Senti-LSSVM: Sentiment-Oriented Multi-Relation Extraction with Latent Structural SVM

Lizhen Qu
Max Planck Institute
for Informatics
lqu@mpi-inf.mpg.de

Yi Zhang
Nuance Communications
yi.zhang@nuance.com

Rui Wang
DFKI GmbH
mars198356@hotmail.com

Lili Jiang
Max Planck Institute
for Informatics
ljiang@mpi-inf.mpg.de

Rainer Gemulla
Max Planck Institute
for Informatics
rgemulla@mpi-inf.mpg.de

Gerhard Weikum
Max Planck Institute
for Informatics
weikum@mpi-inf.mpg.de

Abstract

Extracting instances of sentiment-oriented relations from user-generated web documents is important for online marketing analysis. Unlike previous work, we formulate this extraction task as a structured prediction problem and design the corresponding inference as an integer linear program. Our latent structural SVM based model can learn from training corpora that do not contain explicit annotations of sentiment-bearing expressions, and it can simultaneously recognize instances of both binary (polarity) and ternary (comparative) relations with regard to entity mentions of interest. The empirical evaluation shows that our approach significantly outperforms state-of-the-art systems across domains (cameras and movies) and across genres (reviews and forum posts). The gold standard corpus that we built will also be a valuable resource for the community.

1 Introduction

Sentiment-oriented relation extraction (Choi et al., 2006) is concerned with recognizing sentiment polarities and comparative relations between entities from natural language text. Identifying such relations often requires syntactic and semantic analysis at both sentence and phrase level. Most prior work on sentiment analysis consider either i) subjective sentence detection (Yu and Kübler, 2011), ii) polarity classification (Johansson and Moschitti, 2011; Wilson et al., 2005), or iii) comparative relation identification (Jindal and Liu, 2006; Ganapathibhotla and Liu, 2008). In practice, however, different types of sentiment-oriented relations frequently coexist in documents. In particular, we found that more than 38% of the sentences in our test corpus contain more than one type of relations. The isolated analysis approach is inappropriate because i) it sacrifices acuracy by ignoring the intricate interplay among different types of relations; ii) it could lead to conflicting predictions such as estimating a relation candidate as both negative and comparative. Therefore, in this paper, we identify instances of both sentiment polarities and comparative relations for entities of interest *simultaneously*. We assume that all the mentions of entities and attributes are given, and entities are disambiguated. It is a widely used assumption when evaluating a module in a pipeline system that the outputs of preceding modules are error-free.

To the best of our knowledge, the only existing system capable of extracting both comparisons and sentiment polarities is a rule-based system proposed by Ding et al. (2009). We argue that it is better to tackle the task by using a *unified model* with *structured outputs*. It allows us to consider a set of correlated relation instances *jointly* and characterize their interaction through a set of soft and hard constraints. For example, we can encode constraints to discourage an attribute to participate in a polarity relation and a comparative relation at the same time. As a result, the system extracts a set of correlated instances of sentiment-oriented relations from a given sentence. For example, with the sentence about the camera Canon 7D, *"The sensor is great, but the price is higher than Nikon D7000."* the expected output is *positive*(`Canon 7D`, *sensor*)

155

and *preferred*(`Nikon D7000`, `Canon 7D`, textit-price).

However, constructing a fully annotated training corpus for this task is labor-intensive and requires strong linguistic background. We minimize this overhead by applying a simplified annotation scheme, in which annotators mark mentions of entities and attributes, disambiguate the entities, and label instances of relations for each sentence. Based on the new scheme, we have created a small Sentiment Relation Graph (SRG) corpus for the domains of cameras and movies, which significantly differs from the corpora used in prior work (Wei and Gulla, 2010; Kessler et al., 2010; Toprak et al., 2010; Wiebe et al., 2005; Hu and Liu, 2004) in the following ways: i) both sentiment polarities and comparative relations are annotated; ii) all mentioned entities are disambiguated; and iii) no subjective expressions are annotated, unless they are part of entity mentions.

The new annotation scheme raises a new challenge for learning algorithms in that they need to automatically find textual evidences for each annotated relation during training. For example, with the sentence *"I like the Rebel a little better, but that is another price jump"*, simply assigning a sentiment-bearing expression to the nearest relation candidate is insufficient, especially when the sentiment is not explicitly expressed.

In this paper, we propose SENTI-LSSVM, a latent structural SVM based model for sentiment-oriented relation extraction. SENTI-LSSVM is applied to find the most likely set of the relation instances expressed in a given sentence, where the latent variables are used to assign the most appropriate textual evidences to the respective instances.

In summary, the contributions of this paper are the following:

- We propose SENTI-LSSVM: the first unified statistical model with the capability of extracting instances of both binary and ternary sentiment-oriented relations.

- We design a task-specific integer linear programming (ILP) formulation for inference.

- We construct a new SRG corpus as a valuable asset for the evaluation of sentiment relation extraction.

- We conduct extensive experiments with online reviews and forum posts, showing that SENTI-LSSVM model can effectively learn from a training corpus without explicitly annotated subjective expressions and that its performance significantly outperforms state-of-the-art systems.

2 Related Work

There are ample works on analyzing sentiment polarities and entity comparisons, but the majority of them studied the two tasks in isolation.

Most prior approaches for fine-grained sentiment analysis focus on polarity classification. Supervised approaches on expression-level analysis require the annotation of sentiment-bearing expressions as training data (Jin et al., 2009; Choi and Cardie, 2010; Johansson and Moschitti, 2011; Yessenalina and Cardie, 2011; Wei and Gulla, 2010). However, the corresponding annotation process is time-consuming. Although sentence-level annotations are easier to obtain, the analysis at this level cannot cope with sentences conveying relations of multiple types (McDonald et al., 2007; Täckström and McDonald, 2011; Socher et al., 2012). Lexicon-based approaches require no training data (Ku et al., 2006; Kim and Hovy, 2006; Godbole et al., 2007; Ding et al., 2008; Popescu and Etzioni, 2005; Liu et al., 2005) but suffer from inferior performance (Wilson et al., 2005; Qu et al., 2012). In contrast, our method requires no annotation of sentiment-bearing expressions for training and can predict both sentiment polarities and comparative relations.

