

# Joint Inference for Fine-grained Opinion Extraction

**Bishan Yang**

Department of Computer Science  
Cornell University  
bishan@cs.cornell.edu

**Claire Cardie**

Department of Computer Science  
Cornell University  
cardie@cs.cornell.edu

## Abstract

This paper addresses the task of fine-grained opinion extraction – the identification of opinion-related entities: the opinion expressions, the opinion holders, and the targets of the opinions, and the relations between opinion expressions and their targets and holders. Most existing approaches tackle the extraction of opinion entities and opinion relations in a pipelined manner, where the interdependencies among different extraction stages are not captured. We propose a joint inference model that leverages knowledge from predictors that optimize subtasks of opinion extraction, and seeks a globally optimal solution. Experimental results demonstrate that our joint inference approach significantly outperforms traditional pipeline methods and baselines that tackle subtasks in isolation for the problem of opinion extraction.

## 1 Introduction

Fine-grained opinion analysis is concerned with identifying opinions in text at the expression level; this includes identifying the subjective (i.e., opinion) expression itself, the opinion holder and the target of the opinion (Wiebe et al., 2005). The task has received increasing attention as many natural language processing applications would benefit from the ability to identify text spans that correspond to these key components of opinions. In question-answering systems, for example, users may submit questions in the form “What does entity A think about target B?”; opinion-oriented summarization systems also need to recognize opinions and their targets and holders.

In this paper, we address the task of identifying opinion-related entities and opinion relations. We

consider three types of opinion entities: *opinion expressions* or direct subjective expressions as defined in Wiebe et al. (2005) — expressions that explicitly indicate emotions, sentiment, opinions or other private states (Quirk et al., 1985) or speech events expressing private states; *opinion targets* — expressions that indicate what the opinion is about; and *opinion holders* — mentions of whom or what the opinion is from. Consider the following examples in which opinion expressions (O) are underlined and targets (T) and holders (H) of the opinion are bracketed.

S1: [The workers]<sub>[H<sub>1,2</sub>]</sub> were irked<sub>[O<sub>1</sub>]</sub>  
by [the government report]<sub>[T<sub>1</sub>]</sub> and  
were worried<sub>[O<sub>2</sub>]</sub> as they went about  
their daily chores.

S2: From the very start it could be  
predicted<sub>[O<sub>1</sub>]</sub> that on the subject of  
economic globalization, [the developed  
states]<sub>[T<sub>1,2</sub>]</sub> were going to come across  
fierce opposition<sub>[O<sub>2</sub>]</sub>.

The numeric subscripts denote *linking relations*, one of IS-ABOUT or IS-FROM. In S1, for instance, opinion expression “were irked” ( $O_1$ ) IS-ABOUT “the government report” ( $T_1$ ). Note that the IS-ABOUT relation can contain an empty target (e.g. “were worried” in S1); similarly for IS-FROM w.r.t. the opinion holder (e.g. “predicted” in S2). We also allow an opinion entity to be involved in multiple relations (e.g. “the developed states” in S2).

Not surprisingly, fine-grained opinion extraction is a challenging task due to the complexity and variety of the language used to express opinions and their components (Pang and Lee, 2008). Nevertheless, much progress has been made in extracting opinion information from text. Sequence labeling models have been successfully employed to identify opinion expressions (e.g. (Breck et al.,

2007; Yang and Cardie, 2012)) and relation extraction techniques have been proposed to extract opinion holders and targets based on their linking relations to the opinion expressions (e.g. Kim and Hovy (2006), Kobayashi et al. (2007)). However, most existing work treats the extraction of different opinion entities and opinion relations in a pipelined manner: the interaction between different extraction tasks is not modeled jointly and error propagation is not considered. One exception is Choi et al. (2006), which proposed an ILP approach to jointly identify opinion holders, opinion expressions and their IS-FROM linking relations, and demonstrated the effectiveness of joint inference. Their ILP formulation, however, does not handle *implicit linking relations*, i.e. opinion expressions with no explicit opinion holder; nor does it consider IS-ABOUT relations.

In this paper, we present a model that jointly identifies opinion-related entities, including *opinion expressions*, *opinion targets* and *opinion holders* as well as the associated opinion linking relations, IS-ABOUT and IS-FROM. For each type of opinion relation, we allow implicit (i.e. empty) arguments for cases when the opinion holder or target is not explicitly expressed in text. We model entity identification as a sequence tagging problem and relation extraction as binary classification. A joint inference framework is proposed to jointly optimize the predictors for different sub-problems with constraints that enforce global consistency. We hypothesize that the ambiguity in the extraction results will be reduced and thus, performance increased. For example, uncertainty w.r.t. the spans of opinion entities can adversely affect the prediction of opinion relations; and evidence of opinion relations might provide clues to guide the accurate extraction of opinion entities.

