

MPQA 3.0: An Entity/Event-Level Sentiment Corpus

Lingjia Deng

Intelligent Systems Program

University of Pittsburgh

lid29@pitt.edu

Janyce Wiebe

Intelligent Systems Program

Department of Computer Science

University of Pittsburgh

wiebe@cs.pitt.edu

Abstract

This paper presents an annotation scheme for adding entity and event target annotations to the MPQA corpus, a rich span-annotated opinion corpus. The new corpus promises to be a valuable new resource for developing systems for entity/event-level sentiment analysis. Such systems, in turn, would be valuable in NLP applications such as Automatic Question Answering. We introduce the idea of entity and event targets (*eTargets*), describe the annotation scheme, and present the results of an agreement study.

1 Introduction

Much work in sentiment analysis and opinion mining is at the document level (Pang et al., 2002; Turney, 2002). There is increasing interest in more fine-grained levels - sentence-level (Yu and Hatzivassiloglou, 2003; McDonald et al., 2007), phrase-level (Choi and Cardie, 2008), aspect-level (Hu and Liu, 2004; Titov and McDonald, 2008), etc. We specifically address sentiments toward entities and events (i.e., *eTargets*) expressed in data such as blogs, newswire, and editorials. A system that could recognize sentiments toward entities and events would be valuable in an application such as Automatic Question Answering, to support answering questions such as “Toward whom/what is *X* negative/positive?” “Who is negative/positive toward *X*?” (Stoyanov et al., 2005). Or, to augment an automatic wikification system (Ratinov et al., 2011) – in addition to relationships such as *spouse* and *parents*, the system could include information about

whom or what the subject supports or opposes. A recent NIST evaluation – The Knowledge Base Population (KBP) Sentiment track¹ — aims at using corpora to collect information regarding sentiments expressed toward or by named entities.

Annotated corpora of reviews (e.g., (Hu and Liu, 2004; Titov and McDonald, 2008)), widely used in NLP, often include target annotations. Such targets are often aspects or features of products or services, and as such are somewhat limited.²

Recently, to create the Sentiment Treebank (Socher et al., 2013), researchers crowdsourced annotations of movie review data and then overlaid the annotations onto syntax trees. Thus, the targets are not limited to aspects of products/services. However, annotators were asked to annotate small and then increasingly larger segments of the sentence. Thus, the annotations are mixed in the degree to which context was considered when making the judgements. Previously, we (Deng et al., 2013) annotated a corpus of non-review data with sentiments toward entities, but only for those that participate in certain types of events. In all of the above corpora, the only sentiments considered are those of the writer, excluding sentiments attributed to other entities.

The MPQA opinion annotated corpus (Wiebe et al., 2005; Wilson, 2007) is entirely span-based, and contains no *eTarget* annotations. However, it provides an infrastructure for sentiment annotation that is not provided by other sentiment NLP corpora, and

¹<http://www.nist.gov/tac/2014/KBP/Sentiment/index.html>

²For example, as stated in SemEval-2014: “We annotate only aspect terms naming particular aspects.”

is much more varied in topic, genre, and publication source. This paper addresses adding eTarget annotations to the MPQA corpus; we believe that the result will be a valuable new resource for the community.

2 From MPQA 2.0 to MPQA 3.0

To create MPQA 3.0, entity-target and event-target (*eTarget*) annotations are added to the MPQA 2.0 annotations.³ The MPQA annotations consist of *private states*, which are states of *sources* holding *attitudes* toward *targets*. In the MPQA 2.0 annotations, the top-level annotations are *direct subjective (DS)* and *objective speech event* annotations. DS annotations are for private states, and objective speech event annotations are for objective statements attributed to a source. An important property of sources is that they are nested, reflecting the fact that private states and speech events are often embedded in one another.

As shown in Figure 1, one DS may contain links to multiple *attitude* annotations, meaning that all of the attitudes share the same nested source. The attitudes differ from one another in their attitude types, polarities, and/or targets. There are several types of attitudes included in MPQA 2.0 (Wilson, 2007; Somasundaran et al., 2007), including sentiment and arguing. This work focuses on sentiments, which are defined in (Wilson, 2007) as positive and negative evaluations, emotions, and judgements.

