Expert Systems With Applications 118 (2019) 65–79

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# Sentiment analysis based on rhetorical structure theory:Learning deep neural networks from discourse trees

Mathias Kraus*, Stefan Feuerriegel

*ETH Zurich, Weinbergstr. 56/58, Zurich, 8092, Switzerland*

ARTICLE INFO

*Article history:*
Received 9 July 2018
Revised 1 October 2018
Accepted 2 October 2018
Available online 4 October 2018

*Keywords:*
Sentiment analysis
Rhetorical structure theory
Discourse tree
Tree-structured network
Long short-term memory
Tensor-based network

ABSTRACT

Prominent applications of sentiment analysis are countless, covering areas such as marketing, customer service and communication. The conventional bag-of-words approach for measuring sentiment merely counts term frequencies; however, it neglects the position of the terms within the discourse. As a remedy, we develop a discourse-aware method that builds upon the discourse structure of documents. For this purpose, we utilize rhetorical structure theory to label (sub-)clauses according to their hierarchical relationships and then assign polarity scores to individual leaves. To learn from the resulting rhetorical structure, we propose a tensor-based, tree-structured deep neural network (named Discourse-LSTM) in order to process the complete discourse tree. The underlying tensors infer the salient passages of narrative materials. In addition, we suggest two algorithms for data augmentation (node reordering and artificial leaf insertion) that increase our training set and reduce overfitting. Our benchmarks demonstrate the superior performance of our approach. Moreover, our tensor structure reveals the salient text passages and thereby provides explanatory insights.



## 1. Introduction

Sentiment analysis reveals personal opinions towards entities such as products, services or events, which can benefit organizations and businesses in improving their marketing, communication, production and procurement. For this purpose, sentiment analysis quantifies the positivity or negativity of subjective information in narrative materials (Chen, Xu, He, & Wang, 2017; Feldman, 2013; Kratzwald, Ilic, Kraus, Feuerriegel, & Prendinger, 2018; Pang & Lee, 2008). Among the many applications of sentiment analysis are tracking customer opinions (Araque, Corcuera-Platas, Sánchez-Rada, & Iglesias, 2017; Bohanec, Kljajić Borštnar, & Robnik-Šikonja, 2017; Tanaka, 2010), mining user reviews (Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013; Mostafa, 2013; Ye, Zhang, & Law, 2009), trading upon financial news (Khadjeh Nassirtoussi, Aghabozorgi, Ying Wah, & Ngo, 2015; Kraus & Feuerriegel, 2017; Weng, Lu, Wang, Megahed, & Martinez, 2018), detect social events (Yoo, Song, & Jeong, 2018) and predicting sales (Rui, Liu, & Whinston, 2013; Yu, Liu, Huang, & An, 2012).

Sentiment analysis traditionally utilizes bag-of-words approaches, which merely count the frequency of words (and tuples thereof) to obtain a mathematical representation of documents in matrix form (Dey, Jenamani, & Thakkar, 2018; Guzella & Caminhas, 2009; Manning & Schütze, 1999; Pang & Lee, 2008). As such, these approaches are not capable of taking into consideration semantic relationships between sections and sentences of a document. In naïve bag-of-words models, all clauses are assigned the same level of relevance, which cannot mark certain subordinate clauses more than others for purposes of inferring the sentiment. Conversely, the objective of this paper is to develop a discourse-aware method for sentiment analysis that can recognize differences in salience between individual subordinate clauses, as well as the discriminate the relevance of sentences based on their function (e. g.whether it introduces a new fact or elaborates upon an existing one).

Let us, for instance, consider the two examples in Fig. 1, which express opposite polarities. By simply counting the frequency of positive and negative words, we cannot discriminate between the texts, as both contain the same number of polarity terms. To reliably analyze the sentiment, it is essential to account for the semantic structure and the variable importance across passages. That is, we can identify the main clauses and then infer the correct tone of the examples by looking at them. Similarly, RST trees can locate relevant parts in lengthy texts. For instance, the concluding section of a newspaper article is typically relevant as it reports the opinion of the author.

Our method is based on rhetorical structure theory (RST), which incorporates the discourse structures of natural language. RST structures documents hierarchically (Mann & Thompson, 1988) by

* Corresponding author.
*E-mail addresses:* mathiaskraus@ethz.ch (M. Kraus), sfeuerriegel@ethz.ch (S. Feuerriegel).

https://doi.org/10.1016/j.eswa.2018.10.002

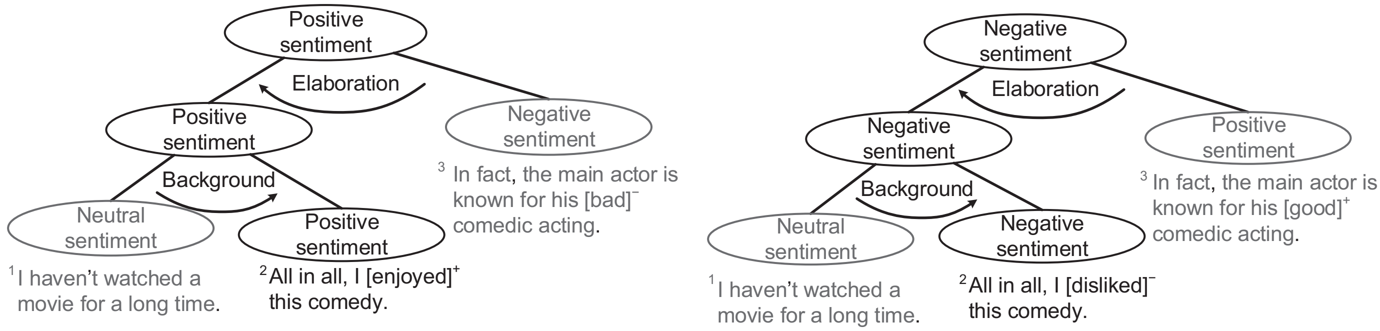


**Fig. 1.** Illustrative examples in which the discourse tree helps identify the conveyed sentiment from the main clause (highlighted in black). Here relation type additionally denotes the rhetorical function. The original inputs are: *"I haven't watched a movie for a long time. All in all, I liked/disliked this comedy. In fact, the main actor is known for his bad/good comedic acting."*.