We evaluate our approach using a standard corpus for fine-grained opinion analysis (the MPQA corpus (Wiebe et al., 2005)) and demonstrate that our model outperforms by a significant margin traditional baselines that do not employ joint inference for extracting opinion entities and different types of opinion relations.

## 2 Related Work

Significant research effort has been invested into fine-grained opinion extraction for open-domain text such as news articles (Wiebe et al., 2005; Wilson et al., 2009). Many techniques were proposed

to identify the text spans for opinion expressions (e.g. (Breck et al., 2007; Johansson and Moschitti, 2010b; Yang and Cardie, 2012)), opinion holders (e.g. (Choi et al., 2005)) and topics of opinions (Stoyanov and Cardie, 2008). Some consider extracting opinion targets/holders along with their relation to the opinion expressions. Kim and Hovy (2006) identifies opinion holders and targets by using their semantic roles related to opinion words. Ruppenhofer et al. (2008) argued that semantic role labeling is not sufficient for identifying opinion holders and targets. Johansson and Moschitti (2010a) extract opinion expressions and holders by applying reranking on top of sequence labeling methods. Kobayashi et al. (2007) considered extracting “aspect-evaluation” relations (relations between opinion expressions and targets) by identifying opinion expressions first and then searching for the most likely target for each opinion expression via a binary relation classifier. All these methods extract opinion arguments and opinion relations in separate stages instead of extracting them jointly.

Most similar to our method is Choi et al. (2006), which jointly extracts opinion expressions, holders and their IS-FROM relations using an ILP approach. In contrast, our approach (1) also considers the IS-ABOUT relation which is arguably more complex due to the larger variety in the syntactic structure exhibited by opinion expressions and their targets, (2) handles implicit opinion relations (opinion expressions without any associated argument), and (3) uses a simpler ILP formulation.

There has also been substantial interest in opinion extraction from product reviews (Liu, 2012). Most existing approaches focus on the extraction of opinion targets and their associated opinion expressions and usually employ a pipeline architecture: generate candidates of opinion expressions and opinion targets first, and then use rule-based or machine-learning-based approaches to identify potential relations between opinions and targets (Hu and Liu, 2004; Wu et al., 2009; Liu et al., 2012). In addition to pipeline approaches, bootstrapping-based approaches were proposed (Qiu et al., 2009; Qiu et al., 2011; Zhang et al., 2010) to identify opinion expressions and targets iteratively; however, they suffer from the problem of error propagation.

There is much work demonstrating the benefit of performing global inference. Roth and Yih

(2004) proposed a global inference approach in the formulation of a linear program (LP) and applied it to the task of extracting named entities and relations simultaneously. Their problem is similar to ours — the difference is that Roth and Yih Roth and Yih (2004) assume that named entity spans are known a priori and only their labels need to be assigned. Joint inference has also been applied to semantic role labeling (Punyakanok et al., 2008; Srikumar and Roth, 2011; Das et al., 2012), where the goal is to jointly identify semantic arguments for given lexical predicates. The problem is conceptually similar to identifying opinion arguments for opinion expressions, however, we do not assume prior knowledge of opinion expressions (unlike in SRL, where predicates are given).

### 3 Model

As proposed in Section 1, we consider the task of jointly identifying opinion entities and opinion relations. Specifically, given a sentence, our goal is to identify spans of opinion expressions, opinion arguments (targets and holders) and their associated linking relations. Training data consists of text with manually annotated opinion expression and argument spans, each with a list of relation ids specifying the linking relation between opinion expressions and their arguments.

In this section, we will describe how we model opinion entity identification and opinion relation extraction, and how we combine them in a joint inference model.

#### 3.1 Opinion Entity Identification

We formulate the task of opinion entity identification as a sequence labeling problem and employ conditional random fields (CRFs) (Lafferty et al., 2001) to learn the probability of a sequence assignment  $\mathbf{y}$  for a given sentence  $\mathbf{x}$ . Through inference we can find the best sequence assignment for sentence  $x$  and recover the opinion entities according to the standard “IOB” encoding scheme. We consider four entity labels:  $D, T, H, N$ , where  $D$  denotes opinion expressions,  $T$  denotes opinion targets,  $H$  denotes opinion holders and  $N$  denotes “NONE” entities.

We define potential function  $f_{iz}$  that gives the probability of assigning a span  $i$  with entity label  $z$ , and the probability is estimated based on the learned parameters from CRFs. Formally, given a within-sentence span  $i = (a, b)$ , where  $a$  is the

starting position and  $b$  is the end position, and label  $z \in \{D, T, H\}$ , we have

$$f_{iz} = p(\mathbf{y}_a = B_z, \mathbf{y}_{a+1} = I_z, \dots, \mathbf{y}_b = I_z, \mathbf{y}_{b+1} \neq I_z | \mathbf{x})$$

$$f_{iN} = p(\mathbf{y}_a = O, \dots, \mathbf{y}_b = O | \mathbf{x})$$

These probabilities can be efficiently computed using the forward-backward algorithm.

