, w_1+) \wedge \text{word}(j, w_2+) \wedge \text{subjLexicon}(w_1) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{pos}(i, w_1+, t_+) \wedge \text{pos}(j, w_2+, t_+) \wedge \text{subjLexicon}(w_1) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{dep}(i, h_1+, m_1+) \wedge \text{dep}(j, h_2+, m_2+) \wedge \text{subjLexicon}(m_1) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{firstWord}(j, w_+) \wedge \text{relationLexicon}(w) \wedge \text{clauseDistance}(i, j, m_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{firstWord}(j, w_+) \wedge \text{relationLexicon}(w) \wedge \text{sentenceDistance}(i, j, n_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{firstWord}(j, w_+) \wedge \text{relationLexicon}(w) \wedge \text{pos}(j, w, t_+) \wedge \text{clauseDistance}(i, j, m_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{firstWord}(j, w_+) \wedge \text{relationLexicon}(w) \wedge \text{pos}(j, w, t_+) \wedge \text{sentenceDistance}(i, j, n_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{firstWord}(j, w_+) \wedge \text{relationLexicon}(w) \wedge \text{word}(i, w_1+) \wedge \text{clauseDistance}(i, j, m_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{firstWord}(j, w_+) \wedge \text{relationLexicon}(w) \wedge \text{word}(i, w_1+) \wedge \text{sentenceDistance}(i, j, n_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{firstWord}(j, w_+) \wedge \text{relationLexicon}(w) \wedge \text{word}(j, w_1+) \wedge \text{clauseDistance}(i, j, m_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$
 $\text{firstWord}(j, w_+) \wedge \text{relationLexicon}(w) \wedge \text{word}(j, w_1+) \wedge \text{sentenceDistance}(i, j, n_+) \wedge j \neq i \Rightarrow \text{explain}(i, j)$

Table 2: Descriptions of local formulas.

also construct local formulas based on predicates extracted from dependency trees of clauses.

For explanatory relation extraction, we firstly use formulas to capture lexical and syntactic information from both of the clauses. Since distances between clauses are helpful in determining the relation, we incorporate two kinds of distance features with lexical and syntactic predicates. Connective words such as *for example*, *since*, explicitly signal the presence of the explanation relation. Although some connective words are ambiguous in terms of relation they mark (Pitler and Nenkova, 2009), they may still be useful for explanation relation extraction. Hence, we construct local formulas with relation lexicon and other predicates.

3.3.2 Global Formulas

Local formulas are designed to deal with subjective classification of a single clause or relation determination of a single pair of clauses. Global formulas are designed to handle global constraints of multiple clauses. From the definition of explanatory relation and corpus statistics, we observe the following properties:

Property 1: One clause can only serve as the explanation of one subjective clause.

Property 2: Explanatory clauses occur immediately before or after their corresponding subjective clauses.

Property 3: The positions of explanatory clauses are consecutive. In other words, if clause c_k and c_{k+2} explain clause c_j , the clause c_{k+1} would also be explanatory clause of c_j .

For property 1, we use the following global formula to make sure that one clause only explains at most one another clause.

$$\text{explain}(i, j) \Rightarrow \neg \text{explain}(k, j) \quad \forall k \neq i, j \quad (1)$$

Based on the property 2 and 3, explanatory clauses are consecutive and immediately before or after their corresponding subjective clauses. We use the following formulas to guarantee the property:

$$\begin{aligned} \text{explain}(i, i+k) \Rightarrow \text{explain}(i, i+m), \\ 1 \leq m \leq k-1 \end{aligned} \quad (2)$$

$$\begin{aligned} \text{explain}(i, i-k) \Rightarrow \text{explain}(i, i-m), \\ 1 \leq m \leq k-1 \end{aligned} \quad (3)$$

Since our aim is to extract explanatory for subjective clauses, we also use the following formulas to make sure that the clauses which are explained are subjective ones.

$$\text{explain}(i, j) \Rightarrow \text{subj}(i) \quad (4)$$

4 Experiments

4.1 Data Set

We crawled a number of reviews about digital cameras from Buzzillions⁶, which is a product review site and contains more than 16 million reviews. We randomly select 100 reviews whose usefulness ratings are 5 on a 5-point scale. They contain 1137 sentences, which are composed by 1665 clauses. Amazon’s Mechanical Turk is used to deploy two tasks for labeling the corpus. 694 clauses are labeled subjective and 478 clauses explain other ones. More than 56.1% opinion expressions are explained by their corresponding explanatory sentences.

