

# Detecting Turnarounds in Sentiment Analysis: Thwarting

**Ankit Ramteke**

Dept. of Computer Science & Engg.,  
Indian Institute of Technology  
Bombay, Mumbai, India.  
ankitr@cse.iitb.ac.in

**Akshat Malu**

Dept. of Computer Science & Engg.,  
Indian Institute of Technology  
Bombay, Mumbai, India.  
akshatmalu@cse.iitb.ac.in

**Pushpak Bhattacharyya**

Dept. of Computer Science & Engg.,  
Indian Institute of Technology  
Bombay, Mumbai, India.  
pb@cse.iitb.ac.in

**J. Saketha Nath**

Dept. of Computer Science & Engg.,  
Indian Institute of Technology  
Bombay, Mumbai, India.  
saketh@cse.iitb.ac.in

## Abstract

Thwarting and sarcasm are two uncharted territories in sentiment analysis, the former because of the lack of training corpora and the latter because of the enormous amount of world knowledge it demands. In this paper, we propose a working definition of thwarting amenable to machine learning and create a system that detects if the document is thwarted or not. We focus on identifying thwarting in product reviews, especially in the camera domain. An ontology of the camera domain is created. *Thwarting is looked upon as the phenomenon of polarity reversal at a higher level of ontology compared to the polarity expressed at the lower level.* This notion of thwarting defined with respect to an ontology is novel, to the best of our knowledge. A rule based implementation building upon this idea forms our baseline. We show that machine learning with annotated corpora (thwarted/non-thwarted) is more effective than the rule based system. Because of the skewed distribution of thwarting, we adopt the Area-under-the-Curve measure of performance. To the best of our knowledge, this is the first attempt at the difficult problem of thwarting detection, which we hope will at

least provide a baseline system to compare against.

## 1 Credits

The authors thank the lexicographers at Center for Indian Language Technology (CFILT) at IIT Bombay for their support for this work.

## 2 Introduction

Although much research has been done in the field of sentiment analysis (Liu *et al.*, 2012), *thwarting* and *sarcasm* are not addressed, to the best of our knowledge. Thwarting has been identified as a common phenomenon in sentiment analysis (Pang *et al.*, 2002, Ohana *et al.*, 2009, Brooke, 2009) in various forms of texts but no previous work has proposed a solution to the problem of identifying thwarting. We focus on identifying thwarting in product reviews.

The definition of an opinion as specified in Liu (2012) is

“An opinion is a quintuple,  $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ , where  $e_i$  is the name of an entity,  $a_{ij}$  is an aspect of  $e_i$ ,  $s_{ijkl}$  is the sentiment on aspect  $a_{ij}$  of entity  $e_i$ ,  $h_k$  is the opinion holder, and  $t_l$  is the time when the opinion is expressed by  $h_k$ .”

If the sentiment towards the entity or one of its important attribute contradicts the sentiment towards all other attributes, we can say that the document is thwarted.

A domain ontology is an ontology of various features pertaining to a domain, arranged in a hierarchy. Subsumption in this hierarchy implies that the child is a part or feature of the parent. Domain ontology has been used by various works in NLP (Saggion *et al.*, 2007 and Polpinij *et al.*, 2008). In our work, we use domain ontology of camera. We look upon thwarting as the phenomenon of *reversal of polarity* from the lower level of the ontology to the higher level. At the higher level of ontology the entities mentioned are the whole product or a large critical part of the product. So while statements about entities at the lower level of the ontology are on “details”, statements about entities at higher levels are on the “big picture”. **Polarity reversal from details to the big picture is at the heart of thwarting.**

The motivation for our study on thwarting comes from the fact that: a) Thwarting is a challenging NLP problem and b) Special ML machinery is needed in view of the fact that the training data is so skewed. Additionally large amount of world and domain knowledge maybe called for to solve the problem. In spite of the relatively fewer occurrence of the thwarting phenomenon the problem poses an intellectually stimulating exercise. We may also say that in the limit, thwarting approaches the very difficult problem of sarcasm detection (Tsur *et al.* 2010).