#### 3.2 Opinion Relation Extraction

We consider extracting the IS-ABOUT and IS-FROM opinion relations. In the following we will not distinguish these two relations, since they can both be characterized as relations between opinion expressions and opinion arguments, and the methods for relation extraction are the same.

We treat the relation extraction problem as a combination of two binary classification problems: *opinion-arg classification*, which decides whether a pair consisting of an opinion candidate  $o$  and an argument candidate  $a$  forms a relation; and *opinion-implicit-arg classification*, which decides whether an opinion candidate  $o$  is linked to an implicit argument, i.e. no argument is mentioned. We define a potential function  $r$  to capture the strength of association between an opinion candidate  $o$  and an argument candidate  $a$ ,

$$r_{oa} = p(y = 1|x) - p(y = 0|x)$$

where  $p(y = 1|x)$  and  $p(y = 0|x)$  are the logistic regression estimates of the positive and negative relations. Similarly, we define potential  $r_{o\emptyset}$  to denote the confidence of predicting opinion span  $o$  associated with an implicit argument.

##### 3.2.1 Opinion-Arg Relations

For *opinion-arg* classification, we construct candidates of opinion expressions and opinion arguments and consider each pair of an opinion candidate and an argument candidate as a potential opinion relation. Conceptually, all possible subsequences in the sentence are candidates. To filter out candidates that are less reasonable, we consider the opinion expressions and arguments obtained from the  $n$ -best predictions by CRFs<sup>1</sup>. We also employ syntactic patterns from dependency

<sup>1</sup>We randomly split the training data into 10 parts and obtained the 50-best CRF predictions on each part for the generation of candidates. We also experimented with candidates generated from more CRF predictions, but did not find any performance improvement for the task.

trees to generate candidates. Specifically, we selected the most common patterns of the shortest dependency paths<sup>2</sup> between an opinion candidate  $o$  and an argument candidate  $a$  in our dataset, and include all pairs of candidates that satisfy at least one dependency pattern. For the IS-ABOUT relation, the top three patterns are (1)  $o \uparrow_{dobj} a$ , (2)  $o \uparrow_{ccomp} x \uparrow_{nsubj} a$  ( $x$  is a word in the path that is not covered by either  $o$  nor  $a$ ), (3)  $o \uparrow_{ccomp} a$ ; for the IS-FROM relation, the top three patterns are (1)  $o \uparrow_{nsubj} a$ , (2)  $o \uparrow_{poss} a$ , (3)  $o \downarrow_{ccomp} x \uparrow_{nsubj} a$ .

Note that generating candidates this way will give us a large number of negative examples. Similar to the preprocessing approach in (Choi et al., 2006), we filter pairs of opinion and argument candidates that do not overlap with any gold standard relation in our training data.

Many features we use are common features in the SRL tasks (Punyakanok et al., 2008) due to the similarity of opinion relations to the predicate-argument relations in SRL (Ruppenhofer et al., 2008; Choi et al., 2006). In general, the features aim to capture (a) local properties of the candidate opinion expressions and arguments and (b) syntactic and semantic attributes of their relation.

**Words and POS tags:** the words contained in the candidate and their POS tags.

**Lexicon:** For each word in the candidate, we include its WordNet hypernyms and its strength of subjectivity in the Subjectivity Lexicon<sup>3</sup> (e.g. weaksubj, strongsubj).

**Phrase type:** the syntactic category of the deepest constituent that covers the candidate in the parse tree, e.g. NP, VP.

**Semantic frames:** For each verb in the opinion candidate, we include its frame types according to FrameNet<sup>4</sup>.

**Distance:** the relative distance (number of words) between the opinion and argument candidates.

**Dependency Path:** the shortest path in the dependency tree between the opinion candidate and the target candidate, e.g.  $ccomp \uparrow nsubj \uparrow$ . We also include word types and POS types in the paths, e.g.  $opinion \uparrow_{ccomp} suffering \uparrow_{nsubj} patient$ ,

<sup>2</sup>We use the Stanford Parser to generate parse trees and dependency graphs.

<sup>3</sup>[http://mpqa.cs.pitt.edu/lexicons/subj\\_lexicon/](http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/)

<sup>4</sup><https://framenet.icsi.berkeley.edu/fndrupal/>

$NN \uparrow_{ccomp} VBG \uparrow_{nsubj} NN$ . The dependency path has been shown to be very useful in extracting opinion expressions and opinion holders (Johansson and Moschitti, 2010a).

### 3.2.2 Opinion-Implicit-Arg Relations

When the *opinion-arg* relation classifier predicts that there is no suitable argument for the opinion expression candidate, it does not capture the possibility that an opinion candidate may associate with an implicit argument. To incorporate knowledge of implicit relations, we build an *opinion-implicit-arg* classifier to identify an opinion candidate with an implicit argument based on its own properties and context information.