The two projects we deployed on Amazon’s Mechanical Turk are: 1) Determine whether a clause contains opinion expressions or not; 2) Determine whether a clause clarifies causes, reasons, or consequences of another given clause. In order to control the labeling quality, we configured parameters of the project to make sure that all the tasks should be judged by at least 20 annotators. Most of the annotators can complete a task within 25 seconds. Figure 2 shows the screenshots of the two projects.

Over all, 127 workers participated in the project. About 72% of them submitted more than 5 tasks. Although we listed several examples on the project descriptions, different people may have their own understanding and criteria for those tasks. In order to measure the quality of the labeling task, we use perplexity to evaluate each task. If the perplexity of a task is below 0.51, which means that more than 80% of the workers submitted the same decision, the result of the task will be used as training or testing data. From the statistic of the corpus, we observe that only 6.2% of the clauses’ subjectiveness and 15.6% of explanation relations can not be certainly decided. For the first project, we treated those clauses as objective one. And, those clause pairs in the second project were not considered as explanation relations.

⁶www.buzzillions.com

Task1: Help us determine whether a sentence is subjective or objective.

The following sentences are extracted from product reviews. Please help us check whether the following sentences expressing opinion towards some attributes/parts of a product.

The battery life is something I come to expect from this line of camera.
()Subjective
()Objective

I have the camera set to shut off the sensor after about 30 seconds
()Subjective
()Objective

Task2: Help us check whether a sentence is an explanation of the opinion sentence.

The opinion sentence (red one) is extracted from product reviews and express opinion towards some attributes/parts of a product. Please help us check whether the following blue sentences describe a set of facts which clarifies the causes, reason, and consequences of the opinion given in the opinion sentence.

click "yes" if there is an explanation relation between them, "no" otherwise.

The battery life is something I come to expect from this line of camera.
I can leave the camera on for better than 8 hours shooting
()YES
()NO

The battery life is something I come to expect from this line of camera.
and I have the camera set to shut off the sensor after about 30 seconds
()YES
()NO

Figure 2: Screenshots of the two tasks on Amazon Mechanical Turk.

4.2 Experiments Configurations

Stanford parser (Klein and Manning, 2003) is used for extracting features from dependency parse trees. For resolving Markov logic network, we use the toolkit *thebeast*⁷. The detailed setting of thebeast engine is as follows: The inference algorithm is the MAP inference with a cutting plane approach. For parameter learning, the weights for formulas are updated by an online learning algorithm with MIRA update rule. All the initial weights are set to zeros. The number of iterations is set to 10 epochs.

Evaluation metrics used for subjectivity classification and relation extraction throughout the experiments include: Precision, Recall, and F1-score. We randomly select 80% reviews as training set and the others as testing set.

Since the dataset is newly created for this task, to compare the performance of the proposed method to other models, we also reimplemented several state-

⁷<http://code.google.com/p/thebeast>

of-the-art methods for comparison.

- **CRF-Subj:** We follow the method proposed by Zhao et al. (2008), which regard the subjectivity of all clauses throughout a paragraph as a sequential flow of sentiments and use CRFs to model it. The feature sets are similar as the local formulas for MLN including words, POS tags, dependency relations, and opinion lexicon.
- **RAE-Subj:** Socher et al. (2011) proposed to use recursive autoencoders for sentence-level predication of sentiment label distributions. To compare with it, we also reimplement their method without any hand designed lexicon.
- **PDTB-Rel:** For discourse relation extraction, we use "PDTB-Styled End-to-End Discourse Parser" (Lin et al., 2010) to extract discourse level relations as baseline. Since it is a general discourse relations identification algorithms, "Cause", "Pragmatic Cause", "Instantiation", and "Restatement" relation types are treated as explanatory relation in this work.
- **SVM-Rel:** We also use LibSVM (Chang and Lin, 2011) to classify the relations between clauses. Following the configurations reported by Feng and Hirst (2012), we use linear kernel and probability estimation to model it.

4.3 Results

Table 3 shows the comparisons of the proposed method with the state-of-the-art systems on subjectivity classification and explanatory relation extraction. From the results, we can observe that recursive autoencoders based subjectivity classification method achieves slightly better performance than our method and conditional random fields based method. The performances of the proposed method are similar as CRFs'. We think that the main reason is that only lexical features are used in MLN models for subjective classification. However, conditional random fields consider not only lexical information but also inference of the contexts of sentences. RAE method learns vector space representations for multi-word phrases and uses compositional semantics to understand sentiment.