We start by defining and understanding the problem of thwarting in section 2. In section 3, we describe a method to create the domain ontology. In section 4, we propose a naïve rule based approach to detect thwarting. In section 5 we discuss a machine learning based approach which could be used to identify whether a document is thwarted or not. This is followed by experimental results in section 6. Section 7 draws conclusions and points to future work.

### 3 Definition

Thwarting is defined by Pang *et al.*, (2008) as follows:

*“Thwarted expectations basically refer to the phenomenon wherein the author of the text first builds up certain expectations for the topic, only to produce a deliberate contrast to the earlier discussion.”*

For our computational purposes, we define thwarting as:

*“The phenomenon wherein the overall polarity of the document is in contrast with the polarity of majority of the document.”*

This definition emphasizes thwarting as piggy-backing on sentiment analysis to improve the latter’s performance. The current work however only addresses the problem of whether a document is thwarted or not and does not output the sentiment of the document. The basic block diagram for our system is shown in figure 1.

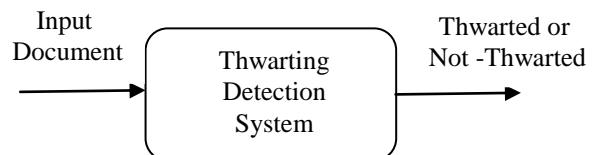


Figure 1: Basic Block Diagram

An example of a thwarted document is:

*“I love the sleek design. The lens is impressive. The pictures look good but, somehow this camera disappoints me. I do not recommend it.”*

While thwarting occurs in various forms of sentiment bearing texts, it is not a very frequent one. It accounts for hardly 1-2% of any given corpus. Thus, it becomes hard to find sufficient number of examples of thwarting to train a classifier.

Since thwarting is a complex natural language phenomenon we require basic NLP tools and resources, whose accuracy in turn can affect the overall performance of a thwarting detection system.

### 4 Building domain ontology

Domain ontology comprises of features and entities from the domain and the relationships between them. The process thus has two steps, *viz.* (a) identify the features and entities, and (b) connect them in the form of a hierarchy. We decided to use a combination of review corpora mining and manual means for identifying key features. Our approach to building the domain ontology is as follows:

**Step 1:** We use Latent Dirichlet Allocation (LDA) (Blei *et al.*, 2003) on a corpus containing reviews of a particular product (camera, in our case) to identify key features from the domain. The output is then analyzed manually to finally select the key features. Some additional features get added by human annotator to increase the coverage of the ontology. For Example, in the camera domain, the corpus may include words

like *memory, card, gb, etc.* but, may not contain the word *storage*. The abstract concept of *storage* is contributed by the human annotator through his/her world knowledge.