For training, we consider all gold-standard opinion expressions as training examples — including those with implicit arguments — as positive examples and those associated with explicit arguments as negative examples. For features, we use words, POS tags, phrase types, lexicon and semantic frames (see Section 3.2.1 for details) to capture the properties of the opinion expression, and also features that capture the context of the opinion expression:

**Neighboring constituents:** The words and grammatical roles of neighboring constituents of the opinion expression in the parse tree — the left and right sibling of the deepest constituent containing the opinion expression in the parse tree.

**Parent Constituent:** The grammatical role of the parent constituent of the deepest constituent containing the opinion expression.

**Dependency Argument:** The word types and POS types of the arguments of the dependency patterns in which the opinion expression is involved. We consider the same dependency patterns that are used to generate candidates for *opinion-arg* classification.

### 3.3 Joint Inference

The inference goal is to find the optimal prediction for both opinion entity identification and opinion relation extraction. For a given sentence, we denote  $\mathcal{O}$  as a set of opinion candidates,  $\mathcal{A}_k$  as a set of argument candidates, where  $k$  denotes the type of opinion relation — IS-ABOUT or IS-FROM — and  $\mathcal{S}$  as a set of within-sentence spans that cover all of the opinion candidates and argument can-

didates. We introduce binary variable  $x_{iz}$ , where  $x_{iz} = 1$  means span  $i$  is associated with label  $z$ . We also introduce binary variable  $u_{ij}$  for every pair of opinion candidate  $i$  and argument candidate  $j$ , where  $u_{ij} = 1$  means  $i$  forms an opinion relation with  $j$ , and binary variable  $v_{ik}$  for every opinion candidate  $i$  in relation type  $k$ , where  $v_{ik} = 1$  means  $i$  associates with an implicit argument in relation  $k$ . Given the binary variables  $x_{iz}, u_{ij}, v_{ik}$ , it is easy to recover the entity and relation assignment by checking which spans are labeled as opinion entities, and which opinion span and argument span form an opinion relation.

The objective function is defined as a linear combination of the potentials from different predictors with a parameter  $\lambda$  to balance the contribution of two components: opinion entity identification and opinion relation extraction.

$$\begin{aligned} & \arg \max_{x,u,v} \lambda \sum_{i \in \mathcal{S}} \sum_z f_{iz} x_{iz} \\ & + (1 - \lambda) \sum_k \sum_{i \in \mathcal{O}} \left( \sum_{j \in \mathcal{A}_k} r_{ij} u_{ij} + r_{i\emptyset} v_{ik} \right) \end{aligned} \quad (1)$$

It is subject to the following linear constraints:

Constraint 1: *Uniqueness*. For each span  $i$ , we must assign one and only one label  $z$ , where  $z \in \{H, D, T, N\}$ .

$$\sum_z x_{iz} = 1$$

Constraint 2: *Non-overlapping*. If two spans  $i$  and  $j$  overlap, then at most one of the spans can be assigned to a non-NONE entity label:  $H, D, T$ .

$$\sum_{z \neq N} x_{iz} + \sum_{z \neq N} x_{jz} \leq 1$$

Constraint 3: *Consistency between the opinion-arg and opinion-implicit-arg classifiers*. For an opinion candidate  $i$ , if it is predicted to have an implicit argument in relation  $k$ ,  $v_{ik} = 1$ , then no argument candidate should form a relation with  $i$ . If  $v_{ik} = 0$ , then there exists some argument candidate  $j \in \mathcal{A}_k$  such that  $u_{ij} = 1$ . We introduce two auxiliary binary variables  $a_{ik}$  and  $b_{ik}$  to limit the maximum number of relations associated with each opinion candidate to be less than or equal to

three<sup>5</sup>. When  $v_{ik} = 1$ ,  $a_{ik}$  and  $b_{ik}$  have to be 0.

$$\sum_{j \in \mathcal{A}_k} u_{ij} = 1 - v_{ik} + a_{ik} + b_{ik}$$

$$a_{ik} \leq 1 - v_{ik}, b_{ik} \leq 1 - v_{ik}$$

Constraint 4: *Consistency between opinion-arg classifier and opinion entity extractor*. Suppose an argument candidate  $j$  in relation  $k$  is assigned an argument label by the entity extractor, that is  $x_{jz} = 1$  ( $z = T$  for IS-ABOUT relation and  $z = H$  for IS-FROM relation), then there exists some opinion candidates that associate with  $j$ . Similar to constraint 3, we introduce auxiliary binary variables  $c_j$  and  $d_j$  to enforce that an argument  $j$  links to at most three opinion expressions. If  $x_{jz} = 0$ , then no relations should be extracted for  $j$ .

