Opinion Graphs for Polarity and Discourse Classification *

Swapna Somasundaran
Univ. of Pittsburgh
Pittsburgh, PA 15260
swapna@cs.pitt.edu

Galileo Namata
Univ. of Maryland
College Park, MD 20742
namatag@cs.umd.edu

Lise Getoor
Univ. of Maryland
College Park, MD 20742
getoor@cs.umd.edu

Janyce Wiebe
Univ. of Pittsburgh
Pittsburgh, PA 15260
wiebe@cs.pitt.edu

Abstract

This work shows how to construct discourse-level opinion graphs to perform a joint interpretation of opinions and discourse relations. Specifically, our opinion graphs enable us to factor in discourse information for polarity classification, and polarity information for discourse-link classification. This inter-dependent framework can be used to augment and improve the performance of local polarity and discourse-link classifiers.

1 Introduction

Much research in opinion analysis has focused on information from words, phrases and semantic orientation lexicons to perform sentiment classification. While these are vital for opinion analysis, they do not capture discourse-level associations that arise from relations between opinions. To capture this information, we propose discourse-level opinion graphs for classifying opinion polarity.

In order to build our computational model, we combine a linguistic scheme *opinion frames* (Somasundaran et al., 2008) with a collective classification framework (Bilgic et al., 2007). According to this scheme, two opinions are related in the discourse when their targets (what they are about) are related. Further, these pair-wise discourse-level relations between opinions are either *reinforcing* or *non-reinforcing* frames. Reinforcing frames capture reinforcing discourse scenarios where the individual opinions reinforce one another, contributing to the same opinion polarity or stance. Non-reinforcing frames, on the other hand, capture discourse scenarios where the individual opinions do not support the same stance. The individual opinion polarities and the type of relation between their targets determine whether the discourse frame is reinforcing or non-reinforcing.

Our polarity classifier begins with information from opinion lexicons to perform polarity classification locally at each node. It then uses discourse-level links, provided by the opinion frames, to transmit the polarity information between nodes. Thus the opinion classification of a node is not just dependent on its local features, but also on the class labels of related opinions and the nature of these links. We design two discourse-level link classifiers: the target-link classifier, which determines if a given node pair has unrelated targets (*no link*), or if their targets have a *same* or *alternative* relation, and the frame-link classifier, which determines if a given node pair has *no link*, *reinforcing* or *non-reinforcing* link relation. Both these classifiers too first start with local classifiers that use local information. The opinion graph then provides a means to factor in the related opinion information into the link classifiers. Our approach enables using the information in the nodes (and links) to establish or remove links in the graph. Thus information flows to and fro between all the opinion nodes and discourse-level links to achieve a joint inference.
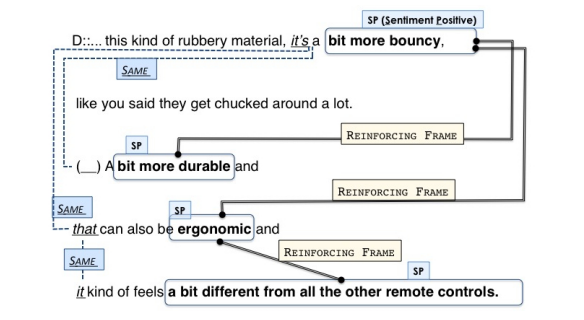
The paper is organized as follows: We first describe opinion graphs, a structure that can capture discourse-level opinion relationships in Section 2, and then describe our joint interpretation approach to opinion analysis in Section 3. Next, we describe our algorithm for joint interpretation in Section 4. Our experimental results are reported in Section 5. We discuss related work in Section 6 and conclude in Section 7.

2 Discourse-Level Opinion Graphs

The pairwise relationships that compose opinion frames can be used to construct a graph over opinion expressions in a discourse, which we refer to as the *discourse-level opinion graph (DLOG)*.

In this section, we describe these graphs and illustrate their applicability to goal-oriented multi-party conversations.