Sentiment-oriented comparative relations have been studied in the context of user-generated discourse (Jindal and Liu, 2006; Ganapathibhotla and Liu, 2008). Approaches rely on linguistically motivated rules and assume the existence of independent keywords in sentences which indicate comparative relations. Therefore, these methods fall short of extracting comparative relations based on domain dependent information.

Both Johansson and Moschitti (2011) and Wu et al. (2011) formulate fine-grained sentiment analysis as a learning problem with structured outputs. However, they focus only on polarity classification

of expressions and require annotation of sentiment-bearing expressions for training as well.

While ILP has been previously applied for inference in sentiment analysis (Choi and Cardie, 2009; Somasundaran and Wiebe, 2009; Wu et al., 2011), our task requires a complete ILP reformulation due to 1) the absence of annotated sentiment expressions and 2) the constraints imposed by the joint extraction of both sentiment polarity and comparative relations.

3 System Overview

This section gives an overview of the whole system for extracting sentiment-oriented relation instances. Prior to presenting the system architecture, we introduce the essential concepts and the definitions of two kinds of directed hypergraphs as the representation of correlated relation instances extracted from sentences.

3.1 Concepts and Definitions

Entity. An *entity* is an abstract or concrete thing, which needs not be of material existence. An entity in this paper refers to either a product or a brand.

Attribute. An *attribute* is an object closely associated with or belonging to an entity, such as the lens of digital camera.

Sentiment-Oriented Relation. A *sentiment-oriented relation* is either a sentiment polarity or a comparative relation, defined on tuples of entities and attributes. A *sentiment polarity* relation conveys either a positive or a negative attitude towards entities or their attributes, whereas a *comparative relation* indicates the preference of one entity over the other entity w.r.t. an attribute.

Relation Instance. An instance of *sentiment polarity* takes the form r(entity, attribute) with $r \in$ {*positive, negative*}, such as *positive*(Canon 7D, *sensor*). The polarity instances expressed in the form of unary relations, such as "Nikon D7000 is excellent.", are denoted as binary relations r(entity, *whole*), where the attribute *whole* indicates the entity as a whole. In contrast, an instance of *comparative relation* is in the form of *preferred*{entity, entity, attribute}, e.g. *preferred*(Canon 7D, Nikon D7000, *price*). For brevity, we refer to an instance set of sentiment-oriented relations extracted from a

sentence as an SSOR. To represent the instances of the remaining relations, we represent them as *other*{entity, attribute}, such as textitpartOf{*wheel*, car}. These relations include objective relations and the subjective relations other than sentiment-oriented relations.

Mention-Based Relation Instances. A mention-based relation instance refers to a tuple of entity mentions with a certain relation. This concept is introduced as the representation of instances in a sentence by replacing entities with the corresponding entity mentions, such as *positive*("Canon SD880i", "wide angle view").
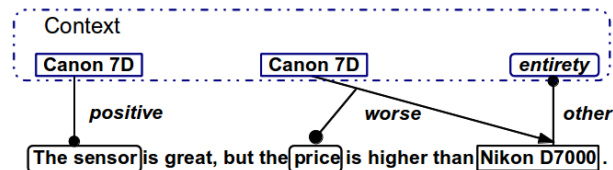

Figure 1: An example of MRG.

Mention-Based Relation Graph. A mention-based relation graph (or **MRG**) represents a collection of mention-based relation instances expressed in a sentence. As illustrated in Figure 1, an MRG is a directed hypergraph $G = \langle M, E \rangle$ with a vertex set M and an edge set E. A vertex $m_i \in M$ denotes a mention of an entity or an attribute occurring either within the sentence or in its context. We say that a mention is from the context if it is mentioned in the previous sentence or is an attribute implied in the current sentence. An instance of a binary relation in an MRG takes the form of a binary edge $e_l = (m_i, m_a)$, where m_i and m_a denote an entity mention and an attribute mention respectively, and the type $l \in$ {*positive, negative, other*}. A ternary edge e_l indicating comparative relation is represented as $e_l = (m_i, m_j, m_a)$, where two entity mentions m_i and m_j are compared with respect to the attribute mention m_a. We define the type $l \in$ {*better,worse*} to indicate two possible directions of the relation and assume m_i occurs before m_j. As a result, we have a set L of five *relation types*: *positive, negative, better, worse* or *other*. According to these definitions, the annotations in the SRG corpus are actually MRGs and disambiguated entities. If there are multiple mentions referring to the same entity, annotators are asked to choose the

most obvious one because it saves annotation time and is less demanding for the entity recognition and diambiguation modules.
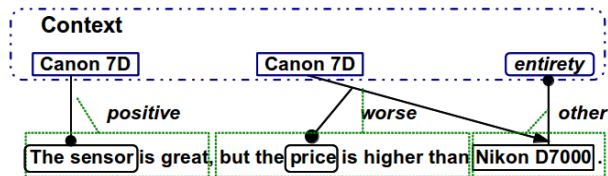

Figure 2: An example of eMRG. The textual evidences are wrapped by green dashed boxes.

Evidentiary Mention-Based Relation Graph. An *evidentiary* mention-based relation graph, coined **eMRG** , extends an MRG by associating each edge with a textual evidence to support the corresponding relation assertions (see Figure 2). Consequently, an edge in an eMRG is denoted by a pair (a, c), where a represents a mention-based relation instance and c is the associated textual evidence. It is also referred to as an *evidentiary* edge. represented as $e_l = (m_i, m_j, m_a)$, an MRG as an **evidentiary MRG** (**eMRG**) and the edges of eMRGs as **evidentiary** edges, as shown in Figure 2.
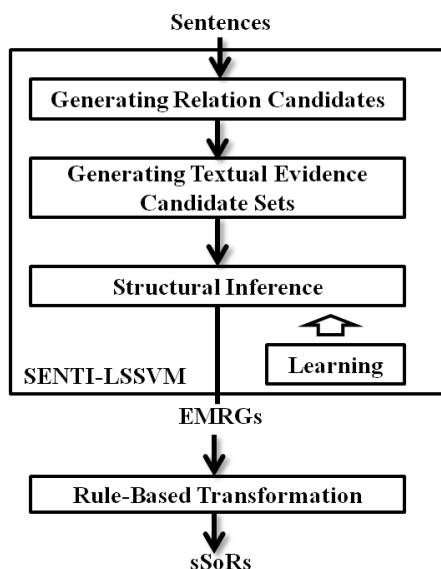
3.2 System Architecture

Figure 3: System architecture.