$$\sum_{i \in \mathcal{O}} u_{ij} = x_{jz} + c_{jk} + d_{jk}$$

$$c_{jk} \leq x_{jz}, d_{jk} \leq x_{jz}$$

Constraint 5: *Consistency between the opinion-implicit-arg classifier and opinion entity extractor*. When an opinion candidate  $i$  is predicted to associate with an implicit argument in relation  $k$ , that is  $v_{ik} = 1$ , then we allow  $x_{iD}$  to be either 1 or 0 depending on the confidence of labeling  $i$  as an opinion expression. When  $v_{ik} = 0$ , there exists some opinion argument associated with the opinion candidate, and we enforce  $x_{iD} = 1$ , which means the entity extractor agrees to label  $i$  as an opinion expression.

$$v_{ik} + x_{iD} \geq 1$$

Note that in our ILP formulation, the label assignment for a candidate span involves one multiple-choice decision among different opinion entity labels and the ‘‘NONE’’ entity label. The scores of different label assignments are comparable for the same span since they come from one entity extraction model. This makes our ILP formulation advantageous over the ILP formulation proposed in Choi et al. (2006), which needs  $m$  binary decisions for a candidate span, where  $m$  is the number of types of opinion entities, and the score for each possible label assignment is obtained by

<sup>5</sup>It is possible to add more auxiliary variables to allow more than three arguments to link to an opinion expression, but this rarely happens in our experiments. For the IS-FROM relation, we set  $a_{ik} = 0, b_{ik} = 0$  since an opinion expression usually has only one holder.

the sum of raw scores from  $m$  independent extraction models. This design choice also allows us to easily deal with multiple types of opinion arguments and opinion relations.

## 4 Experiments

For evaluation, we used version 2.0 of the MPQA corpus (Wiebe et al., 2005; Wilson, 2008), a widely used data set for fine-grained opinion analysis.<sup>6</sup> We considered the subset of 482 documents<sup>7</sup> that contain attitude and target annotations. There are a total of 9,471 sentences with opinion-related labels at the phrase level. We set aside 132 documents as a development set and use 350 documents as the evaluation set. All experiments employ 10-fold cross validation on the evaluation set; the average over the 10 runs is reported.

Our gold standard opinion expressions, opinion targets and opinion holders correspond to the direct subjective annotations, target annotations and agent annotations, respectively. The IS-FROM relation is obtained from the *agent* attribute of each opinion expression. The IS-ABOUT relation is obtained from the attitude annotations: each opinion expression is annotated with attitude frames and each attitude frame is associated with a list of targets. The relations may overlap: for example, in the following sentence, the target of relation 1 contains relation 2.

[John]<sub>H<sub>1</sub></sub> is happy<sub>O<sub>1</sub></sub> because [[he]<sub>H<sub>2</sub></sub>  
loves<sub>O<sub>2</sub></sub> [being at Enderly Park]<sub>T<sub>2</sub></sub>]<sub>T<sub>1</sub></sub>.

We discard relations that contain sub-relations because we believe that identifying the sub-relations usually is sufficient to recover the discarded relations. (Prediction of overlapping relations is considered as future work.) In the example above, we will identify (*loves, being at Enderly Park*) as an IS-ABOUT relation and *happy* as an opinion expression associated with an implicit target. Table 1 shows some statistics of the corpus.

We adopted the evaluation metrics for entity and relation extraction from Choi et al. (2006), which include precision, recall, and F1-measure according to *overlap* and *exact* matching metrics.<sup>8</sup> We

<sup>6</sup>Available at <http://www.cs.pitt.edu/mpqa/>.

<sup>7</sup>349 news articles from the original MPQA corpus, 84 Wall Street Journal articles (Xbank), and 48 articles from the American National Corpus.

<sup>8</sup>Overlap matching considers two spans to match if they overlap, while exact matching requires two spans to be exactly the same.

	Opinion	Target	Holder
TotalNum	5849	4676	4244
	Opinion-arg Relations		Implicit Relations
IS-ABOUT	4823		1302
IS-FROM	4662		1187

Table 1: Data Statistics of the MPQA Corpus.

will focus our discussion on results obtained using overlap matching, since the exact boundaries of opinion entities are hard to define even for human annotators (Wiebe et al., 2005).

We trained CRFs for opinion entity identification using the following features: indicators for words, POS tags, and lexicon features (the subjectivity strength of the word in the Subjectivity Lexicon). All features are computed for the current token and tokens in a  $[-1, +1]$  window. We used L2-regularization; the regularization parameter was tuned using the development set. We trained the classifiers for relation extraction using L1-regularized logistic regression with default parameters using the LIBLINEAR (Fan et al., 2008) package. For joint inference, we used GLPK<sup>9</sup> to provide the optimal ILP solution. The parameter  $\lambda$  was tuned using the development set.

