Author-Specific Sentiment Aggregation for Polarity Prediction of Reviews

Subhabrata Mukherjee and Sachindra Joshi
Max-Planck-Institut für Informatik, Saarbrücken, Germany
IBM Research, India
smukherjee@mpi-inf.mpg.de, jsachind@in.ibm.com

Abstract

In this work, we propose an *author-specific* sentiment aggregation model for polarity prediction of reviews using an ontology. We propose an approach to construct a *Phrase annotated Author specific Sentiment Ontology Tree (PASOT)*, where the facet nodes are annotated with opinion phrases of the author, used to describe the facets, as well as the author's preference for the facets. We show that an author-specific aggregation of sentiment over an ontology fares better than a flat classification model, which does not take the domain-specific facet importance or author-specific facet preference into account. We compare our approach to supervised classification using Support Vector Machines, as well as other baselines from previous works, where we achieve an accuracy improvement of 7.55% over the SVM baseline. Furthermore, we also show the effectiveness of our approach in capturing *thwarting* in reviews, achieving an accuracy improvement of 11.53% over the SVM baseline.

Keywords: Sentiment Ontology Tree, Author-Specific Facet Preference, Sentiment Aggregation

1. Introduction

In recent times there has been an explosion in the volume of data in the web. With the advent of blogs, micro-blogs, online review sites *etc.* there is a huge surge of interest in mining these information sources for popular opinions. Sentiment analysis aims to analyze text to find the user opinion about a given product or its different facets.

The earlier works (Pang and Lee, 2002; Pang and Lee, 2004; Turney, 2002) in sentiment analysis considered a review as a bag-of-words, where the different topics or facets of a product were ignored. The more recent works (Lin and He, 2009; Wang et al., 2011; Mukherjee and Bhattacharyya, 2012a; Mukherjee et al., 2014) consider a review as a bag-of-facets, and use approaches like dependency parsing, topic models to extract feature-specific expressions of opinion. However, the association between the facets influencing the review polarity has been largely ignored. Although these works extract the feature-specific polarities, they do not give any systematic approach to aggregate those polarities to obtain the overall review polarity. For example, consider the following review from IMDB:

"the acting performance in the movie is mediocre. the characters are thin and replaceabale. it has such common figures that it would not have suffered much with a lesser talented cast. it is likely that those pouring into the theater are going to be those anxious to partake of tarantino's quirky dialogue and eccentric directing style. it's good, but it is not anything that made pulp fiction such a revolutionary effort. this is a more conservative tarantino, but not one that will not satiate true fans." ... (1)

A flat classification model considering all features to be equally important will fail to capture the positive polarity of this review, as there are more negative feature polarities than positive ones. The reviewer seems to be impressed with "tarantino's direction style" and "quirky dialogue". However, the "character roles, acting performance, cast" seem to disappoint him. The overall review polarity is positive as the reviewer expresses positive opinion about the director and the movie as a whole. If we consider an ontology tree for the movie, then it can be observed that the positive polarity of the facets higher up in the tree dominate the negative ones at a lower level.

Now, consider the above review from the point of view of different users. Some may prefer the "character aspects" in the movie over the "director". Such users may consider the above review to be negative. Hence, the polarity of the above review will differ for users having varying facet-specific preferences. The affective polarity of phrases also depend on the authors. For example, the affective value of "mediocre" refering to the "acting performance" will have a different affective polarity for different reviewers. The sentiment aggregation approach over the ontology, thus, should not only capture the domain-specific importance of the facet, given by its depth in the ontology tree, but also the author-specific preference for the facet.

In this work, we show that an author-specific sentiment aggregation over the ontology fares better than the generic sentiment aggregation, which is a global model capturing only popular facet opinions. We propose an approach to construct a *Phrase annotated Author specific Sentiment Ontology Tree (PASOT)*, where each facet node of the domain-specific product ontology is annotated with opinion phrases in the review pertaining to that facet, extracted using Dependency Parsing. Given a review, we map it to the ontology using a WordNet based similarity measure. Thereafter, we propose a learning algorithm to find the node polarities and aggregate them bottom-up to find the overall review polarity. In the process, we learn the ontology weights on a *per-author* basis, where the node weights in the ontology tree capture the author-specific preference as well as the domain-specific importance of the facet.

The rest of the paper is organized as follows: In Section 2., we describe an approach to create the phrase annotated autho specific sentiment ontology tree. Section 3. discusses the algorithm to learn the ontology weights on a per-author basis and perform a bottom-up sentiment aggregation over the tree to find the overall review polarity. We present the experimental evaluation of the model on the IMDB movie review dataset in Section 4.. We also present an interesting