As illustrated by Figure 3, at the core of our system is the SENTI-LSSVM model, which extracts sets of mention-based relationships in the form of eMRGs from sentences. For a given sentence with known entity mentions, we select all possible mention sets as *relation candidates*, where each set includes at least one entity mention. Then we associate each relation candidate with a set of constituents or the whole sentence as the textual evidence candidates (cf. Section 6.1). Subsequently, the inference component aims to find the most likely eMRG from all possible combinations of mention-based relation instances and their textual evidences (cf. Section 6.2). The representation eMRG is chosen because it characterizes exactly the model outputs by letting each edge correspond to an instance of mention-based relation and the associated textual evidence. Finally, the model parameters of this model are learned by an online algorithm (cf. Section 7).

Since instance sets of sentiment-oriented relations (sSoRs) are the expected outputs, we can obtain sSoRs from MRGs by using a simple rule-based algorithm. The algorithm essentially maps the mentions from an MRG into entities and attributes in an sSoR and label the corresponding tuples with the relation types of the edges from an MRG. For instances of comparative relation, the label *better* or *worse* is mapped to the relation type *preferred*.

4   SENTI-LSSVM Model

The task of sentiment-oriented relation extraction is to determine the most likely sSoR in a sentence. Since sSoRs are derived from the corresponding MRGs as described in Section 3, the task is reduced to find the most likely MRG for each sentence. Since an MRG is created by assigning relation types to a subset of all *relation candidates*, which are possible tuples of mentions with unknown relation types, the number of MRGs can be extremely high.

To tackle the task, one solution is to employ an edge-factored linear model in the framework of structural SVM (Martins et al., 2009; Tsochantaridis et al., 2004). The model suggests that a bag of features should be specified for each relation candidate, and then the model predicts the most likely candidate sets along with their relation types to form the optimal MRGs. As we observed, for a relation candidate, the most informative features are the words near its entity mentions in the original text. How-

158

ever, if we represent a candidate by all these words, it is very likely that the instances of different relation types share overly similar features, because a mention is often involved in more than one relation candidate, as shown in Figure 2. As a consequence, the instances of different relations represented by overly similar features can easily confuse the learning algorithm. Thus, it is critical to select proper constituents or sentences as textual evidences for each relation candidate in both training and testing.

Consequently, we divide the task of sentiment-oriented relation extraction into two subtasks : i) identifying the most likely MRGs; ii) assigning proper textual evidences to each edge of MRGs to support their relation assertions. It is desirable to carry out the two subtasks *jointly* as these two subtasks could enhance each other. First, the identification of relation types requires proper textual evidences; second, the soft and hard constraints imposed by the correlated relation instances facilitate the recognition of the corresponding textual evidences. Since the eMRGs are created by attaching every MRG with a set of textual evidences, tackling the two subtasks simultaneously is equivalent to selecting the most likely eMRG from a set of eMRG candidates. It is challenging because our SRG corpus does not contain any annotation of textual evidences.

Formally, let $\mathcal{X}$ denote the set of all available sentences, and we define $y \in \mathcal{Y}(x)(x \in \mathcal{X})$ as the set of labeled edges of an MRG and $\mathcal{Y} = \cup_{x \in \mathcal{X}} \mathcal{Y}(x)$. Since the assignments of textual evidences are not observed, an assignment of evidences to y is denoted by a latent variable $h \in \mathcal{H}(x)$ and $\mathcal{H} = \cup_{x \in \mathcal{X}} \mathcal{H}(x)$. Then (y, h) corresponds to an eMRG, and $(a, c) \in (y, h)$ is a labeled edge a attached with a textual evidence c. Given a labeled dataset $D = \{(x_1, y_1), ..., (x_n, y_n)\} \in (\mathcal{X} \times \mathcal{Y})^n$, we aim to learn a discriminant function $f : \mathcal{X} \rightarrow \mathcal{Y} \times \mathcal{H}$ that outputs the optimal eMRG $(y, h) \in \mathcal{Y}(x) \times \mathcal{H}(x)$ for a given sentence x.

Due to the introduction of latent variables, we adopt the latent structural SVM (Yu and Joachims, 2009) for structural classification. Our discriminant function is defined as

$$f(x) = \text{argmax}_{(y,h) \in \mathcal{Y}(x) \times \mathcal{H}(x)} \boldsymbol{\beta}^\top \Phi(x, y, h) \quad (1)$$

where $\Phi(x, y, h)$ is the feature function of an eMRG (y, h) and $\boldsymbol{\beta}$ is the corresponding weight vector.

To ensure tractability, we also employ edge-based factorization for our model. Let M_p denote a set of entity mentions and $y_r(m_i)$ be a set of edges labeled with sentiment-oriented relations incident to m_i, the factorization of $\Phi(x, y, h)$ is given as

$$\Phi(x, y, h) \quad = \sum_{(a,c) \in (y,h)} \Phi_e(x, a, c) + \quad (2)$$
$$\sum_{m_i \in M_p} \sum_{a, a' \in y_r(m_i), a \neq a'} \Phi_c(a, a')$$

where $\Phi_e(x, a, c)$ is a local edge feature function for a labeled edge a attached with a textual evidence c and $\Phi_c(a, a')$ is a feature function capturing co-occurrence of two labeled edges a_{m_i} and a'_{m_i} incident to an entity mention m_i.

5 Feature Space

The following features are used in the feature functions (Equation 2):

Unigrams: As mentioned before, a textual evidence attached to an edge in MRG is either a word, phrase or sentence. We consider all lemmatized unigrams in the textual evidence as unigram features.

Context: Since web users usually express related sentiments about the same entity across sentence boundaries, we describe the sentiment flow using a set of contextual binary features. For example, if entity A is mentioned in both the previous sentence and the current sentence, a set of contextual binary features are used to indicate all possible combinations of the current and the previous mentioned sentiment-oriented relations regarding to entity A.