### 4.1 Baseline Methods

We compare our approach to several pipeline baselines. Each extracts opinion entities first using the same CRF employed in our approach, and then predicts opinion relations on the opinion entity candidates obtained from the CRF prediction. Three relation extraction techniques were used in the baselines:

- Adj: Inspired by the adjacency rule used in Hu and Liu (2004), it links each argument candidate to its nearest opinion candidate. Arguments that do not link to any opinion candidate are discarded. This is also used as a strong baseline in Choi et al. (2006).
- Syn: Links pairs of opinion and argument candidates that present prominent syntactic patterns. (We consider the syntactic patterns listed in Section 3.2.1.) Previous work also demonstrates the effectiveness of syntactic information in opinion extraction (Johansson and Moschitti, 2012).

<sup>9</sup><http://www.gnu.org/software/glpk/>

Method	Opinion Expression			Opinion Target			Opinion Holder		
	P	R	F1	P	R	F1	P	R	F1
CRF	82.21	66.15	73.31	73.22	48.58	58.41	72.32	49.09	58.48
CRF+Adj	82.21	66.15	73.31	80.87	42.31	55.56	75.24	48.48	58.97
CRF+Syn	82.21	66.15	73.31	81.87	30.36	44.29	78.97	40.20	53.28
CRF+RE	83.02	48.99	61.62	85.07	22.01	34.97	78.13	40.40	53.26
Joint-Model	71.16	77.85	<b>74.35*</b>	75.18	57.12	<b>64.92**</b>	67.01	66.46	<b>66.73**</b>
CRF	66.60	52.57	58.76	44.44	29.60	35.54	65.18	44.24	52.71
CRF+Adj	66.60	52.57	58.76	49.10	25.81	33.83	68.03	43.84	53.32
CRF+Syn	66.60	52.57	58.76	50.26	18.41	26.94	74.60	37.98	50.33
CRF+RE	69.27	40.09	50.79	60.45	15.37	24.51	75	38.79	51.13
Joint-Model	57.39	62.40	<b>59.79*</b>	49.15	38.33	<b>43.07**</b>	62.73	62.22	<b>62.47**</b>

Table 2: Performance on opinion entity extraction using *overlap* and *exact* matching metrics (the top table uses *overlap* and the bottom table uses *exact*). Two-tailed t-test results are shown on F1 measure for our method compared to the other baselines (statistical significance is indicated with \* ( $p < 0.05$ ), \*\* ( $p < 0.005$ )).

Method	IS-ABOUT			IS-FROM		
	P	R	F1	P	R	F1
CRF+Adj	73.65	37.34	49.55	70.22	41.58	52.23
CRF+Syn	76.21	28.28	41.25	77.48	36.63	49.74
CRF+RE	78.26	20.33	32.28	74.81	37.55	50.00
CRF+Adj-merged-10-best	25.05	61.18	35.55	30.28	62.82	40.87
CRF+Syn-merged-10-best	41.60	45.66	43.53	48.08	54.03	50.88
CRF+RE-merged-10-best	51.60	33.09	40.32	47.73	54.40	50.84
Joint-Model	64.38	51.20	<b>57.04**</b>	64.97	58.61	<b>61.63**</b>

Table 3: Performance on opinion relation extraction using the *overlap* metric.

- RE: Predicts opinion relations by employing the *opinion-arg* classifier and *opinion-implicit-arg* classifier. First, the *opinion-arg* classifier identifies pairs of opinion and argument candidates that form valid opinion relations, and then the *opinion-implicit-arg* classifier is used on the remaining opinion candidates to further identify opinion expressions without explicit arguments.

We report results using opinion entity candidates from the best CRF output and from the merged 10-best CRF output.<sup>10</sup> The motivation of merging the 10-best output is to increase recall for the pipeline methods.

## 5 Results

Table 2 shows the results of opinion entity identification using both *overlap* and *exact* metrics. We compare our approach with the pipeline baselines and CRF (the first step of the pipeline). We can see that our joint inference approach significantly outperforms all the baselines in F1 measure on extracting all types of opinion entities. In general,

<sup>10</sup>It is similar to the *merged 10-best* baseline in Choi et al. (2006). If an entity  $E_i$  extracted by the  $i$ th-best sequence overlaps with an entity  $E_j$  extracted by the  $j$ th-best sequence, where  $i \leq j$ , then we discard  $E_j$ . If  $E_i$  and  $E_j$  do not overlap, then we consider both entities.

by adding the relation extraction step, the pipeline baselines are able to improve precision over the CRF but fail at recall. CRF+Syn and CRF+Adj provide the same performance as CRF, since the relation extraction step only affects the results of opinion arguments. By incorporating syntactic information, CRF+Syn provides better precision than CRF+Adj on extracting arguments at the expense of recall. This indicates that using simple syntactic rules would mistakenly filter many correct relations. By using binary classifiers to predict relations, CRF+RE produces high precision on opinion and target extraction but also results in very low recall. Using the *exact* metric, we observe the same general trend in the results as the *overlap* metric. The scores are lower since the metric is much stricter.