The nodes in the DLOG represent opinions, and there are two kinds of links: target links and frame links. Each opinion node has a polarity (*positive*, *negative* or *neutral*) and type (*sentiment* or *arguing*). *Sentiment* opinions are evaluations, feelings or judgments about the target. *Arguing* opinions argue for or against something. Target links are labeled as either *same* or *alternatives*. *Same* links hold between targets that refer to the same entity or proposition, while *alternative* links hold between targets that are related by virtue of being opposing (mutually exclusive) options in the context of the discourse. The frame links correspond to the *opinion frame* relation between opinions.

We illustrate the construction of the opinion graph with an example (Example 1, from Somasundaran et al. (2008)) from a multi-party meeting corpus where participants discuss and design a new TV remote control. The opinion expressions are in bold and their targets are in italics. Notice here that speaker D has a positive sentiment towards the rubbery material for the TV remote.

(1)   D:: ... this kind of rubbery material, *it's* a **bit more bouncy**, like you said they get chucked around a lot. A **bit more durable** and *that* can also be **ergonomic** and *it* kind of feels **a bit different from all the other remote controls**.

All the individual opinions in this example are essentially regarding the same thing – the rubbery material. The speaker's positive sentiment is apparent from the text spans **bit more bouncy**, **bit more durable**, **ergonomic** and **a bit different from all the other remote controls**. The explicit targets of these opinions (*it's*, *that*, and *it*) and the implicit target of "a bit more durable" are thus all linked with *same* relations.

Figure 1 illustrates the individual opinion annotations, target annotations (shown in italics) and

the relations between the targets (shown in dotted lines). Note that the target of **a bit more durable** is a zero span ellipsis that refers back to the rubbery material. The opinion frames resulting from the individual annotations make pairwise connections between opinion instances, as shown in bold lines in the figure. For example, the two opinions **bit more bouncy** and **ergonomic**, and the *same* link between their targets (*it's* and *that*), make up an opinion frame. An opinion frame type is derived from the details (type and polarity) of the opinions it relates and the target relation involved. Even though the different combinations of opinion type (sentiment and arguing), polarity (positive and negative) and target links (same and alternative) result in many distinct frames types (32 in total), they can be grouped, according to their discourse-level characteristics, into the two categories reinforcing and non-reinforcing. In this work, we only make this category distinction for opinion frames and the corresponding frame links.

The next example (Example 2, also from Somasundaran et al. (2008)) illustrates an *alternative* target relation. In the domain of TV remote controls, the set of all shapes are alternatives to one another, since a remote control may have only one shape at a time. In such scenarios, a positive opinion regarding one choice may imply a negative opinion toward competing choices, and vice versa. In this passage, speaker C's positive stance towards the curved shape is brought out even more strongly with his negative opinions toward the alternative, square-like, shapes.

(2)   C:: . . . shapes **should be** *curved*, so round shapes. **Nothing** *square-like*.
⋮
C:: . . . So we **shouldn't have too** *square corners* and that kind of thing.

The reinforcing frames characteristically show a reinforcement of an opinion or stance in the discourse. Both the examples presented above depict a reinforcing scenario. In the first example, the opinion towards the rubbery material is reinforced by repeated positive sentiments towards it, while in the second example the positive stance towards the curved shapes is further reinforced by negative opinions toward the alternative option. Examples of non-reinforcing scenarios are ambivalence between alternative options (for e.g., "I **like** *the rubbery material* but *the plastic* will be **much**

cheaper") or mixed opinions about the same target (for e.g., weighing pros and cons "*The rubbery material* is **good** but *it* will be **just so expensive**").

3   Interdependent Interpretation

Our interdependent interpretation in DLOGs is motivated by the observation that, when two opinions are related, a clear knowledge of the polarity of one of them makes interpreting the other much easier. For instance, suppose an opinion classifier wants to find the polarity of all the opinion expressions in Example 1. As a first step, it can look up opinion lexicons to infer that words like **"bouncy"**, **"durable"** and **" ergonomic"** are positive. However, **"a bit different "** cannot be resolved via this method, as its polarity can be different in different scenarios.

