Exploiting ontologies for deep learning: a case for sentiment mining[*]

Stephan Raaijmakers and Christopher Brewster

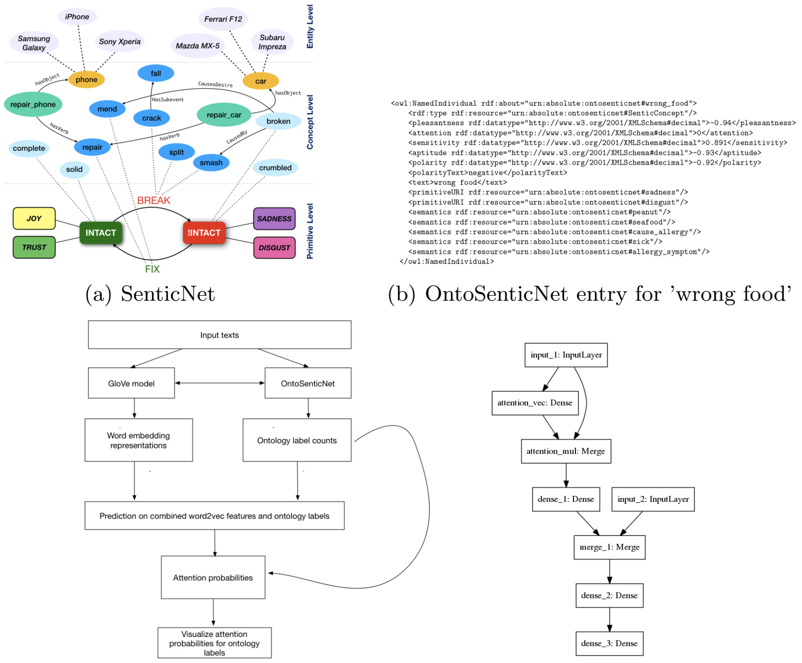
TNO, Data Science Department, Anna van Buerenplein 1, The Hague, 2595 DA, The Netherlands

Abstract. We present a practical method for explaining deep learning-based text mining with ontology-based information. Our approach uses the recently proposed OntoSenticNet ontology for sentiment mining, and consists of a composite deep learning classifier for sentiment mining, endowed with an ontology-driven attention module. The attention module analyzes the attention the neural network pays to semantic labels assigned to bigrams in input texts.

1 Introduction and approach

Deep learning continues to achieve state of the art performance in a variety of domains, such as image analysis and text mining. Despite this success, deep learning models remain elusive, and it is quite hard to understand what knowledge is represented in them, and how they generate decisions (see [3] for discussion). The field of *explainable AI* is increasingly gaining traction. Promising results have been reported with attention-based models [4] and latent-space analysis [7]. The link between ontologies and deep learning is actively being expored. For instance, [6] addresses the extraction of OWL information with deep learning from raw text and [2] applies deep learning to ontology extraction. Our approach attempts to leverage the semantic information in ontologies for explaining deep text mining, using neural attention and word embeddings. Ontologies usually contain structured, encyclopedic knowledge, arranged in a semantic, conceptual structure. One such ontology is the recently proposed sentiment ontology OntoSenticNet [1], an extension of the SenticNet ontology. SenticNet (Figure 1(a)) links entities via an intermediate concept level (consisting of semantic categories and relations) to an affective level describing sentiment-based associations, like sadness or joy. OntoSenticNet uses SenticNet to derive affective associations for words and phrases. It is automatically compiled from affective analyses performed with WordNet-Affect, Open Mind Common Sense and GECKA. Figure 1(b) lists the OntoSenticNet entry for “wrong food”. The `primitiveURI` nodes contain the affective labels associated with the multi-word expression “wrong food”. The `semantics` nodes express associations with other `NamedIndividuals` (expressions), based on corpus-based evidence such as collocations, and the static knowledge contained in SenticNet. We embed the ontology information into the

[*] The research reported in this paper has been carried out within the Research Programme Applied AI of TNO.



```
<owl:NamedIndividual rdf:about="urn:absolute:ontosenticnet#wrong_food">
    <rdf:type rdf:resource="urn:absolute:ontosenticnet#SenticConcept"/>
    <pleasantness rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">-0.94</pleasantness>
    <attention rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0</attention>
    <sensitivity rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0.891</sensitivity>
    <aptitude rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">-0.93</aptitude>
    <polarity rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">-0.92</polarity>
    <polarityText>negative</polarityText>
    <text>wrong food</text>
    <primitiveURI rdf:resource="urn:absolute:ontosenticnet#sadness"/>
    <primitiveURI rdf:resource="urn:absolute:ontosenticnet#disgust"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#peanut"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#seafood"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#cause_allergy"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#sick"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#allergy_symptom"/>
  </owl:NamedIndividual>
```

(a) SenticNet

(b) OntoSenticNet entry for 'wrong food'

(c) Process flow

(d) Model

Fig. 1. SenticNet, OntoSenticNet, our processing pipeline and model architecture.

sentiment analysis process directly, combining it with non-ontological information such as textual features. Taking advantage of the attention a neural network pays to the extra ontology-based information will allow us to decompose its decisions semantically. We start (Figure 1(c)) with generating vector representations of our input data, using 100-dimensional GloVe vectors [5], which were derived on the basis of 6 billion words coming from a 2014 English fragment of Wikipedia. Every document is represented as the sum of the GloVe vectors of its constituting words, normalized for the length of the document. Subsequently, we chunk up every document in bigrams, and perform a beam search over the semantically labeled bigrams in the OntoSenticNet ontology. As semantic labels for bigrams, we use the `primitiveURI` labels, and every combination in OntoSenticNet generates a unique label. In order to cater for bigrams without overt affective labels, we randomly took 5,000 bigrams from a BBC news corpus[1], and labeled these bigrams as 'bbc'. This approach yields, for every dataset we use, a unique set of semantic labels. Restricting our use of OntoSenticNet to bigrams allows us to look for contextual matches rather than for word-based matches, without run-