Table 3 shows the results of opinion relation extraction using the *overlap* metric. We compare our approach with pipelined baselines in two settings: one employs relation extraction on 1-best output of CRF (top half of table) and the other employs the merged 10-best output of CRF (bottom half of table). We can see that in general, using merged 10-best CRF outputs boosts the recall while sacrificing precision. This is expected since merging the 10-best CRF outputs favors candidates that are

Method	IS-ABOUT Relation Extraction			IS-FROM Relation Extraction		
	P	R	F1	P	R	F1
ILP-W/O-ENTITY	49.10	40.48	44.38	44.77	58.24	50.63
ILP-W-SINGLE-RE	63.88	49.35	55.68	53.64	65.02	58.78
ILP-W/O-IMPLICIT-RE	62.00	44.73	51.97	73.23	51.28	60.32
Joint-Model	64.38	51.20	<b>57.04**</b>	64.97	58.61	<b>61.63*</b>

Table 4: Comparison between our approach and ILP baselines that omit some potentials in our approach.

believed to be more accurate by the CRF predictor. If CRF makes mistakes, the mistakes will propagate to the relation extraction step. The poor performance on precision further confirms the error propagation problem in the pipeline approaches. In contrast, our joint-inference method successfully boosts the recall while maintaining reasonable precision. This demonstrates that joint inference can effectively leverage the advantage of individual predictors and limit error propagation.

To demonstrate the effectiveness of different potentials in our joint inference model, we consider three variants of our ILP formulation that omit some potentials in the joint inference: one is ILP-W/O-ENTITY, which extracts opinion relations without integrating information from opinion entity identification; one is ILP-W-SINGLE-RE, which focuses on extracting a single opinion relation and ignores the information from the other relation; the third one is ILP-W/O-IMPLICIT-RE, which omits the potential for *opinion-implicit-arg* relation and assumes every opinion expression is linked to an explicit argument. The objective function of ILP-W/O-ENTITY can be represented as

$$\arg \max_u \sum_k \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{A}_k} r_{ij} u_{ij} \quad (2)$$

which is subject to constraints on  $u_{ij}$  to enforce relations to not overlap and limit the maximum number of relations that can be extracted for each opinion expression and each argument. For ILP-W-SINGLE-RE, we simply remove the variables associated with one opinion relation in the objective function (1) and constraints. The formulation of ILP-W/O-IMPLICIT-RE removes the variables associated with potential  $r_i$  in the objective function and corresponding constraints. It can be viewed as an extension to the ILP approach in Choi et al. (2006) that includes opinion targets and uses simpler ILP formulation with only one parameter and fewer binary variables and constraints to represent entity label assignments<sup>11</sup>.

<sup>11</sup>We compared the proposed ILP formulation with the ILP

Table 4 shows the results of these methods on opinion relation extraction. We can see that without the knowledge of the entity extractor, ILP-W/O-ENTITY provides poor performance on both relation extraction tasks. This confirms the effectiveness of leveraging knowledge from entity extractor and relation extractor. The improvement yielded by our approach over ILP-W-SINGLE-RE demonstrates the benefit of jointly optimizing different types of opinion relations. Our approach also outperforms ILP-W/O-IMPLICIT-RE, which does not take into account implicit relations. The results demonstrate that incorporating knowledge of implicit opinion relations is important.

## 6 Discussion

We note that the joint inference model yields a clear improvement on recall but not on precision compared to the CRF-based baselines. Analyzing the errors, we found that the joint model extracts comparable number of opinion entities compared to the gold standard, while the CRF-based baselines extract significantly fewer opinion entities (around 60% of the number of entities in the gold standard). With more extracted opinion entities, the precision is sacrificed but recall is boosted substantially, and overall we see an increase in F-measure. We also found that a good portion of errors were made because the generated candidates failed to cover the correct solutions. Recall that the joint model finds the global optimal solution over a set of opinion entity and relation candidates, which are obtained from the n-best CRF predictions and constituents in the parse tree that satisfy certain syntactic patterns. It is possible that the generated candidates do not contain the gold standard answers. For example, our model failed to identify the IS-ABOUT relation (*offers, general aid*) from the following sentence *Powell had contacted ... and received offers<sub>O1</sub> of [gen-*

formulation in Choi et al. (2006) on extracting opinion holders, opinion expressions and IS-FROM relations, and showed that the proposed ILP formulation performs better on all three extraction tasks.



eral aid] $T_1$ ... because both the CRF predictor and syntactic heuristics fail to capture (*offers, general aid*) as a potential relation candidate. By applying simple heuristics such as treating all verbs or verb phrases as opinion candidates would not help because it would introduce a large number of negative candidates and lower the accuracy of relation extraction (only 52% of the opinion expressions are verbs or verb phrases and 64% of the opinion targets are noun or noun phrases in the corpus we used). Therefore a more effective candidate generation method is needed to allow more candidates while limiting the number of negative candidates. We also observed incorrect parsing to be a cause of error. We hope to study ways to account for such errors in our approach as future work.