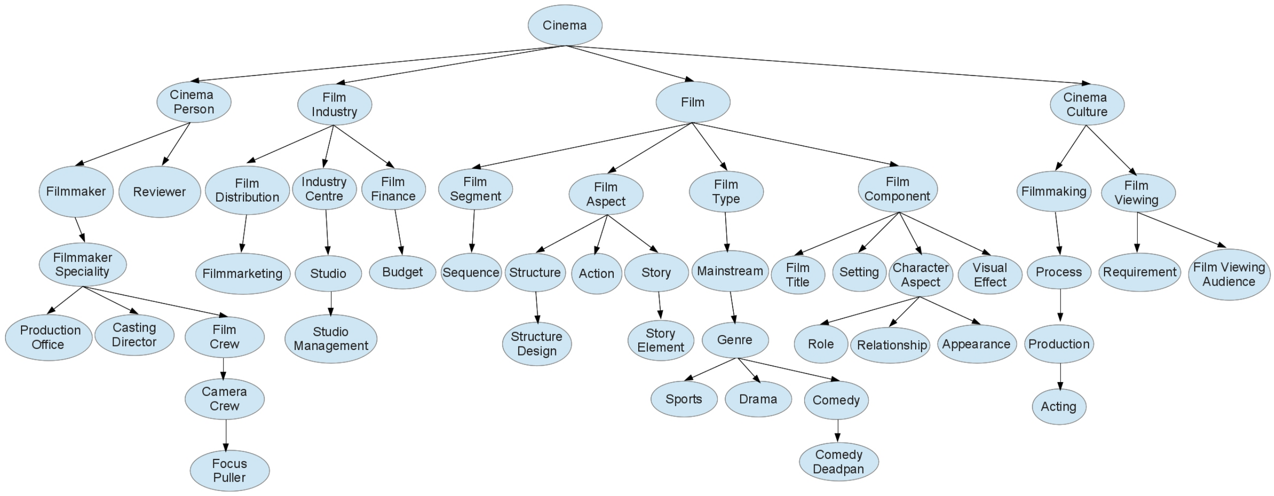

Figure 1: Snapshot of Cinema Ontology Tree

use-case to detect *thwarting* in reviews using our approach. Related work is discussed in Section 5., followed by conclusions.

2. Phrase Annotated Author Specific Sentiment Ontology Tree

An ontology can be viewed as a data structure that specifies terms, their properties and relations among them for a richer knowledge representation. A domain-specific ontology tree consists of domain-specific concepts (E.g. 'movie', 'direction', 'actor', 'editor' *etc.* are concepts in the movie domain) and relations between the concepts (E.g. "movie has_a actor", "actor has_a acting_performance", "movie has_a editorial_department", "editorial_department has_a colorist" *etc.*).

Consider the following review from IMDB:

"as with any gen-x mtv movie (like last year's dead man on campus), the movie is marketed for a primarily male audience as indicated by its main selling points: sex and football. those two items are sure to snare a sizeable box office chunk initially, but sales will decline for two reasons. first, the football sequences are nothing new; the sports genre isn't mainstream and it's been retread to death. second, the sex is just bad. despite the appearance of a whipped cream bikini or the all-night strip-club party, there's nothing even remotely tantalizing. the acting is mostly mediocre, not including the fantastic jon voight. cultivating his usual sliminess, voight gives an unexpectedly standout performance as west canaan coyotes head coach bud kilmer ... these elements (as well as the heavy drinking and carousing) might be more appropriate on a college campus – but mtv's core audience is the high school demographic. this focus is further emphasized by the casting: james van der beek, of tv' s "dawson's creek", is an understandable choice for the reluctant hero..." ... (2)

Figure 1 shows a snapshot of a movie domain ontology tree for Review 2.. Only the facets which are present in the review are shown in the ontology.

2.1. Sentiment Ontology Tree (SOT)

A sentiment ontology tree has been used in (Wei and Gulla, 2010; Mukherjee and Joshi, 2013) for capturing facet-specific sentiments in a domain. A Sentiment Ontology Tree (SOT) bears all the facets or concepts in a given domain as nodes, with edges between nodes capturing the relationship between the facets. For a given review, the nodes are annotated with polarities which represent the review polarity with respect to the facet. The tree captures componential relationship between the product features in a given domain (E.g. "movie has_a producer", "film_aspect has_a story" *etc.*), and how the children facet polarities come together to influence the parent facet polarity. Figure 2 shows a snapshot of the sentiment ontology tree for Review 2.. It shows the review polarity to be positive with respect to "acting performance, box office, casting" *etc.*, and negative with respect to "film character appearance, film setting, structure design" *etc.* and the overall movie.

2.2. Phrase Annotated Sentiment Ontology Tree (PSOT)

A review may consist of many facets with varying opinions about each facet. Even a single review sentence can bear varying opinions about different facets, like *"The acting was fine in the movie but the direction was mediocre"*. Here, the polarity with respect to 'acting' is positive and that with respect to 'direction' is negative. Hence, an SOT considering the sentence as a whole will assign a neutral polarity to both nodes 'actor' and 'director'.

In our previous work (Mukherjee and Bhattacharyya, 2012a), we used a dependency parsing based feature-specific sentiment extraction approach to evaluate the polarity of a sentence with respect to a given facet. Dependency parsing captures the association between any specific feature and the expressions of opinion that come together to describe that feature. A set of significant dependency parsing relations (like "nsubj, dobj, advmod, amod" *etc.*) are used to capture important associations between words in the review, followed by clustering to retrieve words associated to the target feature.

Consider a review r consisting of $< s_i >$ sentences, and $< f_j >$ facets. Let p_i^j be the phrase in the i^{th} sentence associated to the j^{th} facet as given by the above dependency parsing algorithm. In the phrase annotated SOT, we associate each node f_j to all the phrases $< p_i^j >$ associated to it in the review, that are extracted by the dependency parser. Figure 2 shows a snapshot of the phrase annotated sentiment ontology tree (PSOT) for Review 2..