Suppose now we relate the targets of opinions. There are clues in the passage that the targets are related via the *same* relation; for instance they are all third person pronouns occurring in adjacent clauses and sentences. Once we relate the targets, the opinions of the passage are related via target links in the discourse opinion graph. We are also able to establish frames using the opinion information and target link information wherever they are available, i.e., a reinforcing link between **bit more bouncy** and **ergonomic**. For the places where all the information is not available (between **ergonomic** and **a bit different**) there are multiple possibilities. Depending on the polarity, either a reinforcing frame (if **a bit different** has positive polarity) or a non-reinforcing frame (if **a bit different** has negative polarity) can exist. There are clues in the discourse that this passage represents a reinforcing scenario. For instance there are reinforcing frames between the first few opinions, the repeated use of "and" indicates a *list, conjunction* or *expansion* relation between clauses (according to the Penn Discourse TreeBank (PDTB) (Prasad et al., 2008)), and there is a lack of contrastive clues that would indicate a change in the opinion. Thus the reinforcing frame link emerges as being the most likely candidate. This in turn disambiguates the polarity of **a bit different**. Thus, by establishing target links and frame links between the opinion instances, we are able to perform a joint interpretation of the opinions.

The interdependent framework of this example is iterative and dynamic — the information in the nodes can be used to change the structure (i.e.,

establish new links), and the structure provides a framework to change node polarity. We build our classification framework and feature sets with respect to this general framework, where the node labels as well as the structure of the graph are predicted in a joint manner.

Thus our interdependent interpretation framework has three main units: an instance polarity classifier (IPC), a target-link classifier (TLC), and a frame-link classifier (FLC). IPC classifies each node (instance), which may be a sentence, utterance or an other text span, as *positive*, *negative* or *neutral*. The TLC determines if a given node pair has related targets and whether they are linked by a *same* or *alternative* relation. The FLC determines if a given node pair is related via frames, and whether it is a reinforcing or non-reinforcing link. As we saw in the example, there are local clues available for each unit to arrive at its classification. The discourse augments this information to aid in further disambiguation.

4   Collective Classification Framework

For our collective classification framework, we use a variant of the iterative classification algorithm (ICA) proposed by Bilgic et al (2007). It combines several common prediction tasks in graphs: object classification (predicting the label of an object) and link prediction (predicting the existence and class of a link between objects). For our tasks, object classification directly corresponds to predicting opinion polarity and the link prediction corresponds to predicting the existence of a *same* or *alternative* target link or a *reinforcing* or *non-reinforcing* frame link between opinions. We note that given the nature of our problem formulation and approach, we use the terms link prediction and link classification interchangeably.

In the collective classification framework, there are two sets of features to use. The first are local features which can be generated for each object or link, independent of the links in which they participate, or the objects they connect. For example, the opinion instance may contain words that occur in sentiment lexicons. The local features are described in Section 4.2. The second set of features, the relational features, reflect neighborhood information in the graph. For frame link classification, for example, there is a feature indicating whether the connected nodes are predicted to have the same polarity. The relational features are de-

scribed in Section 4.3.

4.1 DLOG-ICA Algorithm

Our variant of the ICA algorithm begins by predicting the opinion polarity, and link type using only the local features. We then randomly order the set of all opinions and links and, in turn, predict the polarity or class using the local features and the values of the currently predicted relational features based on previous predictions. We repeat this until some stopping criterion is met. For our experiments, we use a fixed number of 30 iterations which was sufficient, in most of our datasets, for ICA to converge to a solution. The pseudocode for the algorithm is shown in Algorithm 4.1.

Algorithm 1 DLOG-ICA Algorithm

---

  for each opinion o **do** {bootstrapping}
    Compute polarity for o using local attributes
  end for
  for each target link t **do** {bootstrapping}
    Compute label for t using local attributes
  end for
  for each frame link f **do** {bootstrapping}
    Compute label for f using local attributes
  end for
  repeat {iterative classification}
    Generate ordering I over all nodes and links
    for each i in I **do**
      if i is an opinion instance **then**
        Compute polarity for i using local and relational attributes
      else if i is a target link **then**
        Compute class for i using local and relational attributes
      else if i is a frame link **then**
        Compute class for i using local and relational attributes
      end if
    end for
  until Stopping criterion is met

The algorithm is one very simple way of making classifications that are interdependent. Once the local and relational features are defined, a variety of classifiers can be used. For our experiments, we use SVMs. Additional details are provided in the experiments section.