For computational time, our ILP formulation can be solved very efficiently using advanced ILP solvers. In our experiment, using GLPK's branch-and-cut solver took 0.2 seconds to produce optimal ILP solutions for 1000 sentences on a machine with Intel Core 2 Duo CPU and 4GB RAM.

## 7 Conclusion

In this paper we propose a joint inference approach for extracting opinion-related entities and opinion relations. We decompose the task into different subproblems, and jointly optimize them using constraints that aim to encourage their consistency and reduce prediction uncertainty. We show that our approach can effectively integrate knowledge from different predictors and achieve significant improvements in overall performance for opinion extraction. For future work, we plan to extend our model to handle more complex opinion relations, e.g. nesting or cross-sentential relations. This can be potentially addressed by incorporating more powerful predictors and more complex linguistic constraints.

## Acknowledgments

This work was supported in part by DARPA-BAA-12-47 DEFT grant 12475008 and NSF grant BCS-0904822. We thank Igor Labutov for helpful discussion and suggestions, Ainur Yessenalina for early discussion of the work, as well as the reviews for helpful comments.

## References

- E. Breck, Y. Choi, and C. Cardie. 2007. Identifying expressions of opinion in context. In *Proceedings of the 20th international joint conference on Artificial intelligence*, pages 2683–2688. Morgan Kaufmann Publishers Inc.
- Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 355–362. Association for Computational Linguistics.
- Y. Choi, E. Breck, and C. Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 431–439. Association for Computational Linguistics.
- D. Das, A.F.T. Martins, and N.A. Smith. 2012. An exact dual decomposition algorithm for shallow semantic parsing with constraints. *Proceedings of\* SEM.[ii, 10, 50]*.
- Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874.
- M. Hu and B. Liu. 2004. Mining opinion features in customer reviews. In *Proceedings of the National Conference on Artificial Intelligence*, pages 755–760. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
- Richard Johansson and Alessandro Moschitti. 2010a. Reranking models in fine-grained opinion analysis. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 519–527. Association for Computational Linguistics.
- Richard Johansson and Alessandro Moschitti. 2010b. Syntactic and semantic structure for opinion expression detection. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 67–76. Association for Computational Linguistics.
- Richard Johansson and Alessandro Moschitti. 2012. Relational features in fine-grained opinion analysis.
- 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*, pages 1–8. Association for Computational Linguistics.
- N. Kobayashi, K. Inui, and Y. Matsumoto. 2007. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural*

- Language Learning (EMNLP-CoNLL)*, pages 1065–1074.
- John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- K. Liu, L. Xu, and J. Zhao. 2012. Opinion target extraction using word-based translation model. In *Proceedings of the conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Bo Pang and Lillian Lee. 2008. *Opinion mining and sentiment analysis*. Now Pub.
- V. Punyakanok, D. Roth, and W. Yih. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2):257–287.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009. Expanding domain sentiment lexicon through double propagation. In *Proceedings of the 21st international joint conference on Artificial intelligence*, pages 1199–1204. Morgan Kaufmann Publishers Inc.
- G. Qiu, B. Liu, J. Bu, and C. Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27.
- R. Quirk, S. Greenbaum, G. Leech, J. Svartvik, and D. Crystal. 1985. *A comprehensive grammar of the English language*, volume 397. Cambridge Univ Press.
- D. Roth and W. Yih. 2004. *A linear programming formulation for global inference in natural language tasks*. Defense Technical Information Center.
- J. Ruppenhofer, S. Somasundaran, and J. Wiebe. 2008. Finding the sources and targets of subjective expressions. In *Proceedings of LREC*.
- Vivek Srikumar and Dan Roth. 2011. A joint model for extended semantic role labeling. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 129–139. Association for Computational Linguistics.
- V. Stoyanov and C. Cardie. 2008. Topic identification for fine-grained opinion analysis. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 817–824. Association for Computational Linguistics.
- J. Wiebe, T. Wilson, and C. Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2):165–210.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, 35(3):399–433.
- Theresa Wilson. 2008. *Fine-Grained Subjectivity Analysis*. Ph.D. thesis, Ph. D. thesis, University of Pittsburgh. Intelligent Systems Program.
- Y. Wu, Q. Zhang, X. Huang, and L. Wu. 2009. Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pages 1533–1541. Association for Computational Linguistics.
- B. Yang and C. Cardie. 2012. Extracting opinion expressions with semi-markov conditional random fields. In *Proceedings of the conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- 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*, pages 1462–1470. Association for Computational Linguistics.