Co-occurrence: We have mentioned the co-occurrence feature in Equation 2, indicated by $\Phi_c(a, a')$. It captures the co-occurrence of two labeled edges incident to the same entity mention. Note that the co-occurrence feature function is considered only if there is a contrast conjunction such as *"but"* between the non-shared entity mentions incident to the two labeled edges.

Senti-predictors: Following the idea of (Qu et al., 2012), we encode the prediction results from the rule-based phrase-level multi-relation predictor (Ding et al., 2009) and from the bag-of-opinions predictor (Qu et al., 2010) as features based on the textual evidence. The output of the first predictor is an integer value, while the output of the second predictor is a sentiment relation, such as "positive",

"negative", "better" or "worse". We map the relational outputs into integer values and then encode the outputs from both predictors as senti-predictor features.

Others: The commonly used part-of-speech tags are also included as features. Moreover, for an edge candidate, a set of binary features are used to denote the types of the edge and its entity mentions. For instance, a binary feature indicates whether an edge is a binary edge related to an entity mentioned in context. To characterize the syntactic dependencies between two adjacent entity mentions, we use the path in the dependency tree between the heads of the corresponding constituents, the number of words and other mentions in-between as features. Additionally, if the textual evidence is a constituent, its feature w.r.t. an edge is the dependency path to the closest mention of the edge that does not overlap with this constituent.

6 Structural Inference

In order to find the best eMRG for a given sentence with a well trained model, we need to determine the most likely relation type for each relation candidate and support the corresponding assertions with proper textual evidences. We formulate this task as an Integer Linear Programming (ILP). Instead of considering all constituents of a sentence, we empirically select a subset as textual evidences for each relation candidate.

6.1 Textual Evidence Candidates Selection

Textual evidences are selected based on the constituent trees of sentences parsed by the Stanford parser (Klein and Manning, 2003). For each mention in a sentence, we first locate a constituent in the tree with the maximal overlap by Jaccard similarity. Starting from this constituent, we consider two types of candidates: *type I* candidates are constituents at the highest level which contain neither any word of another mention nor any contrast conjunctions such as *"but"*; *type II* candidates are constituents at the highest level which cover exactly two mentions of an edge and do not overlap with any other mentions. For a binary edge connecting an entity mention and an attribute mention, we consider a *type I* candidate starting from the attribute men-

tion. For a binary edge connecting two entity mentions, we consider *type I* candidates starting from both mentions. Moreover, for a comparative ternary edge, we consider both *type I* and *type II* candidates starting from the attribute mention. This strategy is based on our observation that these candidates often cover the most important information w.r.t. the covered entity mentions.

6.2 ILP Formulation

We formulate the inference problem of finding the best eMRG as an ILP problem due to its convenient integration of both soft and hard constraints.

Given the model parameters $\boldsymbol{\beta}$, we reformulate the score of an eMRG in the discriminant function (1) as follows,

$$\boldsymbol{\beta}^\top \Phi(x, y, h) = \sum_{(a,c) \in (y,h)} s_{ac} z_{ac} + \sum_{m_i \in M_p} \sum_{a,a' \in y_r(m_i), a \neq a'} s_{aa'} z_{aa'}$$

where $s_{ac} = \boldsymbol{\beta}^\top \Phi_e(x, a, c)$ denotes the score of a labeled edge a attached with a textual evidence c, $s_{aa'} = \boldsymbol{\beta}^\top \Phi_c(a, a')$ is the edge co-occurrence score, the binary variable z_{ac} indicates the presence or absence of the corresponding edge, and $z_{aa'}$ indicates if two edges co-occurr. As not every edge set can form an eMRG, we require that a valid eMRG should satisfy a set of linear constraints, which form our *constraint space*. Then function (1) is equivalent to

$$\begin{aligned} \max_{\mathbf{z} \in \mathbb{B}} \quad & \mathbf{s}^\top \mathbf{z} + \mu \mathbf{z}^d \\ \text{s.t.} \quad & \mathbf{A} \begin{bmatrix} \mathbf{z} \\ \boldsymbol{\eta} \\ \boldsymbol{\tau} \end{bmatrix} \leq \mathbf{d} \\ & \mathbf{z}, \boldsymbol{\eta}, \boldsymbol{\tau} \in \mathbb{B} \end{aligned}$$

where $\mathbb{B} = 2^S$ with $S = \{0, 1\}$, and $\boldsymbol{\eta}$ and $\boldsymbol{\tau}$ are auxiliary binary variables that help define the constraint space. The above optimization problem takes exactly the form of an ILP because both the constraints and the objective function are linear, and all variables take only integer values.

In the following, we consider two types of constraint space, 1) an eMRG with only binary edges and 2) an eMRG with both binary and ternary edges.

160

eMRG with only Binary Edges: An eMRG has only binary edges if a sentence contains no attribute mention or at most one entity mention. We expect that each edge has only one relation type and is supported by a single textual evidence. To facilitate the formulation of constraints, we introduce η_{e_l} to denote the presence or absence of a labeled edge e_l, and η_{ec} to indicate if a textual evidence c is assigned to an unlabeled edge e. Then the binary variable for the corresponding evidentiary edge $z_{e_lc} = \eta_{ec} \wedge \eta_{e_l}$, where the ILP formulation of conjunction can be found in (Martins et al., 2009).

Let C_e denote the set of textual evidence candidates of an unlabeled edge e. The constraint of *at most one textual evidence per edge* is formulated as:

$$\sum_{c \in C_e} \eta_{ec} \leq 1 \tag{3}$$

Once a textual evidence is assigned to an edge, their relation labels should match and the number of labeled edges must agree with the number of attached textual evidences. Further, we assume that a textual evidence c conveys at most one relation so that an evidence will not be assigned to the relations of different types, which is the main problem for the structural SVM based model. Let η_{cl} indicate that the textual evidence c is labeled by the relation type l. The corresponding constraints are expressed as,

$$\sum_{l \in L_e} \eta_{e_l} = \sum_{c \in C_e} \eta_{ec}; \quad z_{e_lc} \leq \eta_{cl}; \quad \sum_{l \in L} \eta_{cl} \leq 1$$

where L_e denotes the set of all possible labels for an unlabeled edge e, and L is the set of all relation types of MRGs (cf. Section 3).