splitting the content into (sub-)clauses called elementary discourse units (EDUs). The EDUs are then connected to form a binary discourse tree. Here RST discriminates between a nucleus, which conveys primary, and satellite, which conveys ancillary information. The formalization of nucleus/satellite can be loosely thought of main and subordinate parts of a clause. The edges are further labeled according to the type of discourse – for instance, whether it is an elaboration or an argument. Hence, this method essentially derives the function of a text passage. Both concepts of the RST tree help in localizing essential information within documents. Hence, the goal of this work is to develop a novel approach that identifies salient passages in a document based on their position in the discourse tree and incorporates their importance in the form of weights when computing sentiment scores.

Previous research has demonstrated that discourse-related information can improve the performance of sentiment analysis (see Section 2 for details). The work by Taboada, Voll, and Brooke (2008) is the first to combine rhetorical structure theory and sentiment analysis. In this work, the authors weigh adjectives in a nucleus more heavily than those in a satellite. Beyond that, one can reweigh the importance of passages based on their relation type (Hogenboom, Frasincar, de Jong, & Kaymak, 2015b) or depth (Märkle-Huß, Feuerriegel, & Prendinger, 2017) in the discourse tree. Some methods prune the discourse trees at certain thresholds to yield a tree of fixed depth, e. g.2 or 4 levels (Märkle-Huß et al., 2017). Other approaches train machine learning classifiers based on the relation types as input features (Hogenboom, Frasincar, de Jong, & Kaymak, 2015a). What the previous references have in common is that they try to map the tree structure onto mathematically simpler representations, thereby dropping partial information from the tree.

An alternative strategy is to apply tree-structured neural networks that traverse discourse trees for representation learning. When encountering a node, these networks combine the information from the leaves and pass them on to the next higher level, until reaching the root at which point a prediction is made. Thereby, the approach merely adheres to the tree-structure but does not account for either the relation type or whether it is a nucleus/satellite. To do so, one can extend the network to include different weights for each edge in the tree depending on, e. g., the relation type. This essentially introduces additional degrees of freedom that can weigh the different discourse units by their importance. The work by Fu, Liu, Xu, Yu, and Wang (2016) extends the network by such a mechanism with respect to the nucleus/satellite information but discards the relation type and merely applies the network to individual sentences instead of longer documents. The approach in Ji and Smith (2017) can only exploit the relation type and not the nucleus/satellite information. Furthermore, former approaches are based on traditional recursive neural networks, which are limited by the fact that they can persist information for only a few iterations (Bengio, Simard, & Frasconi, 1994). Therefore, these methods struggle with complex discourses, while we explicitly build upon tree-shaped long short-term memory models, since they are better equipped to handle very deep structures.

We build upon the previous works and advance them by proposing a specific neural network, called *Discourse-LSTM*. The Discourse-LSTM utilizes multiple tensors to localize salient passages within documents by incorporating the full discourse structure including nucleus/satellite information *and* relation types. In brief, our approach is as follows: we utilize rhetorical structure theory to represent the semantic structure of a document in the form of a hierarchical discourse tree. We then obtain sentiment scores for each leaf by utilizing both polarity dictionaries and word embeddings. The resulting tree is subsequently traversed by the Discourse-LSTM, thereby aggregating the sentiment scores based on the discourse structure in order to compute a sentiment score for the document. This approach thus weighs the importance of (sub-)clauses based on their position and relation in the discourse tree, which is learned during the training phase. As a consequence, this allows us to enhance sentiment analysis with discourse information. Another key contribution is that we propose two techniques for data augmentation that facilitate training and yield higher predictive accuracy.

The remainder of this paper is structured as follows. Section 2 reviews discourse parsing and RST-based sentiment analysis. Section 3 then introduces our Discourse-LSTM, as well as our algorithms for data augmentation. Section 4 describes our experimental setup in order to evaluate the performance of our deep learning methods in comparison to common baselines (Section 5). Section 6 concludes with a summary and suggestions for future research.

## 2. Background

### 2.1. Rhetorical structure theory

Rhetorical structure theory formalizes the discourse in narrative materials by organizing sub-clauses, sentences and paragraphs into a hierarchy (Mann & Thompson, 1988). The premise is that a document is split into elementary discourse units, which constitute the smallest, indivisible segments. These EDUs are then connected by one of 18 different relation types, which represent edges in the discourse tree; see Table 1 for a list. Each relation is further labeled by a hierarchy type, i. e.either as a nucleus ($N$) or a satellite ($S$). Here a nucleus denotes a more essential unit of information, while a satellite indicates a supporting or background unit of information. We note that RST also defines cases where both children are labeled as nuclei at the same time. Fig. 2 presents an example of a