4.2 Local Features

For the local polarity classifier, we employ opinion lexicons, dialog information, and unigram fea-

Feature	Task
Time difference between the node pair	TLC, FLC
Number of intervening instances	TLC, FLC
Content word overlap between the node pair	TLC,FLC
Focus space overlap between the node pair	TLC, FLC
Bigram overlap between the node pair *	TLC, FLC
Are both nodes from same speaker *	TLC, FLC
Bag of words for each node	TLC, FLC
Anaphoric indicator in the second node	TLC
Adjacency pair between the node pair	FLC
Discourse relation between node pair *	FLC

Table 1: Features and the classification task it is used for; TLC = target-link classification, FLC = Frame-link classification

tures. We use lexicons that have been successfully used in previous work (the polarity lexicon from (Wilson et al., 2005) and the arguing lexicon (Somasundaran et al., 2007)). Previous work used features based on parse trees, e.g., (Wilson et al., 2005; Kanayama and Nasukawa, 2006), but our data has very different characteristics from monologic texts – the utterances and sentences are much shorter, and there are frequent disfluencies, restarts, hedging and repetitions. Because of this, we cannot rely on parsing features. On the other hand, in this data, we have dialog act information[1] (Dialog Acts), which we can exploit. Note that the IPC uses only the Dialog Act tags (instance level tags like *Inform*, *Suggest*) and not the dialog structure information.

Opinion frame detection between sentences has been previously attempted (Somasundaran et al., 2008) by using features that capture discourse and dialog continuity. Even though our link classification tasks are not directly comparable (the previous work performs binary classification of frame-present/frame-absent between opinion bearing sentences, while this work performs three-way classification: no-link/reinforcing/non-reinforcing between DA pairs), we adapt the features for the link classification tasks addressed here. These features depend on properties of the nodes that the link connects. We also create some new features that capture discourse relations and lexical overlap.

Table 1 lists the link classification features. New features are indicated with a '*'. Continuous discourse indicators, like *time difference between the node pair* and *number of intervening instances* are useful for determining if the two nodes can be related. The *content word over-*

---

[1]Manual annotations for Dialog act tags and *adjacency pairs* are available for the AMI corpus.

lap, and *focus space overlap* features (the focus space for an instance is a list of the most recently used NP chunks; i.e., NP chunks in that instance and a few previous instances) capture the overlap in topicality within the node pair; while the *bigram overlap* feature captures the alignment between instances in terms of function words as well as content words. The entity-level relations are captured by the *anaphoric indicator feature* that checks for the presence of pronouns such as *it* and *that* in the second node in the node pair. The *adjacency pair* and *discourse relation* are actually feature sets that indicate specific dialog-structure and discourse-level relations. We group the list of discourse relations from the PDTB into the following sets: *expansion*, *contingency*, *alternative*, *temporal*, *comparison*. Each discourse relation in PDTB is associated with a list of discourse connective words.[2] Given a node pair, if the first word of the later instance (or the last word first instance) is a discourse connective word, then we assume that this node is connecting back (or forward) in the discourse and the feature set to which the connective belongs is set to true (e.g., if a latter instance is "because we should ...", it starts with the connective "because", and connects backwards via a *contingency* relation). The *adjacency pair* feature indicates the presence of a particular dialog structure (e.g., *support*, *positive-assessment*) between the nodes.

4.3 Relational Features

In addition to the local features, we introduce relational features (Table 2) that incorporate related class information as well as transfer label information between classifiers. As we saw in our example in Figure 1, we need to know not only the polarity of the related opinions, but also the type of the relation between them. For example, if the frame relation between **ergonomic** and **a bit different** is non-reinforcing, then the polarity of **a bit different** is likely to be negative. Thus link labels play an important role in disambiguating the polarity. Accordingly, our relational features transfer information of class labels from other instances of the same classifier as well as between different classifiers. Table 2 lists our relational features. Each row represents a set of features. Features are generated for all combinations of x, y and z for each

---

[2]The PDTB provides a list of discourse connectives and the list of discourse relations each connective signifies.

row. For example, one of the features in the first row is *Number of neighbors with polarity type positive, that are related via a reinforcing frame link*. Thus each feature for the polarity classifier identifies neighbors for a given node via a specific relation (z or y) and factors in their polarity values. Similarly, both link classifiers use polarity information of the node pair, and other link relations involving the nodes of the pair.