2.3. Phrase Annotated Author Specific Sentiment Ontology Tree (PASOT)

For the same review, different authors may give a different rating to it depending on their topic and facet preferences. The overall rating of Review 2. depends on the taste of the reviewer, and other author-specific properties like gender, age, locale *etc.*. A reviewer who is a fan of "Jon Voight" would probably give it a positive rating for his performance, whereas others would mostly find the "acting" mediocre and hence assign a negative rating to the movie. Similarly, teenagers and male audience may be wooed by the main selling points of the movie *i.e.* "sex and football", whereas mature audience would not be impressed by them.

In order to capture the taste of a reviewer, each node f_j of the phrase annotated SOT (PSOT) is further annotated with the author-specific facet preference w_j. This is a personalized PSOT whose annotations differ across reviewers.

In this author-specific PSOT (PASOT), the sentiment annotation of each facet would also differ across reviewers. For example, consider the node 'actor' in the above review, and the associated phrase "acting mostly mediocre" given by dependency parsing. The polarity of this phrase depends on the expectations of a reviewer from a movie. Figure 3 shows a snapshot of the phrase annotated sentiment ontology tree for a given reviewer for Review 2..

2.4. Ontology Tree Construction

In our earlier work (Mukherjee and Joshi, 2013), we had leveraged ConceptNet (Liu and Singh, 2004) to create a domain-specific ontology tree by categorizing its relations into 3 classes namely, *Hierarchical* (E.g. "LocatedNear, HasA, PartOf, MadeOf"), *Synonymous* (E.g. "Synonym, IsA, ConceptuallyRelatedTo, InheritsFrom") and *Functional* (E.g. "UsedFor, CapableOf, HasProperty, DefinedAs"). ConceptNet is a very large semantic network of common sense knowledge constructed using crowdsourcing, which also incorporates noise in the network. We proposed an algorithm to recursively construct an ontology tree by grounding it on the hierarchical relations.

In absence of a semantic knowledge-base to tap into, we proposed (Mukherjee et al., 2014) an approach to construct a domain-specific ontology for the smartphone domain by considering 4 primary relations namely, *Type-Of, Synonymous, Action-On* and *Function-Of.* We leveraged the English Slot Grammar Parser and *Shallow Semantic Relationship Annotation* built over the parser output, in conjunction with the *Hearst patterns* and *Random Indexing*, built on the *Relational Distributional Similarity* hypothesis.

In this work, we make use of an available manually constructed ontology from the *cinema* domain (JedFilm, 2014). It was constructed using representative sampling and a multi-phased procedure. The ontology is based on a purposive sampling of document types produced by the film community. The document subjects are films, randomly-sampled from a large selection of films considered as important by critics and directors. Purposive sampling selects units for analysis based upon judgment about their usefulness in representing the overall population.

The domain concepts (nouns or noun phrases) are stored as Protégé (Protégé, 2014) "classes", and categorized hierarchically (top-down) within four main branches ("cinema culture, cinema person, filmmaking, film industry"). The "attribute, example, synonym" and "relation" terms are represented as Protégé "slots", associated to the concept terms.

2.5. Mapping of Review to the SOT

Given a review, we need to map the words in the review to the constructed SOT. As the review may contain concepts not present in the ontology but synonymous to some of the nodes, we use a WordNet-based similarity measure for the relatedness of two concepts. The Wu-Palmer measure (Wu and Palmer, 1994) calculates relatedness between two concepts by considering their depths in the WordNet taxonomy, along with the depth of their Lowest Common Subsumer (LCS). The Wu-Palmer similarity between two concepts $s1$ and $s2$ is given by $\frac{2\times depth(lcs)}{(depth(s1)+depth(s2)}$. The concept is ignored if the similarity score is less than a threshold.

3. Author Specific Sentiment Aggregation over Ontology

Consider a review r consisting of $< s_i >$ sentences, and $< f_j >$ facets. Let p_i^j be the phrase in the i^{th} sentence associated to the j^{th} facet as given by the feature-specific dependency parsing algorithm in (Mukherjee and Bhattacharyya, 2012a). Consider the phrase annotated sentiment ontology tree $T(V, E)$, where V is a product attribute set represented by the tuple $V_j =< f_j, < p_i^j >, w_j, d_j >$, where f_j is a product facet, w_j is the author-specific facet preference and d_j is the depth of the product attribute in the ontology tree. $E_{j,k}$ is an attribute relation connecting V_j and V_k. Let $V_{j,k}$ be the k^{th} child of V_j. Consider a sentiment predictor function $O(p)$ that finds and maps the polarity of a phrase to $[-1, 1]$.

The author-specific $(PSOT)$ is now equipped with $T^a(V, E)$ and $O^a(p)$ for a given author a.

The expected sentiment weight (ESW) of a node in the $PASOT$ is defined as,

$$ESW^a(V_j) = w_j^a \times \frac{1}{d_j} \times \sum_i O^a(p_i^j) + \sum_k ESW^a(V_{j,k}) \quad (1)$$

where $O^a(p_i^j) \in [-1, 1]$

The expected sentiment weight measures the weighted polarity of a node, taking its self-weight and children weights into consideration. The self-weight of a node is given by the sum of polarities of all the phrases in the review bearing an opinion about the facet associated with the node, weighed by the author preference for the facet and inverse of its depth in the ontology tree. The closer a facet is to the root of the tree, the more important it is to the SOT.