Methods	Subjective Classification		
	P	R	F ₁
CRF-Subj	83.5%	76.9%	80.1%
RAE-Subj	85.3%	79.1%	82.1%
MLN	79.2%	80.6%	79.9%
Methods	Relation Extraction		
	P	R	F ₁
RAE-Subj + PDTB-Rel	28.5%	38.6%	32.8%
RAE-Subj + SVM-Rel	32.4%	89.7%	47.6%
MLN	56.2%	72.9%	63.5%

Table 3: Performance comparisons between the proposed method and state-of-the-art methods. “MLN” represents the method proposed in this work.

For evaluating the performance of relation extraction, we combine the results of RAE with PDTB-Rel and SVM-Rel. For all the subjective clauses identified by RAE, PDTB-Rel and SVM-Rel are used to extract corresponding explanatory clauses. The results are shown in the last three rows in the Table 3. From the results, we can observe that the proposed joint model achieves best F1 score and precision among all methods. Although the proposed method achieve slightly worse result in processing subjectivity classification. We think that the error propagation is the main reason for worse results of cascaded methods. The relative improvement of MLN over SVM-Rel is more than 33.4%.

To show the effectiveness of different observed predicates, we evaluate the performances of the proposed method with different predicate sets. We subtract one observed predicate and its corresponding local formulas from the original sets at a time. The results of both subjectivity classification and relation extraction are shown in Table 4. The first row shows the result of the MLN based method with all observed predicates and local formulas. From the results we can observe that the observed predicates which are not used in the local formulas for subjectivity classification also impact the performance of subjectivity classification. We think that the performance is effected by the global formulas, which combine the procedure of subjectivity classification

and relation extraction. Among all predicates, we observe that words and dependency relations play the most important roles. Without word predicate, the F1 score of subjectivity classification and relation extraction significantly drop to 51.2% and 42.9% respectively. For subjectivity classification, subjective lexicon contributes a lot for recall. For relation extraction, the impacts of clause distance and sentence distance are not as significant as the other features.

5 Related Work

Our work relates to three research areas: sentiment analysis/opinion mining, discourse-level relation extraction, and Markov logic networks. Along with the increasing requirement, subjectivity classification has recently received considerable attention from both the industry and researchers. A variety of approaches and methods have been proposed for this task from different aspects. Among them, a number of approaches focus on classifying sentiments of text in different levels (e.g. words (Kim and Hovy, 2004), phrases (Wilson et al., 2005), sentences (Zhao et al., 2008), documents (Pang et al., 2002) and so on.), and detecting the overall polarity of them.

Another research direction tries to convert the sentiment analysis task into entity identification and relation extraction. Hu and Liu (2004) proposed to use a set of methods to produce feature-based summary of a large number of customer reviews. Kobayashi et al. (2007) assumed that evaluative opinions could be structured as a frame which is composed by opinion holder, subject, aspect, and evaluation. They converted the task to two kinds of relation extraction tasks and proposed a machine learning-based method which used both contextual and statistical clues.

Analysis of some special types of sentences were also introduced in recent years. Jindal and Liu (2006) studied the problem of identifying comparative sentences. They analyzed different types of comparative sentences and proposed learning approaches to identify them. Conditional sentences were studied by Narayanan et al (2009). They analyzed the conditional sentences in both linguistic and computational perspectives and used learning

	Subjective Classification			Relation Extraction		
	P	R	F1	P	R	F1
MLN	79.2%	80.6%	79.9%	56.2%	72.9%	63.5%
$-subjLexicon(w)$	76.6%	70.4%	73.4 %	52.3%	68.6%	59.4%
$-relationLexicon(w)$	78.2%	79.4%	78.8%	53.6%	70.8%	61.0%
$-word(i, w)$	52.8%	49.6%	51.2 %	36.4%	52.1%	42.9%
$-firstWord(i, w)$	76.3%	80.1%	78.2%	56.9%	69.8%	62.7%
$-pos(i, w, t)$	72.6%	76.8%	74.6 %	52.4%	60.2%	56.0%
$-dep(i, h, m)$	57.6%	70.6%	63.4%	41.2%	56.8%	47.8%
$-clauseDistance(i, j, m)$	78.9%	80.2%	79.5%	52.6%	70.6%	60.3%
$-sentenceDistance(i, j, n)$	78.6%	80.3%	79.4%	52.4%	70.8%	60.2%

Table 4: Performance comparisons of different observed predicates

method to do it. They followed the *feature-based sentiment analysis* model (Hu and Liu, 2004), which also use flat frames to represent evaluations.