MPQA 2.0 also contains *expressive subjective element (ESE)* annotations, which pinpoint specific expressions used to express subjectivity (Wiebe et al., 2005). An ESE also has a nested-source annotation. Since we focus on sentiments, we only consider ESEs whose polarity is positive or negative (excluding those marked neutral).

The *target-span* annotations in MPQA 2.0 are linked to from the attitudes. More than one target may be linked to from an attitude, but most attitudes have only one target. The MPQA 2.0 annotators identified the main/most important target(s) they perceive in the sentence. If there is no target, the target-span annotation is “none”. However, there are many other eTargets to be identified. First, while ESE annotations have nested sources, they do not have any target annotations. Second, there are many

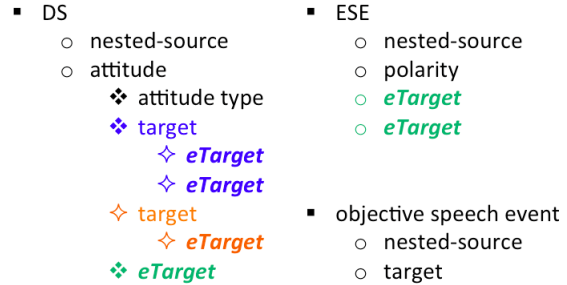


Figure 1: Structure in MPQA 3.0.

more targets that may be marked than the major ones identified in MPQA 2.0. In Figure 1, the eTargets are what we add in MPQA 3.0. We identify the blue (orange) eTargets that are in the span of a blue (orange) target in MPQA 2.0. We also identify the green eTargets that are not in the scope of any target.

Since our priority was to add eTargets to sentiments, no eTargets have yet been added to objective speech events, as shown in Figure 1.

To create MPQA 3.0, the corpus is first parsed, and *potential* eTarget annotations are automatically created from the heads of NPs and VPs. The annotators then consider each sentiment attitude and each polar ESE, and decide for each which eTargets to add. By adding eTargets to the existing annotations, the information in MPQA 2.0 is retained. Before presenting the scheme, we first give some examples.

2.1 Examples

For each example, a subset of the annotations are shown. The phrase in blue is an attitude span, the phrase in red is a target span, the tokens in yellow are the eTargets which are newly annotated in MPQA 3.0. The underlined phrases are ESE spans. Each example is followed by the MPQA structure of the annotations.

In Ex(1), a negative attitude is shown, *issued the fatwa against*. The source is *the Imam*. The target is the event *Rushdie insulting the Prophet*. However, the assertion that the Imam is negative toward the insult event is within the scope of this article. This is captured by an objective speech event annotation (not shown) whose target span includes the insult event, and whose source is the writer (*w*). Thus, the complete interpretation of this negative attitude is, according to the writer, the Imam is negative to-

³Available at <http://mpqa.cs.pitt.edu>

ward the insult event. And, the nested source is *w, imam*.

Ex(1) When the Imam issued the fatwa against Salman Rushdie for insulting the Prophet ...

DS: issued the fatwa
nested-source: *w, imam*
attitude: issued the fatwa against
attitude-type: sentiment-negative
target: Salman...insult...Prophet
eTarget: Rushdie, insulting

We find two eTargets in the target-span: “Rushdie” himself plus his act of “insulting.”

In the same sentence, there is another negative attitude, *insulting*, as shown in Ex(2). The source is *Salman Rushdie* and the target is *the Prophet*. Note that the span covering this event is the target span of the attitude in Ex(1) — the private state of Ex(2) is nested in the private state of Ex(1). Thus, the complete interpretation of the negative attitude in Ex(2) is: according to the writer, the Imam is negative toward Rushdie insulting the Prophet. The nested source is *w, Imam, Rushdie*.

Ex(2) When the Imam issued the fatwa against Salman Rushdie for insulting the Prophet ...

DS: insulting
nested-source: *w, imam, rushdie*
attitude: insulting
attitude-type: sentiment-negative
target: the Prophet
eTarget: Prophet

We add an eTarget for the Prophet, anchored to the head “Prophet.” Interestingly, “Prophet” is an eTarget for *w,Imam,Rushdie* (i.e., Rushdie is negative toward the Prophet), but not for *w,Imam* (i.e., the Imam is not negative toward the Prophet).

In the following example, the target span is short.