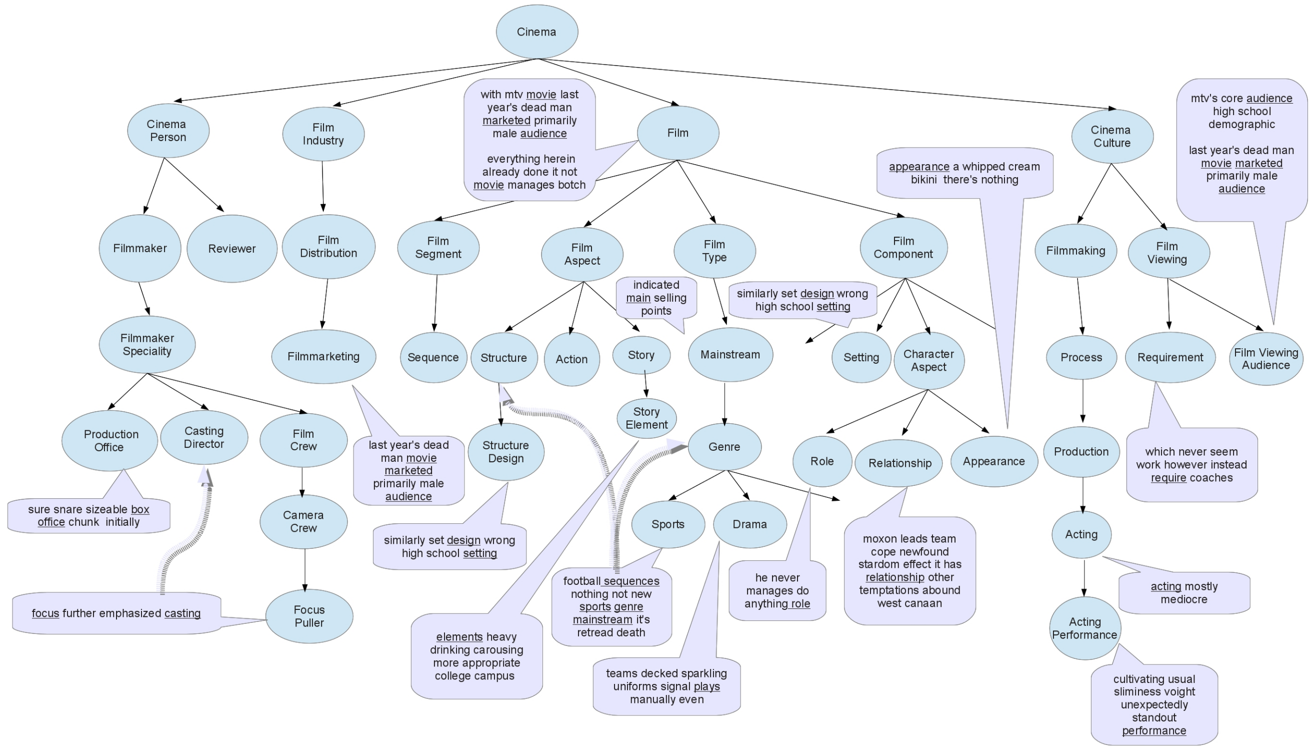

Figure 2: Snapshot of Phrase Annotated Cinema Ontology Tree for Review 2.

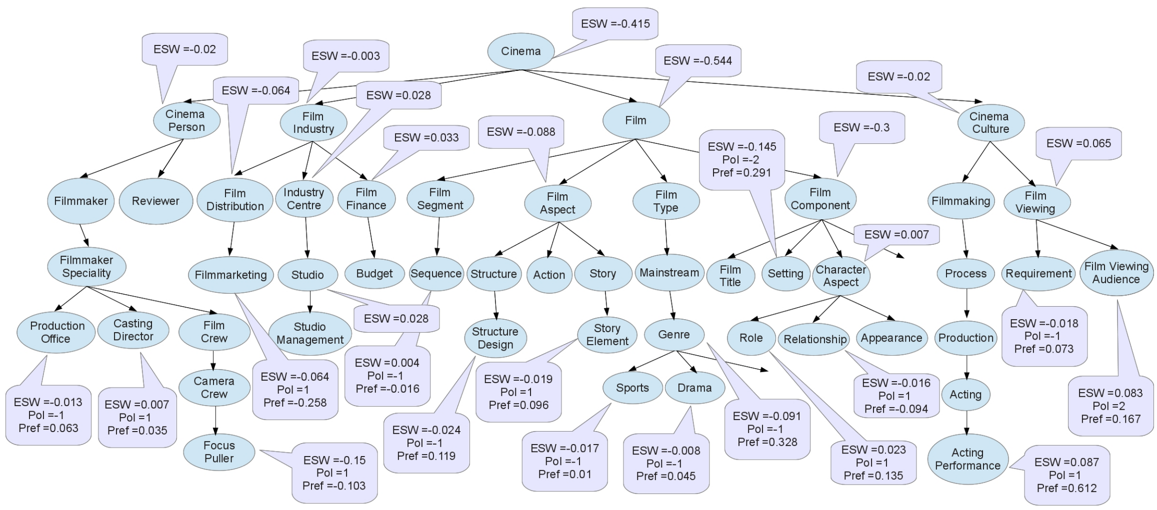

Figure 3: Snapshot of Author-Specific Phrase Annotated Cinema Ontology Tree for Review 2.