In order to avoid a textual evidence being overly reused by multiple relation candidates, we first penalize the assignment of a textual evidence c to a labeled edge a by associating the corresponding z_{ac} with a fixed negative cost $-\mu$ in the objective function. Then the selection of one textual evidence per edge a is encouraged by associating μ to z_c^d in the objective function, where $z_c^d = \bigvee_{e \in S_c} \eta_{ec}$ and S_c is the set of edges that the textual evidence c serves as a candidate. The disjunction z_c^d is expressed as:

$$z_c^d \geq \eta_e, e \in S_c$$
$$z_c^d \leq \sum_{e \in S_c} \eta_e$$

(a) Binary edge structure
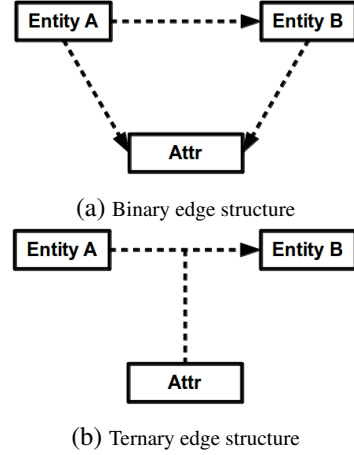

(b) Ternary edge structure

Figure 4: Alternative structures associated with an attribute mention.

This soft constraint not only encourages one textual evidence per edge, but also keeps it eligible for multiple assignments.

For any two labeled edge a and a' incident to the same entity mention, the edge-to-edge co-occurrence is described by $z_{a,a'}^c = z_a \wedge z_{a'}$.

eMRG with both Binary and Ternary Edges: If there are more than one entity mentions and at least one attribute mention in a sentence, an eMRG can potentially have both binary and ternary edges. In this case, we assume that each mention of attributes can participate either in binary relations or in ternary relations. The assumption holds in more than 99.9% of the sentences in our SRG corpus, thus we describe it as a set of hard constraints. Geometrically, the assumption can be visualized as the selection between two alternative structures incident to the same attribute mention, as shown in Figure 4. Note that, in the binary edge structure, we include not only the edges incident to the attribute mention but also the edge between the two entity mentions.

Let $S_{m_i}^b$ be the set of all possible labeled edges in a binary edge structure of an attribute mention m_i. Variable $\tau_{m_i}^b = \bigvee_{e_l \in S_{m_i}^b} \eta_{e_l}$ indicates whether the attribute mention is associated with a binary edge structure or not. In the same manner, we use $\tau_{m_i}^t = \bigvee_{e_l \in S_{m_i}^t} \eta_{e_l}$ to indicate the association of the an attribute mention m_i with an ternary edge structure from the set of all incident ternary edges $S_{m_i}^t$. The selection between two alternative structures is

formulated as $\tau_{m_i}^b + \tau_{m_i}^t = 1$. As this influences only the edges incident to an attribute mention, we keep all the constraints introduced in the previous section unchanged except for constraint (3), which is modified as

$$\sum_{c \in C_e} \eta_{ec} \leq \tau_{m_i}^b; \qquad \sum_{c \in C_e} \eta_{ec} \leq \tau_{m_i}^t$$

Therefore, we can have either binary edges or ternary edges for an attribute mention.

7 Learning Model Parameters

Given a set of training sentences $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, the best weight vector $\boldsymbol{\beta}$ of the discriminant function (1) is found by solving the following optimization problem:

$$\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} [\max_{(\hat{y}, \hat{h}) \in \mathcal{Y}(x) \times \mathcal{H}(x)} (\boldsymbol{\beta}^\top \Phi(x, \hat{y}, \hat{h}) + \delta(\hat{h}, \hat{y}, y))$$
$$- \max_{\bar{h} \in \mathcal{H}(x)} \boldsymbol{\beta}^\top \Phi(x, y, \bar{h})] + \rho |\boldsymbol{\beta}|] \quad (4)$$

where $\delta(\hat{h}, \hat{y}, y)$ is a loss function measuring the discrepancies between an eMRG $(y, \bar{h})$ with gold standard edge labels y and an eMRG $(\hat{y}, \hat{h})$ with inferred labeled edges $\hat{y}$ and textual evidences $\hat{h}$. Due to the sparse nature of the lexical features, we apply L1 regularizer to the weight vector $\boldsymbol{\beta}$, and the degree of sparsity is controlled by the hyperparameter ρ.

Since the L1 norm in the above optimization problem is not differentiable at zero, we apply the online *forward-backward splitting* (FOBOS) algorithm (Duchi and Singer, 2009). It requires two steps for updating the weight vector $\boldsymbol{\beta}$ by using a single training sentence x on each iteration t.

$$\boldsymbol{\beta}_{t+\frac{1}{2}} = \boldsymbol{\beta}_t - \varepsilon_t \Delta_t$$
$$\boldsymbol{\beta}_{t+1} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2} \|\boldsymbol{\beta} - \boldsymbol{\beta}_t\|^2 + \varepsilon_t \rho |\boldsymbol{\beta}|$$

where Δ_t is the subgradient computed without considering the L1 norm and ε_t is the learning rate. For a labeled sentence x, $\Delta_t = \Phi(x, \hat{y}^*, \hat{h}^*) - \Phi(x, y, \bar{h}^*)$, where the feature functions of the corresponding eMRGs are inferred by solving $(\hat{y}^*, \hat{h}^*) = \arg\max_{(\hat{h}, \hat{y}) \in \mathcal{H}(x) \times \mathcal{Y}(x)} [\boldsymbol{\beta}^\top \Phi(x, \hat{y}, \hat{h}) + \delta(\hat{h}, \hat{y}, y)]$ and $(y, \bar{h}^*) = \arg\max_{\bar{h} \in \mathcal{H}(x)} \boldsymbol{\beta}^\top \Phi(x, y, \bar{h})$, as indicated in the optimization problem (4).

The former inference problem is similar to the one we considered in the previous section except the inclusion of the loss function. We incorporate the loss function into the ILP formulation by defining the loss between an MRG (y, h) and a gold standard MRG as the sum of per-edge costs. In our experiments, we consider a positive cost φ for each wrongly labeled edge a, so that if an edge a has a different label from the gold standard, we add φ to the coefficient s_{ac} of the corresponding variable z_{ac} in the objective function of the ILP formulation.