Since the cross sentences relations are considered in this work, the discourse-level relation extraction methods are also related to ours. Marcu and Echiabi (2002) proposed to use an unsupervised approach to recognizing discourse relations. Lin et al.(2009) analyzed the impacts of features extracted from contextual information, constituent parse trees, dependency parse trees, and word pairs. Asher et al.(2009) studied discourse segments containing opinion expressions from the perspective of linguistics. Chen et al. (2010) introduced a multi-label model to detect emotion causes. They developed two sets of linguistic features for this task base on linguistic cues. Zirn et al. (2011) proposed to use MLN framework to capture the context information in analysing (sub-)sentences.

The most similar work to ours was proposed by Somasundaran et al.(2009). They proposed to use iterative classification algorithm to capture discourse-level associations. However different to us, they focused on pairwise relationships between opinion expressions. In this paper, we used MLN framework to capture another different discourse-level relation, which exists between subject clauses or subject clause and objective clause.

Richardson and Domingos (2006) proposed Markov Logic Networks, which combines first-order logic and probabilistic graphical models. In

recent years, MLN has been adopted for several natural language processing tasks and achieved a certain level of success (Singla and Domingos, 2006; Riedel and Meza-Ruiz, 2008; Yoshikawa et al., 2009; Andrzejewski et al., 2011; Jiang et al., 2012; Huang et al., 2012). Singla and Domingos (2006) modeled the entity resolution problem with MLN. They demonstrated the capability of MLN to seamlessly combine a number of previous approaches. Poon and Domingos (2008) proposed to use MLN for joint unsupervised coreference resolution. Yoshikawa et al. (2009) proposed to use Markov logic to incorporate both local features and global constraints that hold between temporal relations. Andrzejewski et al. (2011) introduced a framework for incorporating general domain knowledge, which is represented by First-Order Logic (FOL) rules, into LDA inference to produce topics shaped by both the data and the rules.

6 Conclusions

In this paper, we propose to use Markov logic networks to identify subjective text segments and extract their corresponding explanations in discourse level. We use MLN to jointly model subjectivity classification and explanatory relation extraction. Rich linguistic features and global constraints are incorporated by various logic formulas and global formulas. To evaluate the proposed method, we collected a large number of product reviews and

constructed a labeled corpus through Amazon’s Mechanical Turk. Experimental results demonstrate that the proposed approach achieve better performance than state-of-the-art methods.

7 Acknowledgement

The authors wish to thank the anonymous reviewers for their helpful comments and Kang Han for preparing the corpus. This work was partially funded by National Natural Science Foundation of China (61003092, 61073069), Key Projects in the National Science & Technology Pillar Program(2012BAH18B01), National Major Science and Technology Special Project of China (2014ZX03006005), Shanghai Municipal Science and Technology Commission (12511504502) and “Chen Guang” project supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation(11CG05).

References

- David Andrzejewski, Xiaojin Zhu, Mark Craven, and Benjamin Recht. 2011. A framework for incorporating general domain knowledge into latent dirichlet allocation using first-order logic. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Two*, IJCAI’11, pages 1171–1177. AAAI Press.
- Nicholas Asher, Farah Benamara, and Yannick Mathieu. 2009. Appraisal of opinion expressions in discourse. *Linguisticae Investigationes*.
- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, Valletta, Malta, may. European Language Resources Association (ELRA).
- R. Carston. 1993. Conjunction, explanation and relevance. *Lingua* 90, pages 27–48.
- Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May.
- Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *Proceedings of Coling 2010*.
- Sajib Dasgupta and Vincent Ng. Mine the easy, classify the hard: A semi-supervised approach to automatic sentiment classification. In *Proceedings of ACL-IJCNLP 2009*.
- Eduard Dragut, Hong Wang, Clement Yu, Prasad Sistla, and Weiyi Meng. 2012. Polarity consistency checking for sentiment dictionaries. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 997–1005, Jeju Island, Korea, July. Association for Computational Linguistics.
- Vanessa Wei Feng and Graeme Hirst. 2012. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 60–68, Jeju Island, Korea, July. Association for Computational Linguistics.
- Ahmed Hassan and Dragomir R. Radev. 2010. Identifying text polarity using random walks. In *Proceedings of ACL 2010*, Uppsala, Sweden, July.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of SIGKDD 2004*.
- Minlie Huang, Xing Shi, Feng Jin, and Xiaoyan Zhu. 2012. Using first-order logic to compress sentences. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- Shangpu Jiang, D. Lowd, and Dejing Dou. 2012. Learning to refine an automatically extracted knowledge base using markov logic. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 912–917.
- Nitin Jindal and Bing Liu. 2006. Identifying comparative sentences in text documents. In *Proceedings of SIGIR 2006*.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of COLING 2004*.
- Dan Klein and Christopher D. Manning. 2003. Fast exact inference with a factored model for natural language parsing. In *Proceedings of NIPS 2003*.
- Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. 2007. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proceedings of EMNLP-CoNLL 2007*.
- Paul Kroeger. 2005. *Analyzing Grammar: An Introduction*. Cambridge.
- Alex Lascarides and Nicholas Asher. 1993. Temporal interpretation, discourse relations, and common sense entailment. *Linguistics and Philosophy*, 16(5):437–493.
- Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. 2009. Recognizing implicit discourse relations in the penn discourse treebank. In *Proceedings of EMNLP 2009*.

- Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2010. A pdtb-styled end-to-end discourse parser. *CoRR*, abs/1011.0835.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Daniel Marcu and Abdessamad Echihabi. 2002. An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, pages 368–375.
- Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu, and Houfeng Wang. 2012. Cross-lingual mixture model for sentiment classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 572–581, Jeju Island, Korea, July. Association for Computational Linguistics.
- Rada Mihalcea, Carmen Banea, and Janyce Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of ACL 2007*.
- Ramanathan Narayanan, Bing Liu, and Alok Choudhary. 2009. Sentiment analysis of conditional sentences. In *Proceedings of EMNLP 2009*.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2:1–135, January.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP 2002*.
- Emily Pitler and Ani Nenkova. 2009. Using syntax to disambiguate explicit discourse connectives in text. In *Proceedings of the ACL-IJCNLP 2009*.
- Hoifung Poon and Pedro Domingos. 2008. Joint unsupervised coreference resolution with markov logic. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, pages 650–659, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Alan Lee Eleni Miltsakaki Livio Robaldo Aravind Joshi Rashmi Prasad, Nikhil Dinesh and Bonnie Webber. 2008. The penn discourse treebank 2.0. In Bente Maegaard Joseph Mariani Jan Odjik Stelios Piperidis Daniel Tapias Nicoletta Calzolari (Conference Chair), Khalid Choukri, editor, *Proceedings of LREC'08*, Marrakech, Morocco, may. <http://www.lrec-conf.org/proceedings/lrec2008/>.
- Matthew Richardson and Pedro Domingos. 2006. Markov logic networks. *Machine Learning*, 62(1-2):107–136.
- Sebastian Riedel and Ivan Meza-Ruiz. 2008. Collective semantic role labelling with markov logic. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning, CoNLL '08*, pages 193–197, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of HLT-NAACL 2003*.
- P. Singla and P. Domingos. 2006. Entity resolution with markov logic. In *Data Mining, 2006. ICDM '06. Sixth International Conference on*, pages 572–582.
- Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. 2011. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11*, pages 151–161, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Swapna Somasundaran, Galileo Namata, Lise Getoor, and Janyce Wiebe. 2009. Opinion graphs for polarity and discourse classification. In *Proceedings of TextGraphs-4*.
- Yang Song, Jing Jiang, Wayne Xin Zhao, Sujian Li, and Houfeng Wang. 2012. Joint learning for coreference resolution with markov logic. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1245–1254, Jeju Island, Korea, July. Association for Computational Linguistics.
- Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of ACL 2005*.
- Erik F. Tjong, Kim Sang, and Hervé Déjean. 2001. Introduction to the conll-2001 shared task: clause identification. In *Proceedings of the 2001 workshop on Computational Natural Language Learning - Volume 7, ConLL '01*, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT-EMNLP 2005*.
- Yuanbin Wu, Qi Zhang, Xuangjing Huang, and Lide Wu. 2011. Structural opinion mining for graph-based sentiment representation. In *Proceedings of EMNLP 2011*.
- Liheng Xu, Kang Liu, Siwei Lai, Yubo Chen, and Jun Zhao. 2013. Walk and learn: a two-stage approach for opinion words and opinion targets co-extraction. In *Proceedings of the 22nd international conference on World Wide Web companion, WWW '13 Companion*, pages 95–96, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

- Bishan Yang and Claire Cardie. 2013. Joint inference for fine-grained opinion extraction. In *Proceedings of ACL 2013*.
- Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara, and Yuji Matsumoto. 2009. Jointly identifying temporal relations with markov logic. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1*, ACL '09, pages 405–413, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jun Zhao, Kang Liu, and Gen Wang. 2008. Adding redundant features for crfs-based sentence sentiment classification. In *Proceedings of EMNLP 2008*.
- Cäcilia Zirn, Mathias Niepert, Heiner Stuckenschmidt, and Michael Strube. 2011. Fine-grained sentiment analysis with structural features. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 336–344, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.