Facet importance decreases with increase in distance from the root as it becomes more fine-grained.

The review polarity is given by the expected sentiment-weight (ESW) of the tree given by $ESW^a(root)$. The computation of ESW of the root requires learning of the weights $< w_j^a >$ and the sentiment predictor function O^a for each author a.

80% of the reviews for each author is used for training parameters, and remaining 20% reviews are used for testing.

In absence of labeled training reviews, the sentiment predictor function $O(p)$ can be implemented using a sentiment lexicon that looks up the polarity of the words in a phrase, assigning the majority polarity to the associated node. A supervised computation of this function, requires many reviews *per-author*. Since the IMDB dataset has much less number of reviews per-author, we settle for a global sentiment predictor function by using an L_2-regularized L_2-loss Support Vector Machine and bag-of-words unigram features trained over the movie review corpus in (Maas et al., 2011). This means, for sentiment annotation of the opinion phrases associated to the facet nodes, we consider the general polarity of the opinion phrase. In the earlier example for "the acting is mediocre", a negative polarity is assigned to the facet "actor" irrespective of the author. Supervised classification using SVM's (Pang and Lee, 2002; Pang and Lee, 2004; Mullen and Collier, 2004) is also going to be one of the baselines for this work.

For each author a, every facet f_j is associated with an expected sentiment weight $ESW^a(V_j)$, where $f_j \in V_j$, that encapsulates the self-importance of the facet as well as the weight of its children. In order to learn the author-specific facet preference for each node, the weights $< w_j^a >$ in Equation 1 are initially set to 1, and the expected sentiment weight $ESW^a(V_j)$ of all the nodes are computed. For each review, let y_i be the labeled polarity of a review in the training set for each author. Thereafter, we formulate an L_2-regularized logistic regression problem to find the author-specific weight of each node as follows :

$$\min_{w^a} \frac{1}{2} w^{aT} w^a + C \sum_i \log(1 + \exp^{-y_i \sum_j w_j^a \times ESW^a(V_j)}) \tag{2}$$

Trust region newton method (Lin et al., 2008) is used to learn the weights in the above equation, using an implementation of LibLinear (Fan et al., 2008). After learning the author-specific facet weights, the polarity of an unseen review (given its author) is computed using Equation 1 and $ESW^a(root)$. Figure 3 shows a snapshot of the learnt $PASOT$ for Review 2.. Algorithm 1 gives an overview of the review classification process.

4. Experimental Evaluation

We evaluate the effectiveness of the author-specific sentiment aggregation approach using a phrase annotated sentiment ontology tree over the benchmark IMDB movie review dataset introduced in (Pang and Lee, 2002). Table 1 shows the data statistics.

4.1. Dataset Pre-Processing

The movie review dataset contains 2000 reviews and 312 authors with at least 1 review per author. In order to have sufficient data per author, we retained only those authors with at least 10 reviews. This reduced the number of reviews to 1467 with 65 authors. The number of reviews for the 2 ratings (pos and neg) is balanced in this dataset.

All the words are lemmatized in the reviews so that 'movie' and 'movies' are reduced to the same root word 'movie'. Words like "hvnt, dnt, cnt, shant" *etc.* are replaced with their proper form in both our model and the baselines to capture the negation.

4.2. Baselines

We consider three baselines in this work to judge the effectiveness of our approach.

The *first baseline* is the widely used supervised classification (Pang and Lee, 2002; Pang and Lee, 2004; Mullen and Collier, 2004) using Support Vector Machines with L_2-loss, L_2-regularizer and unigram bag-of-words features.

The *second baseline* is considered to be our earlier author-specific facet preference work in restaurant reviews (Mukherjee et al., 2013). The work considers manually given seed facets like 'food', 'ambience', 'service' *etc.* and uses dependency parsing with a sentiment lexicon (Hu and Liu, 2004) to find the sentiment about each facet. A WordNet similarity metric (Wu and Palmer, 1994) is used

Data: Review Dataset R and its Authorset A
Result: Review Polarities as $+1$ or -1
1. Learn the domain-specific ontology $T(V, E)$ using a knowledge-base (JedFilm, 2014)
2. Learn a global polarity predictor function $O(p)$ over review dataset (Maas et al., 2011) using L_2-regularized L_2-loss SVM
for *each author* $a \in A$ **do**
    for *each review* r *written by* a **do**
        for *each sentence* s *in* r **do**
            for *each word* f *in* s **do**
                Map it to $T(V, E)$ using Wu-Palmer Similarity
                if $f \in V$ **then**
                    1. Use Feature-Specific Dependency Parsing Algorithm (Mukherjee and Bhattacharyya, 2012a) to extract the phrase p_s^f from s that expresses the reviewer opinion about f
                    2. Annotate $V \in T(V, E)$ with p_s^f
                end
            end
        end
        Apply the predictor function $O(p)$ to each $p_s^f \in V$ and annotate the nodes with polarities
    end
    1. Apply Equation 1 to the PSOT bottom-up to find ESW of each node V using Equation 1, with w^a initialized to 1.
    2. Using 80% of the labeled review data (y_i) for a and $< ESW^a(V_j) >$, learn the facet-weights $< w_j^a >$ using Equation 2
    for *each unseen review* r *written by* a **do**
        1. Construct $PASOT$ using the above steps and learnt weights w^a
        2. Use Equation 1 to find $< ESW^a(V_j) >$
        3. Review polarity is given by $Sign(ESW^a(root))$
    end
end

Algorithm 1: Author-Specific Hierarchical Sentiment Aggregation for Review Polarity Prediction

to assign each facet to a seed facet. Thereafter, we used linear regression to learn author preference for the seed facets from review ratings. In this baseline, there is no notion of a domain ontology or hierarchical aggregation.