Ex(3) He is therefore planning to trigger wars ...

DS: (entire sentence)
nested-source: *w*

attitude: planning to trigger wars
attitude-type: sentiment-negative
target: He
eTarget: He
eTarget: planning, trigger, wars

“He” is George W. Bush; this article appeared in the early 2000s. The writer is negative toward Bush because (the writer claims) he is planning to trigger wars. As shown in the example, the MPQA 2.0 target span is only “He,” for which we do create an eTarget. But there are three additional eTargets, which are not included in the target span. The writer is negative toward Bush planning to trigger wars; we make sense of this by inferring that the writer is negative toward the idea of triggering wars and thus toward war itself.

Ex(4) Three leading international organisations warned jointly Thursday that the international fight against terrorism should not be a pretext for the violation of human rights.

DS: warned
nested-source: *w, threeint*
attitude: warned
attitude-type: sentiment-negative
target: the international ... rights
eTarget: be, pretext, violation
ESE: pretext
nested-source: *w, threeint*
polarity: negative
eTarget: pretext

The viewpoints in the article of Ex(4) are not against fighting terrorism (another sentence begins “While we recognize that the threat of terrorism requires specific measures ...”) but against doing so in certain ways. Here the three organizations are against the fight being used as a pretext for civil rights violations. Thus, “be”, “pretext”, and “violation” are eTargets, but “fight” and “terrorism” are not. We mark “be” as an eTarget because the source is negative toward the **state** of the fight being a pretext for the violation of human rights. This makes sense with the source also being negative toward “pretext” and “violation.” The fact that “pre-

text” is identified as a negative ESE annotation in the MPQA 2.0 supports this as well.

There is a difference between ESE and attitude eTarget annotations. Since ESE annotations pinpoint specific wording used to express subjectivity, ESE eTargets are annotated more narrowly than attitude eTargets. For ESEs, the eTargets are the entities and events that are directly evaluated by the expression, while, for attitudes, the eTargets include all entities and events toward which the attitude holds (as we saw in the examples above). For example:

Ex(5) ... because the hard-line wing in the US administration comprising Vice President Dick Cheney ...

DS: (entire sentence)
nested-source: w
attitude: hard-line
attitude-type: sentiment-negative
target: wing in the .. Dick Cheney
eTarget: wing, Cheney
ESE: hard-line wing
nested-source: w
polarity: negative
eTarget: wing

The ESE has only one eTarget, “wing,” while the attitude has two: “wing” and “Cheney.”

2.2 MPQA 3.0 Annotation Scheme

An **eTarget** is an entity or event that is the target of a sentiment (identified in MPQA 2.0 by a sentiment attitude or polar ESE span). The eTarget annotation is anchored to the head word of the NP or VP that refers to the entity or event, and has **three slots**: *id* (unique within the document), *isNegated* (*yes* or *no*), and *type* (*entity* or *event*; note that *event* includes both states and events). The *isNegated* = *yes* option is for the case where the eTarget is the negation of the event referred to by the head word, for example, when the source is positive toward someone **not** doing something.

An **attitude** has one or more target-span annotations in MPQA 2.0. We provide two slots for the k^{th} target annotation. *k-targetSpan* shows the k^{th} target span. *k-eTarget-link* is to be filled with a list of ids of eTargets whose text anchors are within the k^{th}

target span. An additional slot *new-eTarget-link* is to be filled with a list of ids of other eTargets.

Each eTarget of an **ESE** has two slots, one for the eTarget id, and one for an attribute, *isReferredInSpan* (*yes*, or *no*). The value is *yes* if the eTarget is referred to in the ESE span.

3 Agreement Study

We developed the manual via iterative annotation, discussion, and revision. Once the manual was developed, we participated in an agreement study.

For the formal agreement study, one document was randomly selected from each of the four topics of the OPQA subset (Stoyanov et al., 2005) of the MPQA corpus. They were not any of the documents used to develop the manual. We then independently annotated the four documents. There are 292 eTargets in the four documents in total.