5 Evaluation

We experimentally test our hypothesis that discourse-level information is useful and non-redundant with local information. We also wanted to test how the DLOG performs for varying amounts of available annotations: from full neighborhood information to absolutely no neighborhood information.

Accordingly, for polarity classification, we implemented three scenarios: ICA-LinkNeigh, ICA-LinkOnly and ICA-noInfo. The ICA-LinkNeigh scenario measures the performance of the DLOG under ideal conditions (full neighborhood information) — the structure of the graph (link information) as well as the neighbors' class are provided (by an oracle). Here we do not need the TLC, or the FLC to predict links and the Instance Polarity Classifier (IPC) is not dependent on its predictions from the previous iteration. On the other hand, the ICA-noInfo scenario is the other extreme, and has absolutely no neighborhood information. Each node does not know which nodes in the network it is connected to apriori, and also has no information about the polarity of any other node in the network. Here, the structure of the graph, as well as the node classes, have to be inferred via the collective classification framework described in Sections 3 and 4. The ICA-LinkOnly is an intermediate condition, and is representative of scenarios where the discourse relationships between nodes is known. Here we start with the link information (from an oracle) and the IPC uses the collective classification framework to infer neighbor polarity information.

Similarly, we vary the amounts of neighborhood information for the TLC and FLC classifiers. In the ICA-LinkNeigh condition, TLC and FLC have full neighborhood information. In the ICA-noInfo condition, TLC and FLC are fully dependent on the classifications of the previous rounds. In the ICA-Partial condition, the TLC classifier

Feature
Opinion Polarity Classification
Number of neighbors with polarity type x linked via frame link z
Number of neighbors with polarity type x linked via target link y
Number of neighbors with polarity type x and same speaker linked via frame link z
Number of neighbors with polarity type x and same speaker linked via target link y
Target Link Classification
Polarity of the DA nodes
Number of other target links y involving the given DA nodes
Number of other target links y involving the given DA nodes and other same-speaker nodes
Presence of a frame link z between the nodes
Frame Link Classification
Polarity of the DA nodes
Number of other frame links z involving the given DA nodes
Number of other frame links z involving the given DA nodes and other same-speaker nodes
Presence of a target link y between the nodes

Table 2: Relational features: $x \in$ {non-neutral (i.e., positive or negative), positive, negative}, $y \in$ {same, alt}, $z \in$ {reinforcing, non-reinforcing}

uses true frame-links and polarity information, and previous-stage classifications for information about neighborhood target links; the FLC classifier uses true target-links and polarity information, and previous-stage classifications for information about neighborhood frame-links.

5.1 Data

For our experiments, we use the opinion frame annotations from previous work (Somasundaran et al., 2008). These annotations consist of the opinion spans that reveal opinions, their targets, the polarity information for opinions, the labeled links between the targets and the frame links between the opinions. The annotated data consists of 7 scenario-based, multi-party meetings from the AMI meeting corpus (Carletta et al., 2005). The manual Dialog Act (DA) annotations, provided by AMI, segment the meeting transcription into separate dialog acts. We use these DAs as nodes or instances in our opinion graph.

A DA is assigned the opinion orientation of the words it contains (for example, if a DA contains a positive opinion expression, then the DA assigned the positive opinion category). We filter out very small DAs (DAs with fewer than 3 tokens, punctuation included) in order to alleviate data skewness problem in the link classifiers. This gives us a total of 4606 DA instances, of which 1935 (42%) have opinions. Out of these 1935, 61.7% are positive, 30% are negative and the rest are neutral. The DAs that do not have opinions are considered neutral, and have no links in the DLOG. We create DA pairs by first ordering the DAs by their start time, and then pairing a DA with five DAs before it, and five DAs after it. The classes for target-link classification are *no-link, same, alt*. The gold standard target-link class is decided for a DA pair based on the target link between the targets of the opinions contained in that pair. Similarly, the labels for the frame-link labeling task are *no-link, reinforcing, non-reinforcing*. The gold standard frame link class is decided for a DA pair based on the frame between opinions contained by that pair. In our data, of the 4606 DAs, 1118 (24.27%) participate in target links with other DAs, and 1056 (22.9%) form frame links. The gold standard data for links, which has pair-wise information, has a total of 22,925 DA pairs, of which 1371 (6%) pairs have target links and 1264 (5.5%) pairs have frame links.