Our earlier work (Mukherjee and Joshi, 2013) in sentiment aggregation using ontology ignored the identity of the authors. It only took the domain-specific facet associations into consideration while deciding the overall review polarity. We consider it to be the *third baseline* for our work.

It is well-established from earlier works that supervised prediction of polarity fares better than the lexicon-based approaches. Hence, in the last two baselines we use Support Vector Machines with L_2-loss, L_2-regularizer and unigram bag-of-words features trained over the dataset in (Maas et al., 2011) to find the polarity of the sentence containing a facet, which is assigned to the facet under consideration.

In this work, we propose an approach to do an author-specific hierarchical aggregation of sentiment over a domain ontology tree using supervision. This builds over all the earlier baselines.

We report both the accuracy of the classifier over the entire dataset, as well as the author-specific accuracy. The latter computes the average accuracy of the classifier per-author.

Dataset	Authors	Avg Rev/ Author	Rev/ Rating			Avg Rev Length	Avg Words/ Rev
			Pos	**Neg**	**Total**		
Movie Review*	312	7	1000	1000	2000	32	746
			Pos	**Neg**	**Total**		
Movie Review⊥	65	23	705	762	1467	32	711

Table 1: Movie Review Dataset Statistics (* denotes the original data, ⊥ indicates processed data)

Model	Author Acc.	Overall Acc.
Bag-of-words Support Vector Machine (Pang and Lee, 2002; Pang and Lee, 2004; Mullen and Collier, 2004)	80.23	78.49
Author-Specific Analysis using Regression (Mukherjee et al., 2013)	79.31	79.07
Ontological Sentiment Aggregation (Mukherjee and Joshi, 2013)	81.4	79.51
PASOT	**86.32**	**86.04**

Table 2: Accuracy Comparison with Baselines

4.3. Results

Table 2 shows the accuracy comparison of our approach with different baselines. We also compare our approach to other works in the domain on the same dataset and report five-fold cross validation results in Table 3.

Figure 4 shows the variation of the Expected Sentiment Weight of different features with the overall review rating for the author of Review 2.. The expected sentiment weight of a feature encapsulates the feature polarity in the review, the feature depth in the ontology and the author-preference for the feature. The following movie features are considered for analysis : "film story, film type, film crew, film character aspect, film dialogue, film visual effect, film crew" and "camera crew".

Figure 5 shows the variation of the Expected Sentiment Weight of different features with the overall review rating for 10 authors.

4.4. Thwarting

The concept of "thwarted expectations" was first introduced by (Pang and Lee, 2002), and since then it has been considered to be a difficult and challenging problem to deal with (Pang and Lee, 2002; Mullen and Collier, 2004; Mukherjee and Bhattacharyya, 2012b). Thwarting phenomenon is observed where the overall review polarity is different from that of the majority of the opinion words in the review. The authors argued that some sophisticated technique is required to determine the focus of each review sentence and its relatedness to the review, as "the whole is not necessarily the sum of the parts" (Turney, 2002).

Consider the classical example of thwarting from (Pang and Lee, 2002) :

"This film sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."

Models	Acc.
Eigen Vector Clustering (Dasgupta and Ng, 2009)	70.9
Semi Supervised, 40% doc. Label (Li et al., 2009)	73.5
LSM Unsupervised with prior info (Lin et al., 2010)	74.1
SO-CAL Full Lexicon (Taboada et al., 2011)	76.37
RAE Semi Supervised Recursive Auto Encoders with random word initialization (Socher et al., 2011)	76.8
WikiSent: Extractive Summarization with Wikipedia + Lexicon (Mukherjee and Bhattacharyya, 2012b)	76.85
Supervised Tree-CRF (Nakagawa et al., 2010)	77.3
RAE: Supervised Recursive Auto Encoders with 10% cross-validation (Socher et al., 2011)	77.7
JST: Without Subjectivity Detection using LDA (Lin and He, 2009)	82.8
Pang *et al.* (2002): Supervised SVM (Pang and Lee, 2002)	82.9
JST: With Subjectivity Detection (Lin and He, 2009)	84.6
PASOT	**86.04**
Kennedy *et al.* (2006): Supervised SVM (Kennedy and Inkpen, 2006)	86.2
Supervised Subjective MR, SVM (Pang and Lee, 2004)	87.2
JAST: Joint Author Sentiment Topic Model (Mukherjee et al., 2014)	87.69
Appraisal Group: Supervised (Whitelaw et al., 2005)	90.2

Table 3: Comparison of Existing Models with PASOT in the IMDB Dataset

The overall review sentiment is negative despite having more positive sentiment words than negative ones. This implies that the overall review sentiment should not be a simple aggregation over all the polarities in a review. Here, the author sentiment is positive about "plot, actors" and "cast", which is not as important as his negative sentiment about the most important feature of the review, *i.e.* the "film". Thus the review rating should be a weighted function of the individual feature-specific polarities; where the domain importance and author preference of a feature should be considered to find the overall review polarity.