**Step 2:** The features thus obtained are arranged in the form of a hierarchy by a human annotator.

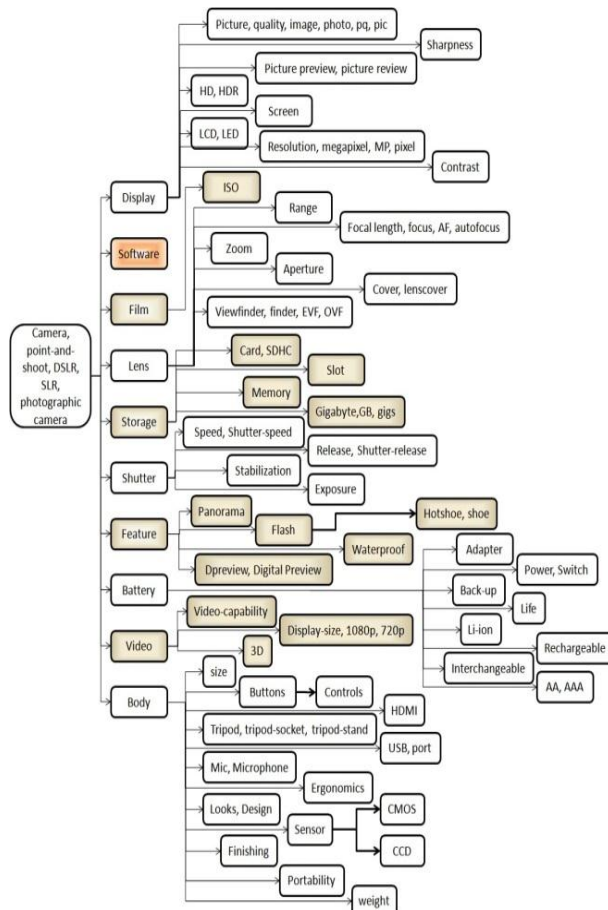


Figure 2: Ontology for the camera domain

## 5 A rule based approach to thwarting recognition

As per the definition of thwarting, most of the thwarted document carries a single sentiment; however, a small but critical portion of the text, carrying the contrary sentiment, actually decides the overall polarity. The critical statement, thus, should be strongly polar (either positive or negative), and it should be on some critical feature of the product.

*From the perspective of the domain ontology, the sentiment towards the overall product or towards some critical feature mentioned near the root of the ontology should be opposite to the sentiment towards features near the leaves.*

Based on these observations we propose the following naïve approach to thwarting detection:

For each sentence in a review to be tested

1. Get the dependency parse of the sentence. This step is essential. It makes explicit the adjective noun dependencies, which in turn uncovers the sentiment on a specific part or feature of the product.

2. Identify the polarities towards all nouns, using the dependency parse and sentiment lexicons.

3. If a domain feature, identified using the domain ontology, exists in the sentence, annotate/update the ontology node, containing the feature, using the polarity obtained.

Once the entire review is processed, we obtain the domain ontology, with polarity marking on nodes, for the corresponding review.

*The given review is thwarted if there is a contradiction of sentiment among different levels of the domain ontology with polarity marking on nodes.*

The sentiment lexicons used are SentiWordNet (Esuli et al., 2006), Taboada (Taboada et al., 2004), BL lexicon (Hu et al., 2004) and Inquirer (Stone et al., 1966).

The procedure is illustrated by an example. *“I love the sleek design. The lens is impressive. The pictures look good but, somehow this camera disappoints me. I do not recommend it.”*

A part of the ontology, with polarity marking on nodes, for this example is shown in figure 3.

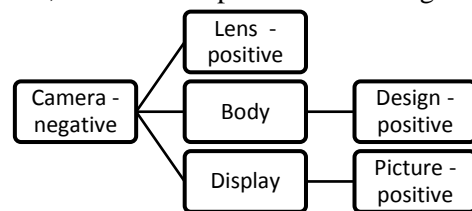


Figure 3: ontology with polarity marking on nodes: example

Based on this ontology we see that there is an opposition of sentiment between the root (“camera”) and the lower nodes. We thus determine that this document is thwarted.

However, since the nodes, within the same level, might have different weighting based upon the product under consideration, this method fails to perform well. For example, the *body* and *video capability* might be subjective whereas any fault in the *lens* or the *battery* will render the camera useless, hence they are more critical. We thus see a need for relative weighting among all features in the ontology.

## 6 A Machine Learning based approach

Manual fixing of relative weightages for the features of the product is possible, but that would be *ad hoc*. We now propose a machine learning based approach to detect thwarting in documents. It uses the domain ontology to identify key features related to the domain. The approach involves two major steps namely learning the weights and building a model that classifies the reviews using the learnt weights.

### 6.1 Learning Weights

The weights are learnt using the loss-regularization framework. The key idea is that the overall polarity of the document is determined by the polarities of individual words in the document. Since, we need to find the weights for the nodes in the domain ontology; we consider only the words belonging to the ontology for further processing. Thus, if  $P$  is the polarity of the review and  $p_i$  is the polarity associated with word  $i$  then  $P = \sum_i w_i p_i$  gives the linear model. The word  $i$  should belong to the ontology as well as the review. Similarly, the hinge loss is given by  $\max(0, 1 - P \cdot w^T x)$  where  $w$  is the weight vector and  $x$  is the feature vector consisting of  $p_i$ 's.