In addition, since the non-positive weights of edge labels in the initial learning phrase often lead to eMRGs with many unlabeled edges, which harms the learning performance, we fix it by adding a constraint for the minimal number of labeled edges in an eMRG,

$$\sum_{a \in A} \sum_{c \in C_a} \eta_{ac} \geq \zeta \qquad (5)$$

where A is the set of all labeled edge candidates and ζ denotes the minimal number of labeled edges.

Empirically, the best way to determine ζ is to make it equal to the maximal number of labeled edges in an eMRG with the restriction that a textual evidence can be assigned to at most one edge. By considering all the edge candidates A and all the textual evidence candidates C as two vertex sets in a bipartite graph $\hat{G} = \langle V = (A, C), E \rangle$ (with edges in E indicating which textual evidence can be assigned to which edge), ζ corresponds to exactly the size of a maximum matching of the bipartite graph[1].

To find the optimal eMRG $(y, \bar{h}^*)$, for the gold label k of each edge, we consider the following set of constraints for inference since the labels of the edges are known for the training data,

$$\sum_{c \in C_e} \eta_{ec} \leq 1; \qquad \eta_{ec} \leq l_{ck}$$
$$\sum_{k' \in L} l_{ck'} \leq 1; \qquad \sum_{e \in S_c} \eta_{ec} \leq 1$$

We include also the soft constraints, which avoid a textual evidence being overly reused by multiple relations, and the constraints similar to (5) to ensure a minimal number of labeled edges and a minimal number of sentiment-oriented relations.

---

[1] It is computed by the Hopcroft-Karp algorithm (Hopcroft and Karp, 1973) in our implementation.

8 SRG Corpus

For evaluation we constructed the SRG corpus, which in total consists of 1686 manually annotated online reviews and forum posts in the digital camera and movie domains[2]. For each domain, we maintain a set of attributes and a list of entity names.

The annotation scheme for the sentiment representation asserts minimal linguistic knowledge from our annotators. By focusing on the meanings of the sentences, the annotators make decisions based on their language intuition, not restricted by specific syntactic structures. Taking the example in Figure 2, the annotators only need to mark the mentions of entities and attributes from both the sentences and the context, disambiguate them, and label ("Canon 7D", "Nikon D7000", *price*) as *worse* and ("Canon 7D", "sensor") as *positive*, whereas in prior work, people have annotated the sentiment-bearing expressions such as "great" and link them to the respective relation instances as well. This also enables them to annotate instances of both sentiment polarity and comparative relaton, which are conveyed by not only explicit sentiment-bearing expressions like *"excellent performance"*, but also factual expressions implying evaluations such as *"The 7V has 10x optical zoom and the 9V has 16x."*.

	Camera		Movie	
	Reviews	Forums	Reviews	Forums
positive	386	1539	879	905
negative	165	363	529	331
comparison	30	480	39	35

Table 1: Distribution of relation instances in SRG corpus.

14 annotators participated in the annotation project. After a short training period, annotators worked on randomly assigned documents one at a time. For product reviews, the system lists all relevant information about the entity and the predefined attributes. For forum posts, the system shows only the attribute list. For each sentence in a document, the annotator first determines if it refers to an entity of interest. If not, the sentence is marked

---

[2]The 107 camera reviews are from *bestbuy.com* and *Amazon.com*; the 667 camera forum posts are downloaded from *forum.digitalcamerareview.com*; the 138 movie reviews and 774 forum posts are from *imdb.com* and *boards.ie* respectively

as *off-topic*. Otherwise, the annotator will identify the most obvious mentions, disambiguate them, and mark the MRGs. We evaluate the inter-annotator agreement on SSoRs in terms of Cohen's Kappa (κ) (Cohen, 1968). An average Kappa value of 0.698 was achieved on a randomly selected set consisting of 412 sentences.

Table 1 shows the corpus distribution after normalizing them into SSoRs. Camera forum posts contain the largest proportion of comparisons because they are mainly about the recommendation of digital cameras. In contrast, web users are much less interested in comparing movies, in both reviews and forums. In all subsets, positive relations play a dominant role since web users intend to express more positive attitudes online than negative ones (Pang and Lee, 2007).

9 Experiments

This section describes the empirical evaluation of SENTI-LSSVM together with two competitive baselines on the SRG corpus.

9.1 Experimental Setup

We implemented a rule-based baseline (DING-RULE) and a structural SVM (Tsochantaridis et al., 2004) baseline (SENTI-SSVM) for comparison. The former system extends the work of Ding et al. (2009), which designed several linguistically-motivated rules based on a sentiment polarity lexicon for relation identification and assumes there is only one type of sentiment relation in a sentence. In our implementation, we keep all the rules of (Ding et al., 2009) and add one phrase-level rule when there are more than one mention in a sentence. The additional rule assigns sentiment-bearing words and negators to its nearest relation candidates based on the absolute surface distance between the words and the corresponding mentions. In this case, the phrase-level sentiment-oriented relations depend only on the assigned sentiment words and negators. The latter system is based on a structural SVM and does not consider the assignment of textual evidences to relation instances during inference. The textual features of a relation candidate are all lexical and sentiment predictor features within a surface distance of four words from the mentions of the candidate.

Thus, this baseline does not need the inference constraints of SENTI-LSSVM for the selection of textual evidences. To gain more insights into the model, we also evaluate the contribution of individual features of SENTI-LSSVM. In addition, to show if identifying sentiment polarities and comparative relations *jointly* works better than tackling each task on its own, we train SENTI-LSSVM for each task separately and combine their predictions according to compatibility rules and the corresponding graph scores.

For each domain and text genre, we withheld 15% documents for development and use the remaining for cross validation. The hyperparameters of all systems are tuned on the development datasets. For all experiments of SENTI-LSSVM, we use $\rho = 0.0001$ for the L1 regularizer in Eq.(4) and $\varphi = 0.05$ for the loss function; and for SENTI-SSVM, $\rho = 0.0001$ and $\varphi = 0.01$. Since the relation type of *off-topic* sentences is certainly *other*, we evaluate all systems with 5-fold cross-validation only on the *on-topic* sentences in the evaluation dataset. Since the same sSOR can have several equivalent MRGs and the relation type *other* is not of our interest, we evaluate the sSORs in terms of precision, recall and F-measure. All reported numbers are averages over the 5 folds.