[1] http://mlg.ucd.ie/datasets/bbc.html

ning into sparsity: OntoSenticNet contains 22,935 bigram expressions, and only 3,104 expressions longer than 2 words. The majority of OntoSenticNet entries consists of unigrams (26,912 entries). The beam search operation attempts to retrieve, for any combination (100 total) of the 10 most similar words per word in the bigram, an existing bigram from OntoSenticNet. As an example, 'bad dinner' is not in OntoSenticNet, but one of its GloVe expansions ('wrong food') is. Once such a hit is found, the beam search stops for the given input bigram, the semantic labels are picked up from OntoSenticNet, and search proceeds with the next bigram in the document. The relation between an OntoSenticNet bigram and its labels is stored as an entry in a dictionary. The attested semantic label for every bigram in a document is counted, and for every document, a count vector (with as its length the total number of labels attested in the corpus) is generated and stored. After processing a labeled text corpus in this manner, every document in the corpus becomes represented by two vectors: a GloVe-based vector, and a count vector describing the counts for the semantic labels that apply to the bigrams in the document. Subsequently, we train a neural network (Figure1(d)) on these joint representations of labeled documents. The network has two branches, each equipped with a separate input layer. First, a branch processes the ontology label vectors, and computes attention scores (probabilities) for the various labels in the vectors. These attention scores indicate the importance ('attention') the network pays to the ontology labels. They are merged with the GloVe vectors by concatenation, and this derived representation is used by a second branch to learn the labeling of documents with sentiment labels. The attention probabilities are optimized during this process in an end-to-end fashion (they are part of the overall weight optimization problem the network is solving). Once learning is complete, for every test case, the attention scores as computed by the trained network for the test document are extracted from the network, and an image is generated that displays the scores. We applied our system to a variety of sentiment labeling datasets: a set of UCI datasets[2] comprising Yelp, Amazon product and IMDB movie reviews. In addition, we trained and tested on a subjectivity dataset[3].

2 Results

Some illustrative results are listed in Figure 2. For the complex emotion expressed in the sentence *The only thing I did like was the prime rib and the dessert section*, the OntoSenticNet labels `anger, sadness, disgust, surprise` score relatively high. Sentences *We'd definitely go back here again* and *Will go back next trip out* both score high for the joint label `joy#surprise`. The negative sentiment of *...least think to refill my water before I struggle to wave you over for 10 minutes* has significant underpinning with `disgust` and `anger` labels. The attention probabilities extracted from our classifier may thus serve to decompose monadic sentiment labels into much more rich and varied descriptions, enhancing

[2] https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences

[3] https://www.cs.cornell.edu/people/pabo/movie-review-data/

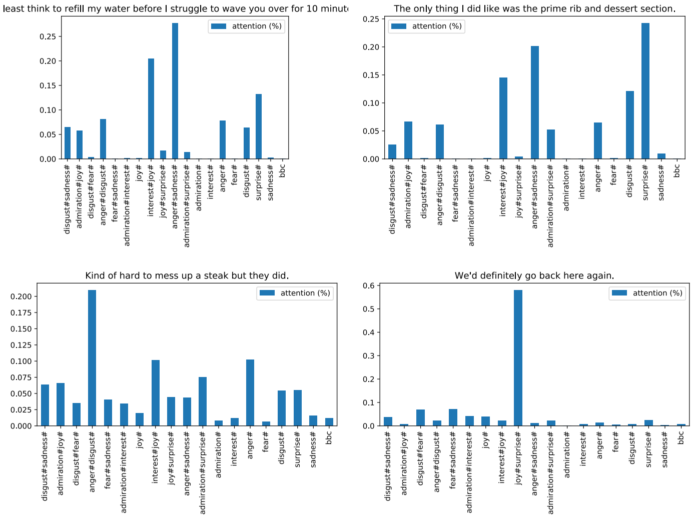


Fig. 2. Sample attention-based analyses.

the explainability of monadic sentiment labeling. The explanatory advantages of our system will be assessed in future work by submitting the generated analyses to human evaluators in a task-based evaluation setting, and by displaying the underlying words and phrases used by the model for sentiment decomposition. Our code will be shared at `https://github.com/stephanraaijmakers/deeptext`.

References

1. Dragoni, M., Poria, S., Cambria, E.: Ontosenticnet: A commonsense ontology for sentiment analysis. IEEE Computational Intelligence Magazine **33(2)** (2018)
2. Hohenecker, P., Lukasiewicz, T.: Deep learning for ontology reasoning. CoRR **abs/1705.10342** (2017)
3. Lipton, Z.C.: The mythos of model interpretability. CoRR **abs/1606.03490** (2016)
4. Liu, S., Wang, X., Liu, M., Zhu, J.: Towards better analysis of machine learning models: A visual analytics perspective. Visual Informatics **1**(1), 48 – 56 (2017)
5. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: EMNLP. vol. 14, pp. 1532–1543 (01 2014)
6. Petrucci, G., Ghidini, C., Rospocher, M.: Ontology learning in the deep. In: 20th International Conference on Knowledge Engineering and Knowledge Management - Volume 10024. pp. 480–495. EKAW 2016, Springer-Verlag New York, Inc., New York, NY, USA (2016)
7. Raaijmakers, S., Sappelli, M., Kraaij, W.: Investigating the interpretability of hidden layers in deep text mining. In: SEMANTICS. pp. 177–180. ACM (2017)