We perform 7-fold cross-validation experiments, using the 7 meetings. In each fold, 6 meetings are used for training and one meeting is used for testing.

5.2 Classifiers

Our baseline (Base) classifies the test data based on the distribution of the classes in the training data. Note that due to the heavily skewed nature of our link data, this classifier performs very poorly for minority class prediction, even though it may achieve good overall accuracy.

For our local classifiers, we used the classifiers from the Weka toolkit (Witten and Frank, 2002). For opinion polarity, we used the Weka's SVM implementation. For the target link and frame link classes, the huge class skew caused SVM to learn a trivial model and always predict the majority class. To address this, we used a cost sensitive classifier in Weka where we set the cost of misclassifying a less frequent class, A, to a more frequent class, B,

	Base	Local	ICA		
			LinkNeigh	LinkOnly	noInfo
Acc	45.9	68.7	78.8	72.9	68.4
Class: neutral (majority class)					
Prec	61.2	76.3	83.9	78.2	73.5
Rec	61.5	83.9	89.6	89.1	86.6
F1	61.1	79.6	86.6	83.2	79.3
Class: positive polarity					
Prec	26.3	56.2	70.9	63.3	57.6
Rec	26.1	46.6	62.0	47.0	42.8
F1	25.8	50.4	65.9	53.5	48.5
Class: negative polarity					
Prec	12.4	52.3	64.6	56.3	55.2
Rec	12.2	44.3	60.2	48.2	38.2
F1	12.2	46.0	61.9	51.2	43.9

Table 3: Performance of Polarity Classifiers

	Base	Local	ICA		
			LinkNeigh	Partial	noInfo
TLC					
Acc	88.5	85.8	98.1	98.2	86.3
P-M	33.3	35.9	76.1	76.1	36.3
R-M	33.3	38.1	78.1	78.1	38.1
F1-M	33.1	36.0	74.6	74.6	36.5
FLC					
Acc	89.3	86.2	98.9	98.9	87.6
P-M	33.3	36.9	81.3	82.8	38.0
R-M	33.4	41.2	82.2	84.4	41.7
F1-M	33.1	37.2	80.7	82.3	38.1

Table 4: Performance of Link Classifiers

as $|B|/|A|$ where $|class|$ is the size of the class in the training set. All other misclassification costs are set to 1.

For our collective classification, we use the above classifiers for local features (l) and use similar, separate classifiers for relational features (r). For example, we learned an SVM for predicting opinion polarity using only the local features and learned another SVM using only relational features. For the ICA-noInfo condition, where we use TLC and FLC classifiers, we combine the predictions using a weighted combination where $P(class|l,r) = \alpha * P(class|l) + (1 - \alpha) * P(class|r)$. This allows us to vary the influence each feature set has to the overall prediction. The results for ICA-noInfo are reported on the best performing α (0.7).

5.3 Results

Our polarity classification results are presented in Table 3, specifically accuracy (Acc), precision (Prec), recall (Rec) and F-measure (F1). As we can see, the results are mixed. First, we notice that the Local classifier shows substantial improvement over the baseline classifier. This shows that the lexical and dialog features we use are informative of opinion polarity in multi-party meetings.

Next, notice that the ICA-LinkNeigh classifier performs substantially better than the Local classifier for all metrics and all classes. The accuracy improves by 10 percentage points, while the F-measure improves by about 15 percentage points for the minority (positive and negative) classes. This result confirms that our discourse-level opinion graphs are useful and discourse-level information is non-redundant with lexical and dialog-act

information.

The results for ICA-LinkOnly follow the same trend as for ICA-LinkNeigh, with a 3 to 5 percentage point improvement. These results show that even when the neighbors' classes are not known a priori, joint inference using discourse-level relations helps reduce errors from local classification.