9.2 Results

Table 2 shows the complete results of all systems. Here our model SENTI-LSSVM outperformed all baselines in terms of the average F-measure scores and recalls by a large margin. The F-measure on movie reviews is about 14% over the best baseline. The rule-based system has higher precision than recall in most cases. However, simply increasing the coverage of the domain independent sentiment polarity lexicon might lead to worse performance (Taboada et al., 2011) because many sentiment oriented relations are conveyed by domain dependent expressions and factual expressions implying evaluations, such as *"This camera does not have manual control."* Compared to DING-RULE, SENTI-SSVM performs better in the camera domain but worse for the movies due to many misclassification of *negative* relation instances as *other*. It also wrongly predicted more *positive* instances as *other* than SENTI-LSSVM. We found that the recalls of these instances are low because they often have overly similar features with the instances of the type

other linking to the same mentions. The problem gets worse in the movie domain since i) many sentences contain no explicit sentiment-bearing words; ii) the prior polarity of the sentiment-bearing words do not agree with their contextual polarity in the sentences. Consider the following example from a forum post about the movie *"Superman Returns"*: *"Have a look at Superman: the Animated Series or Justice League Unlimited . . . that is how the characters of Superman and Lex Luthor should be.".* In contrast, our model minimizes the overlapping features by assigning them to the most likely relation candidates. This leads to significantly better performance. Although SENTI-SSVM has low recall for both *positive* and *negative* relations, it achieves the highest recall for the comparative relation among all systems in the movie domain and camera reviews. Since less than 1% of all instances are for comparative relations in these document sets and all models are trained to optimize the overall accuracy, SENTI-LSSVM intends to trade off the minority class for the overall better performance. This advantage disappears on the camera forum posts, where the number of instances of comparative relation is 12 times more than that in the other data sets.

All systems perform better in predicting positive relations than the negative ones. This corresponds well to the empirical findings in (Wilson, 2008) that people intend to use more complex expressions for negative sentiments than their affirmative counterparts. It is also in accordance with the distribution of these relations in our SRG corpus which is randomly sampled from the online documents. For learning systems, it can also be explained by the fact that the training data for positive relations are considerably more than those for negative ones. The comparative relation is the hardest one to process since we found that many corresponding expressions do not contain explicit keywords for comparison.

To understand the performance of the key feature groups in our model better, we remove each group from the full SENTI-LSSVM system and evaluate the variations with movie reviews and camera forum posts, which have relatively balanced distribution of relation types. As shown in Table 3, the features from the sentiment predictors make significant contributions for both datasets. The different drops of the performance indicate that the po-

		Positive			Negative			Comparison			Micro-average		
		P	R	F	P	R	F	P	R	F	P	R	F
Camera Forum	Ding-Rule	56.4	39.0	46.1	46.2	24.0	31.6	42.6	14.0	21.0	53.4	30.8	39.0
	Senti-SSVM	60.2	35.6	44.8	44.2	38.5	41.2	28.0	**40.1**	32.9	43.7	36.7	39.9
	Senti-LSSVM	**69.2**	38.9	**49.8**	**50.8**	39.3	**44.3**	42.6	35.1	**38.5**	**56.5**	38.0	45.4
Camera Review	Ding-Rule	**83.6**	69.0	75.6	68.6	38.8	49.6	**30.0**	16.9	21.6	**81.1**	58.6	68.1
	Senti-SSVM	72.6	75.4	74.0	63.9	**62.5**	63.2	28.0	**38.9**	**32.5**	68.1	70.4	69.3
	Senti-LSSVM	77.3	**85.4**	**81.2**	68.9	61.3	**64.9**	22.3	20.7	21.6	73.1	**73.4**	**73.7**
Movie Forum	Ding-Rule	63.7	37.4	47.1	27.6	34.3	30.6	8.9	5.6	6.8	48.2	35.9	41.2
	Senti-SSVM	**66.2**	30.1	41.3	25.6	17.3	20.7	**44.2**	**56.7**	**49.7**	53.3	27.9	36.6
	Senti-LSSVM	63.3	**44.2**	**52.1**	29.7	45.6	36.0	40.1	45.0	42.4	49.7	**44.6**	**47.0**
Movie Review	Ding-Rule	**66.5**	47.2	55.2	42.0	39.1	40.5	**31.4**	12.0	17.4	56.2	44.0	49.4
	Senti-SSVM	61.3	54.0	57.4	45.2	13.7	21.1	24.5	**63.3**	**35.3**	54.6	39.2	45.7
	Senti-LSSVM	59.0	**79.1**	67.6	**53.3**	**51.4**	**52.3**	28.3	34.0	30.9	**57.9**	**68.8**	**62.9**

Table 2: Evaluation results for Ding-Rule, Senti-SSVM and Senti-LSSVM. Boldface figures are statistically significantly better than all others in the same comparison group under t-test with $p = 0.05$.

Feature Models	Movie Reviews	Camera Forums
full system	62.9	45.4
¬unigram	63.2 (+0.3)	41.2 (-4.2)
¬context	54.5 (-8.4)	46.0 (+0.6)
¬co-occurrence	62.6 (-0.3)	44.9 (-0.5)
¬senti-predictors	61.3 (-1.6)	34.3 (-11.1)

Table 3: Micro-average F-measure of Senti-LSSVM with different feature models

larities predicted by rules are more consistent in camera forum posts than in movie reviews. Due to the complexity of expressions in the movie reviews our model cannot benefit from the unigram features but these features are a good compensation for the sentiment predictor features in camera forum posts. The sharp drop by removing the context features from our model on movie reviews indicates that the sentiments in movie reviews depend highly on the relations of the previous sentences. In contrast, the sentiment-oriented relations of the previous sentences could be a reason of overfitting for camera forum data. The edge co-occurrence features do not play an important role in our model since the number of co-occurred sentiment-oriented relations in the sentences with contrast conjunctions like *"but"* is small. However, we found that allowing the co-occurrence of any sentiment-oriented relations would harm the performance of the model.