The proper polarity of this review is captured in our approach, as the negative polarity of "movie" at the top of the ontology tree is weighed up by (inverse of) its depth and the author preference, making it dominate other features with positive polarities at a greater depth in the tree.

Table 4 shows the number of reviews for positive thwarted and negative thwarted data used in our experimentation, as well as the accuracy comparison of our approach with an L_2-loss Support Vector Machine baseline using bag-of-words features.

4.5. Discussions

Table 2 shows the gradual performance improvement (in terms of overall accuracy) of each of the models - Author-

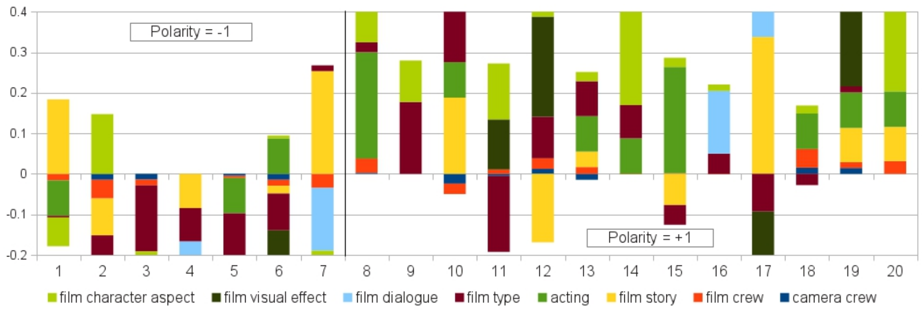

Figure 4: Variation of Expected Sentiment Weight of Facets with Review Rating for a Specific Author

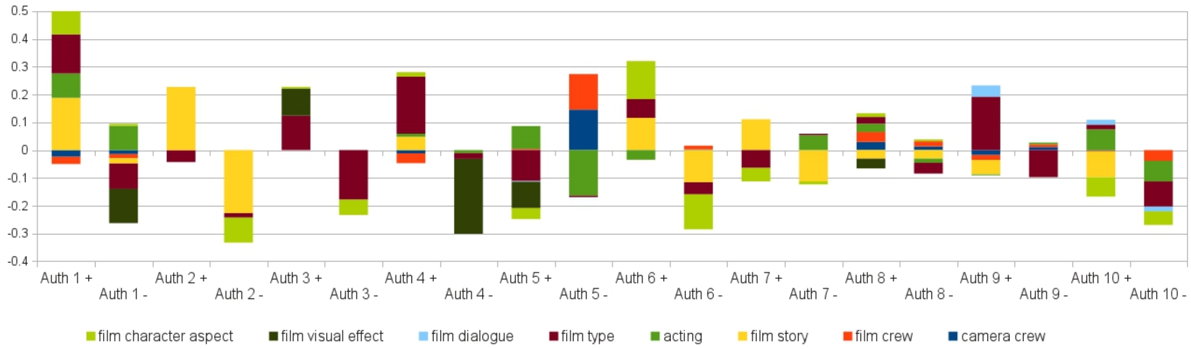

Figure 5: Variation of Expected Sentiment Weight of Facets with Review Rating for 10 Authors

Dataset	Positive Thwarted	Negative Thwarted
1467	279	132

Model	Thwarting Acc.
Bag-of-words SVM	61.54
PASOT	73.07

Table 4: Thwarting Accuracy Comparison

Specific LR, Ontological Sentiment Aggregation and PASOT, over the SVM baseline. The Phrase annotated Author specific Sentiment Ontology Tree (PASOT) approach achieves an overall accuracy improvement of 7.55% and 6% average accuracy improvement for each author, over the bag-of-words SVM baseline.

Table 3 shows the accuracy comparison of our approach with all the state-of-the-art systems in the domain that used the same IMDB dataset as ours. Since the objective of this work has been to show the effectiveness of an author-specific, hierarchical sentiment aggregation approach that can be built over an unigram bag-of-words SVM baseline, we did not experiment with a richer feature representation; for example, a combination of unigrams and bigrams with subjectivity analysis (Pang and Lee, 2004) built into the SVM have been found to be effective features for movie review classification. However, even with simple unigram features our model performs better than many systems using a richer feature representation.

Table 4 shows the effectiveness of our approach in capturing *thwarting* in reviews, where we achieve an accuracy improvement of 11.53% over the SVM baseline.

Figure 4 shows the variation of the Expected Sentiment Weight of different features with the overall review rating for the author of Review 2.. It shows that the overall rating of a movie by this author is highly influenced by the "film type" (genre), the characters in the film ("film character aspects"), "film dialogue" and "acting" of the protagonists, whereas he is quite flexible with the quality of "film story".

Figure 5 shows the variation of the Expected Sentiment Weight of different features with the overall review rating for 10 authors. It shows that, in general, the quality of the "film story" and its genre ("film type") plays a deciding role for the overall rating of the movie.