However, the performance of the ICA-noInfo system, which is given absolutely no starting information, is comparable to the Local classifier for the overall accuracy and F-measure metrics for the neutral class. There is slight improvement in precision for both the positive and negative classes, but there is a drop in their recall. The reason this classifier does no better than the Local classifier is because the link classifiers TLC and FLC predict "none" predominantly due to the heavy class skew.

The performance of the link classifiers are reported in Table 4, specifically the accuracy (Acc) and macro averages over all classes for precision (P-M), recall (R-M) and F-measure (F1-M). Due to the heavy skew in the data, accuracy of all classifiers is high; however, the macro F-measure, which depends on the F1 of the minority classes, is poor for the ICA-noInfo. Note, however, that when we provide some (Partial) or full (LinkNeigh) neighborhood information for the Link classifiers, the performance of these classifiers improve considerably. This overall observed trend is similar to that observed with the polarity classifiers.

6 Related Work

Previous work on polarity disambiguation has used contextual clues and reversal words (Wilson et al., 2005; Kennedy and Inkpen, 2006; Kanayama and Nasukawa, 2006; Devitt and Ahmad, 2007; Sadamitsu et al., 2008). However, these do not capture discourse-level relations.

Polanyi and Zaenen (2006) observe that a central topic may be divided into subtopics in order to perform evaluations. Similar to Somasundaran et al. (2008), Asher et al. (2008) advocate a discourse-level analysis in order to get a deeper understanding of contextual polarity and the strength of opinions. However, these works do not provide an implementation for their insights. In this work we demonstrate a concrete way that discourse-level interpretation can improve recognition of individual opinions and their polarities.

Graph-based approaches for joint inference in sentiment analysis have been explored previously by many researchers. The biggest difference between this work and theirs is in what the links represent linguistically. Some of these are not related to discourse at all (e.g., lexical similarities (Takamura et al., 2007), morphosyntactic similarities (Popescu and Etzioni, 2005) and word based measures like TF-IDF (Goldberg and Zhu, 2006)). Some of these work on sentence cohesion (Pang and Lee, 2004) or agreement/disagreement between speakers (Thomas et al., 2006; Bansal et al., 2008). Our model is not based on sentence cohesion or structural adjacency. The relations due to the opinion frames are based on relationships between targets and discourse-level functions of opinions being mutually reinforcing or non-reinforcing. Adjacent instances need not be related via opinion frames, while long distant relations can be present if opinion targets are same or alternatives. Also, previous efforts in graph-based joint inference in opinion analysis has been text-based, while our work is over multi-party conversations.

McDonald et al. (2007) propose a joint model for sentiment classification based on relations defined by granularity (sentence and document). Snyder and Barzilay (2007) combine an agreement model based on contrastive RST relations with a local aspect (topic) model. Their aspects would be related as same and their high contrast relations would correspond to (a subset of) the non-reinforcing frames.

In the field of product review mining, sentiments and features (aspects or targets) have been mined (for example, Yi et al. (2003), Popescu and Etzioni (2005), and Hu and Liu (2006)). More recently there has been work on creating joint models of topic and sentiments (Mei et al., 2007; Titov and McDonald, 2008) to improve topic-sentiment summaries. We do not model topics; instead we directly model the relations between targets. The focus of our work is to jointly model opinion polarities via target relations. The task of finding co-referent opinion topics by (Stoyanov and Cardie, 2008) is similar to our target link classification task, and we use somewhat similar features. Even though their genre is different, we plan to experiment with their full feature set for improving our TLC system.

Turning to collective classification, there have been various collective classification frameworks proposed (for example, Neville and Jensen (2000), Lu and Getoor (2003), Taskar et al. (2004), Richardson and Domingos (2006)). In this paper, we use an approach proposed by (Bilgic et al., 2007) which iteratively predicts class and link existence using local classifiers. Other joint models used in sentiment classification include the spin model (Takamura et al., 2007), relaxation labeling (Popescu and Etzioni, 2005), and label propagation (Goldberg and Zhu, 2006).

7 Conclusion

This work uses an opinion graph framework, DLOG, to create an interdependent classification of polarity and discourse relations. We employed this graph to augment lexicon-based methods to improve polarity classification. We found that polarity classification in multi-party conversations benefits from opinion lexicons, unigram and dialog-act information. We found that the DLOGs are valuable for further improving polarity classification, even with partial neighborhood information. Our experiments showed three to five percentage points improvement in F-measure with link information, and 15 percentage point improvement with full neighborhood information. These results show that lexical and discourse information are non-redundant for polarity classification, and our DLOG, that employs both, improves performance.