In addition, our experiments showed that the sep-arated approach, which trains a model for sentiment polarities and comparative relations respectively, leads to a decrease by almost 1% in terms of the F-measure averaged over all four datasets. The largest drop of F-measure is 3% on camera forum posts, since this dataset contains the largest proportion of comparative relations. We found that the errors are increased when the trained models make conflicting predictions. In this case, the joint approach can take all factors into account and make more consistent decisions than the separated approaches.

10 Conclusion

We proposed Senti-LSSVM model for extracting instances of both sentiment polarities and comparative relations. For evaluating and training the model, we created an SRG corpus by using a lightweight annotation scheme. We showed that our model can automatically find textual evidences to support its relation predictions and achieves significantly better F-measure scores than alternative state-of-the-art methods.

References

Yejin Choi and Claire Cardie. 2009. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural*

Language Processing: Volume 2 - Volume 2, EMNLP '09, pages 590–598, Stroudsburg, PA, USA. Association for Computational Linguistics.

Yejin Choi and Claire Cardie. 2010. Hierarchical sequential learning for extracting opinions and their attributes. In *Proceedings of the Annual meeting of the Association for Computational Linguistics*, pages 269–274. Association for Computational Linguistics.

Yejin Choi, Eric Breck, and Claire Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 431–439, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jacob Cohen. 1968. Weighted Kappa: Nominal Scale Agreement Provision for Scaled Disagreement or Partial Credit. *Psychological bulletin*, 70(4):213.

Xiaowen Ding, Bing Liu, and Philip S. Yu. 2008. A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 231–240, New York, NY, USA. ACM.

Xiaowen Ding, Bing Liu, and Lei Zhang. 2009. Entity discovery and assignment for opinion mining applications. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1125–1134.

John Duchi and Yoram Singer. 2009. Efficient online and batch learning using forward backward splitting. *The Journal of Machine Learning Research*, 10:2899–2934.

Murthy Ganapathibhotla and Bing Liu. 2008. Mining opinions in comparative sentences. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, pages 241–248, Stroudsburg, PA, USA. Association for Computational Linguistics.

Namrata Godbole, Manjunath Srinivasaiah, and Steven Skiena. 2007. Large-scale sentiment analysis for news and blogs (system demonstration). In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

John E Hopcroft and Richard M Karp. 1973. An nˆ5/2 algorithm for maximum matchings in bipartite graphs. *SIAM Journal on computing*, 2(4):225–231.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 168–177, New York, NY, USA. ACM.

Wei Jin, Hung Hay Ho, and Rohini K. Srihari. 2009. Opinionminer: a novel machine learning system for web opinion mining and extraction. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1195–1204, New York, NY, USA. ACM.

Nitin Jindal and Bing Liu. 2006. Mining comparative sentences and relations. In *Proceedings of the 21st International Conference on Artificial Intelligence - Volume 2*, AAAI'06, pages 1331–1336. AAAI Press.

Richard Johansson and Alessandro Moschitti. 2011. Extracting opinion expressions and their polarities–exploration of pipelines and joint models. In *Proceedings of the Annual meeting of the Association for Computational Linguistics*, volume 11, pages 101–106.

Jason S. Kessler, Miriam Eckert, Lyndsie Clark, and Nicolas Nicolov. 2010. The 2010 icwsm jdpa sentment corpus for the automotive domain. In *4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC 2010)*.

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*, SST '06, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 423–430, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lun-Wei Ku, Yu-Ting Liang, and Hsin-Hsi Chen. 2006. Opinion extraction, summarization and tracking in news and blog corpora. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pages 100–107.

Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pages 342–351, New York, NY, USA. ACM.

André L. Martins, Noah A. Smith, and Eric P. Xing. 2009. Concise integer linear programming formulations for dependency parsing. In *Proceedings of the Annual meeting of the Association for Computational Linguistics*, pages 342–350.

Ryan T. McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeffrey C. Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. In *Proceedings of the Annual meeting of the Association for Computational Linguistics*.

Bo Pang and Lillian Lee. 2007. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

166

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 339–346, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lizhen Qu, Georgiana Ifrim, and Gerhard Weikum. 2010. The bag-of-opinions method for review rating prediction from sparse text patterns. In Chu-Ren Huang and Dan Jurafsky, editors, *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, ACL Anthology, pages 913–921, Beijing, China. Tsinghua University Press.

Lizhen Qu, Rainer Gemulla, and Gerhard Weikum. 2012. A weakly supervised model for sentence-level semantic orientation analysis with multiple experts. In *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 149–159, Jeju Island, Korea, July. Proceedings of the Annual meeting of the Association for Computational Linguistics.

Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1201–1211.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, pages 226–234.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly D. Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.

Oscar Täckström and Ryan McDonald. 2011. Discovering fine-grained sentiment with latent variable structured prediction models. In *Proceedings of the 33rd European conference on Advances in information retrieval*, ECIR'11, pages 368–374, Berlin, Heidelberg. Springer-Verlag.

Cigdem Toprak, Niklas Jakob, and Iryna Gurevych. 2010. Sentence and expression level annotation of opinions in user-generated discourse. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 575–584, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. 2004. Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the International Conference on Machine Learning*, pages 104–112.

Wei Wei and Jon Atle Gulla. 2010. Sentiment learning on product reviews via sentiment ontology tree. In *Proceedings of the Annual meeting of the Association for Computational Linguistics*, pages 404–413.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 347–354, Stroudsburg, PA, USA. Association for Computational Linguistics.

Theresa Ann Wilson. 2008. *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. Ph.D. thesis, UNIVERSITY OF PITTSBURGH.

Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. 2011. Structural opinion mining for graph-based sentiment representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1332–1341.

Ainur Yessenalina and Claire Cardie. 2011. Compositional matrix-space models for sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 172–182.

Chun-Nam John Yu and Thorsten Joachims. 2009. Learning structural svms with latent variables. In *Proceedings of the International Conference on Machine Learning*, page 147.

Ning Yu and Sandra Kübler. 2011. Filling the gap: Semi-supervised learning for opinion detection across domains. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, pages 200–209. Association for Computational Linguistics.