The graph further shows that Author 1 seems to be flexible with the quality of "acting" provided the "film type" is good, whereas Author 10 has a high preference for the quality of "acting" which decides his movie ratings.

This clearly depicts the importance of an author-specific analysis for reviews, where facet preferences vary for different authors leading to different overall ratings.

5. Related Work

Earlier works (Pang and Lee, 2002; Pang and Lee, 2004; Turney, 2002; Mullen and Collier, 2004) in sentiment analysis considered a review as a bag-of-words, where the different topics or facets of a product were ignored. Features like unigrams, bigrams, adjectives *etc.* were used followed by the usage of phrase-based features like part-of-speech sequences (E.g. adjectives followed by nouns).

These works were followed by feature-specific sentiment analysis, where the polarity of a sentence or a review is determined with respect to a given feature. Approaches like dependency parsing (Mukherjee and Bhattacharyya, 2012a), joint sentiment topic model (Lin and He, 2009) have been used to extract feature-specific opinions.

Latter works focused on aspect rating prediction that identifies aspects, aspect ratings, and weights placed on the aspects in a review (Wang et al., 2011).

All of these works attempt to learn a global model over the corpus, independent of the author of the review, and capture only the popular sentiment. In our recent works (Mukherjee et al., 2013; Mukherjee et al., 2014), we focused on learning the effect of author-specific facet preferences and author-writing style in modeling a review from the point of view of an author.

However, these works ignore the association between the features of a product that influence the overall rating of a review. Some recent works have focused on the hierarchical learning of a product's attributes and their associated sentiments in product reviews using a Sentiment Ontology Tree (Wei and Gulla, 2010; Mukherjee and Joshi, 2013).

In this work, we bring together all of the above ideas to propose an author-specific, hierarchical aggregation of sentiment over a product ontology tree.

6. Conclusions

In this work, we show that an author-specific sentiment aggregation of reviews perform better than an author-independent model that does not take the author-specific facet preferences and domain-specific facet relationships into account. We propose an approach to construct a *Phrase annotated Author specific Sentiment Ontology Tree (PASOT)*, where the facet nodes are annotated with opinion phrases of the author in the review and the author's preference for the facets.

We perform experiments in the movie review domain, where we achieve an accuracy improvement of 7.55% over the SVM baseline. As a use-case, we show that our approach is effective in capturing *thwarting* in reviews, achieving an accuracy improvement of 11.53% over the SVM baseline.

7. References

Dasgupta, S. and Ng, V. (2009). Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification. EMNLP '09.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). Liblinear: A library for large linear classification. *J. Mach. Learn. Res.*

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. KDD '04.

JedFilm. (2014). Cinema ontology project, March.

Kennedy, A. and Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*.

Li, T., Zhang, Y., and Sindhwani, V. (2009). A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In *ACL/IJCNLP*.

Lin, C. and He, Y. (2009). Joint sentiment/topic model for sentiment analysis. CIKM '09.

Lin, C.-J., Weng, R. C., and Keerthi, S. S. (2008). Trust region newton method for logistic regression. *J. Mach. Learn. Res.*

Lin, C., He, Y., and Everson, R. (2010). A comparative study of bayesian models for unsupervised sentiment detection. CoNLL '10.

Liu, H. and Singh, P. (2004). Conceptnet — a practical commonsense reasoning tool-kit. *BT Technology Journal*.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C. (2011). Learning word vectors for sentiment analysis. Proceedings of ACL.

Mukherjee, S. and Bhattacharyya, P. (2012a). Feature specific sentiment analysis for product reviews. In *CICLing*.

Mukherjee, S. and Bhattacharyya, P. (2012b). Wikisent: weakly supervised sentiment analysis through extractive summarization with wikipedia. ECML PKDD'12.

Mukherjee, S. and Joshi, S. (2013). Sentiment aggregation using conceptnet ontology. In *IJCNLP*.

Mukherjee, S., Basu, G., and Joshi, S. (2013). Incorporating author preference in sentiment rating prediction of reviews. WWW '13.

Mukherjee, S., Ajmera, J., and Joshi, S. (2014). Unsupervised approach for shallow domain ontology construction from corpus. WWW '14.

Mukherjee et al., S. (2014). Joint author sentiment topic model. In *SDM '14*.

Mullen, T. and Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *EMNLP*.

Nakagawa, T., Inui, K., and Kurohashi, S. (2010). Dependency tree-based sentiment classification using crfs with hidden variables. HLT '10.

Pang, B. and Lee, Lillian, V. S. (2002). Thumbs up?: sentiment classification using machine learning techniques. EMNLP '02.

Pang, B. and Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. ACL '04.

Protégé. (2014). Protégé, March.

Socher et al., R. (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. EMNLP.

Taboada et al., M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*.

Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *ACL*.

Wang, H., Lu, Y., and Zhai, C. (2011). Latent aspect rating analysis without aspect keyword supervision. KDD '11.

Wei, W. and Gulla, J. A. (2010). Sentiment learning on product reviews via sentiment ontology tree. ACL '10.

Whitelaw, C., Garg, N., and Argamon, S. (2005). Using appraisal groups for sentiment analysis. CIKM '05.

Wu, Z. and Palmer, M. (1994). Verbs semantics and lexical selection. In *Proceedings of ACL*.