We discovered that link classification is a difficult problem. Here again, we found that by using the DLOG framework, and using even partial neighborhood information, improvements can be achieved.

References

N. Asher, F. Benamara, and Y. Mathieu. 2008. Distilling opinion in discourse: A preliminary study.

COLING-2008.

M. Bansal, C. Cardie, and L. Lee. 2008. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In *COLING-2008*.

M. Bilgic, G. M. Namata, and L. Getoor. 2007. Combining collective classification and link prediction. In *Workshop on Mining Graphs and Complex Structures at the IEEE International Conference on Data Mining*.

J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, I. McCowan, W. Post, D. Reidsma, and P. Wellner. 2005. The AMI Meetings Corpus. In *Proceedings of the Measuring Behavior Symposium on "Annotating and measuring Meeting Behavior"*.

A. Devitt and K. Ahmad. 2007. Sentiment polarity identification in financial news: A cohesion-based approach. In *ACL 2007*.

A. B. Goldberg and X. Zhu. 2006. Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In *HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing*.

M. Hu and B. Liu. 2006. Opinion extraction and summarization on the Web. In *21st National Conference on Artificial Intelligence (AAAI-2006)*.

H. Kanayama and T. Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *EMNLP-2006*, pages 355–363, Sydney, Australia.

A. Kennedy and D. Inkpen. 2006. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125.

Q. Lu and L. Getoor. 2003. Link-based classification. In *Proceedings of the International Conference on Machine Learning (ICML)*.

R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. In *ACL 2007*.

Q. Mei, X. Ling, M. Wondra, H. Su, and C Zhai. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *WWW '07*. ACM.

J. Neville and D. Jensen. 2000. Iterative classification in relational data. In *In Proc. AAAI-2000 Workshop on Learning Statistical Models from Relational Data*, pages 13–20. AAAI Press.

B. Pang and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *ACl 2004*.

L. Polanyi and A. Zaenen, 2006. *Contextual Valence Shifters*. Computing Attitude and Affect in Text: Theory and Applications.

A.-M. Popescu and O. Etzioni. 2005. Extracting product features and opinions from reviews. In *HLT-EMNLP 2005*.

R. Prasad, A. Lee, N. Dinesh, E. Miltsakaki, G. Campion, A. Joshi, and B. Webber. 2008. Penn discourse treebank version 2.0. Linguistic Data Consortium.

M. Richardson and P. Domingos. 2006. Markov logic networks. *Mach. Learn.*, 62(1-2):107–136.

K. Sadamitsu, S. Sekine, and M. Yamamoto. 2008. Sentiment analysis based on probabilistic models using inter-sentence information. In *LREC'08*.

B. Snyder and R. Barzilay. 2007. Multiple aspect ranking using the good grief algorithm. In *HLT 2007: NAACL*.

S. Somasundaran, J. Ruppenhofer, and J. Wiebe. 2007. Detecting arguing and sentiment in meetings. In *SIGdial Workshop on Discourse and Dialogue 2007*.

S. Somasundaran, J. Wiebe, and J. Ruppenhofer. 2008. Discourse level opinion interpretation. In *Coling 2008*.

V. Stoyanov and C. Cardie. 2008. Topic identification for fine-grained opinion analysis. In *Coling 2008*.

H. Takamura, T. Inui, and M. Okumura. 2007. Extracting semantic orientations of phrases from dictionary. In *HLT-NAACL 2007*.

B. Taskar, M. Wong, P. Abbeel, and D. Koller. 2004. Link prediction in relational data. In *Neural Information Processing Systems*.

M. Thomas, B. Pang, and L. Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *EMNLP 2006*.

I. Titov and R. McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *ACL 2008*.

T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT-EMNLP 2005*.

I. H. Witten and E. Frank. 2002. Data mining: practical machine learning tools and techniques with java implementations. *SIGMOD Rec.*, 31(1):76–77.

J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *ICDM-2003*.