Based on the intuition, that every word contributes some polarity to its parent node in the domain ontology, we also learnt weights on the ontology by percolating polarities towards the root. We experimented with complete percolation, wherein the polarity at a node is its polarity in the document summed with the polarities of all its descendants. We also define controlled percolation, wherein the value added for a particular descendant is a function of its distance from the node. We halved the polarity value percolated, for each edge between the two nodes. Thus, for the example in figure 2, the polarity value of camera would be

$$P_{camera} = p_{camera} + \frac{p_{lens}}{2} + \frac{p_{body}}{2} + \frac{p_{display}}{2} + \frac{p_{design}}{4} + \frac{p_{picture}}{4}$$

Where  $P_{camera}$  is the final polarity for camera and  $p_{word}$  is the polarity of the word  $\in$  {camera, body, display, design, picture}.

### 6.2 Classifier

We use the SVM classifier with features generated using the following steps. We first create a vector of weighted polarity values for each review. This is constructed by generating a value

for each word in the domain ontology encountered while reading the review sequentially. The value is calculated by multiplying the weight, found in the previous step (5.1), with the polarity of the word as determined from the sentence. Since, these vectors will be of different dimensionality for each review, we extract features from these reviews. These features are selected based on our understanding of the problem and the fact that thwarting is a function of the change of polarity values and also the position of change.

#### The Features extracted are:

Document polarity, number of flips of sign (*i.e.* change of polarity from positive to negative and vice versa), the maximum and minimum values in a sequence, the length of the longest contiguous subsequence of positive values (LCSP), the length of the longest contiguous subsequence of negative values (LCSN), the mean of all values, total number of positive values in the sequence, total number of negative values in the sequence, the first and the last value in the sequence, the variance of the moving averages, the difference in the means of LCSP and LCSN.

## 7 Results

Experiments were performed on a dataset obtained by crawling product reviews from Amazon<sup>1</sup>. We focused on the camera domain. We obtained 1196 reviews from this domain. The reviews were annotated for thwarting, *i.e.*, thwarted or non-thwarted as well as polarity. The reviews crawled were given to three different annotators. The instructions given for annotation were as follows:

1. Read the entire review and try to form a mental picture of how sentiment in the document is distributed. Ignore anything that is not the opinion of the writer.
2. Try to determine the overall polarity of the document. The star rating of the document can be used for this purpose.
3. If the overall polarity of the document is negative but, most of the words in the document indicate positive sentiment, or vice versa, then consider the document as thwarted.

Since, identifying thwarting is a difficult task even for humans, we calculated the Cohen's kappa score (Cohen 1960) in order to determine the inter annotator agreement. It was found out to

---

<sup>1</sup>Reviews crawled from <http://www.amazon.com/>

be **0.7317**. The annotators showed high agreement (98%) in the non-thwarted class whereas they agreed on 70% of the thwarted documents.

Out of the 1196 reviews, exactly 21 were thwarted documents, agreed upon by all annotators. We used the Stanford Core NLP tools<sup>2</sup> (Klein *et al.*, 2003, Toutanova *et al.*, 2003) for basic NL processing. The system was tested on the entire dataset.

Since, the data is highly skewed; we used Area under the Curve (AUC) for the ROC curve as the measure of evaluation (Ling *et al.*, 2003). The AUC for a random baseline is expected to be 50%, and the rule based approach is close to the baseline (**56.3%**).

Table 1 shows the results for the experiments with the machine learning model. We used the CVX<sup>3</sup> library in Matlab to solve the optimization problem for learning weights and the LIBSVM<sup>4</sup> library to implement the svm classifier. In order to account for the data skew, we assign a class weight of 50 (determined empirically) to the thwarted instances and 1 for non-thwarted instances in the classifier. All results were obtained using a 10 fold cross validation. The same dataset was used for this set of experiments.

| Loss type for weights | Percolation type for weights | AUC value for classification |
|-----------------------|------------------------------|------------------------------|
| Linear                | Complete                     | 73%                          |
|                       | Controlled                   | <b>81%</b>                   |
| Hinge                 | Complete                     | 70%                          |
|                       | Controlled                   | 76%                          |

Table 1: Results of the machine learning based approach to thwarting detection

We see that the overall system for identification of thwarting performs well for the weights obtained using the linear model with a controlled percolation of polarity values in the ontology. The system outperforms both the random baseline as well as the rule based system. These results though great are to be taken with a pinch of salt. The basic objective for creating a thwarting detection system was to include such a module in the general sentiment analysis framework. Thus, using document polarity as a feature contradicts the objective of sentiment analysis, which is to *find* the document polarity. Without the docu-

ment polarity feature, the values drop by 10% which is not acceptable.

## 8 Conclusions and Future Work

We have described a system for detecting thwarting, based on polarity reversal between opinion on most parts of the product and opinion on the overall product or a critical part of the product. The parts of the product are related to one another through an ontology. This ontology guides a rule based approach to thwarting detection, and also provides features for an SVM based learning system. The ML based system scores over the rule based system. Future work consists in trying out the approach across products and across domains, doing better ontology harnessing from the reviews and investing and searching for distributions and learning algorithms more suitable for the problem.

## References

- Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent Dirichlet allocation. In *the Journal of machine Learning research*, 3, pages 993-1022.
- Brooke, J. 2009. *A Semantic Approach to Automated Text Sentiment Analysis*. Ph.D. thesis, Simon Fraser University.
- Chang, C. C., and Lin, C. J. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 27.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, no. 1, pages 37-46.
- Esuli, A. and Sebastiani, F. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, Volume 6, pages 417-422.
- Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168-177. ACM.
- Klein, D. and Manning, C. D. 2003. Accurate Unlexicalized Parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pages 423-430.
- Ling, C. X., Huang, J. and Zhang, H. 2003. AUC: A better measure than accuracy in comparing learning algorithms. In *Advances in Artificial Intelligence*, pages 329-341, Springer Berlin Heidelberg.

<sup>2</sup><http://nlp.stanford.edu/software/corenlp.shtml>

<sup>3</sup><http://cvxr.com/cvx>

<sup>4</sup><http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

- Liu, B., and Zhang, L. 2012. A survey of opinion mining and sentiment analysis. In *Mining Text Data* (pp. 415-463). Springer US.
- Liu B., 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
- Ohana, B. and Tierney, B. 2009. Sentiment classification of reviews using SentiWordNet. In *9th. IT & T Conference*, page 13.
- Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- Pang, B., Lee, L. and Vaithyanathan S. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of EMNLP* pages 79-86).
- Polpinij, J. and Ghose, A. K. 2008. An ontology-based sentiment classification methodology for online consumer reviews. In *Web Intelligence and Intelligent Agent Technology*.
- Taboada, M. and Grieve, J. 2004. Analyzing appraisal automatically. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS# 04# 07)*, Stanford University, CA, pages. 158-161. AAAI Press.
- Toutanova, K., Klein, D., Manning, C. D. and Singer Y. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL*, pages 252-259.
- Tsur, O., Davidov, D., & Rappoport, A. 2010. ICWSM—A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *Proceedings of the fourth international AAAI conference on weblogs and social media*, pages. 162-169.
- Saggion, H., Funk, A., Maynard, D. and Bontcheva, K. 2007. Ontology-based information extraction for business intelligence. In *The Semantic Web* pages 843-856, Springer Berlin Heidelberg.
- Stone, P. J., Dunphy, D. C., Smith, M. S., Ogilvie, D. M. and Associates. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.