To evaluate the results, the same agreement measure is used for both attitude and ESE eTargets. Given an attitude or ESE, let set *A* be the set of eTargets annotated by annotator *X*, and set *B* be the set of eTargets annotated by annotator *Y*. Following (Wilson and Wiebe, 2003; Johansson and Moschitti, 2013), which treat each set *A* and *B* in turn as the gold-standard, we calculate the average F-measure, denoted $agr(A, B)$. The $agr(A, B)$ is 0.82 on average over the four documents, showing good agreement: $agr(A, B) = (|A \cap B|/|B| + |A \cap B|/|A|)/2$.

4 Disagreement Analysis

One issue is whether an attitude toward an entity or event is indeed communicated in the sentence. Consider this sentence: “President **Mugabe’s reelection** has been *praised by* OAU.” The OAU is positive toward “reelection,” which is an eTarget both annotators mark. The question is whether it is also communicated in this sentence that the OAU is also positive toward “Mugabe.” *X* did not mark Mugabe as an eTarget, whereas *Y* did. During the subsequent discussion, *X* now agrees that it should be marked. In general, *X* was using what we now consider to be a too conservative policy. Overall, 29% of all disagreements are of this type of borderline case.

8% of the disagreements arise when there are multiple attitudes with overlapping spans, the same source, the same polarity, but different targets and

intensities⁴. When there are new eTargets which are not in any target span, annotator X splits the new eTargets into different attitudes based on the intensity, while annotator Y adds the new eTargets to all the attitudes regardless of intensity. Later the annotators discuss to decide which attitude each new eTarget should be linked to in the final version.

31% of the disagreements are caused by negligence, meaning an annotator realized, during later discussion, that she should have included an eTarget when she saw that the other annotator had included it.

The remaining disagreements are due to annotator mistakes such as filling in the wrong id.

5 Current Corpus

The current corpus consists of 70 documents, including the subset of the documents in MPQA 2.0 that come from English-language sources (i.e., that are not translations) and a subset of the OPQA subset in MPQA 2.0. A subset contains consensus annotations of X and Y and the rest were annotated by Y . The 70 documents have 1,029 ESEs, 1,287 attitudes, and 1,213 target spans of attitudes (excluding the target span that are marked as “none”) from MPQA 2.0; they have 4,459 eTargets in total. We added 1,366 eTargets to the ESEs and 1,608 eTargets to the target spans. We added 1,485 eTargets which are not in any target span.

6 An Example

In this section, we present an example from the OPQA subset (Stoyanov et al., 2005) to demonstrate how eTargets could help to automatically answer a question. There are opinion and fact questions for each document in the OPQA subset. The sentence below is annotated in MPQA 2.0 to answer the question, “Is the US Annual Human Rights Report received with universal approval around the world?” Here the writer is negative toward the report.

It is due to this hegemony, which the United States wants to maintain, that its State Department makes an assessment of the human rights situation in different

countries and prepares a report on their violations all over the world.

The annotations in MPQA 2.0:

S1: ⟨writer-US, positive, hegemony⟩

S2: ⟨writer, negative, the United States⟩

ESE1: ⟨writer, negative, N/A⟩

First, it is possible for a state-of-the-art system to be trained to recognize the sentiment S1, by the *maintaining* phrase and syntax information. But it would be difficult to find S2. There is no direct sentiment modifying the US, nor is there any sentiment or ESE annotation toward *maintain* or *hegemony* in MPQA 2.0. Now, in MPQA 3.0, we add the eTarget of the ESE1, so that it becomes ⟨writer, negative, hegemony⟩. This is a critical step, because the complete ESE bridges the two sentiments together.

Second, even though we have the two sentiments and the ESE, there is still a gap between *the United States* in the sentence and *report* in the question. One of the eTargets we add is “report.” It is more feasible for a co-reference system to recognize *report* in both the sentence and the question as the same thing, than recognizing that *the United States* and *report* refer to the same concept.

Third, in this sentence, according to the newly added eTargets, the system knows the writer is negative toward both *the United States* and *State Department*. When building a knowledge base about the human rights report, this reveals that the two entities have the same stance toward this topic, even without any world knowledge.

7 Conclusion

This paper presents an annotation scheme for adding entity and event target annotations to the MPQA corpus. A subset of MPQA has already been annotated according to the new scheme. We believe that the corpus will be a valuable new resource for developing entity/event-level sentiment analysis systems to facilitate NLP applications such as Automatic Question Answering.

Acknowledgements. This work was supported in part by DARPA-BAA-12-47 DEFT grant #12475008. We thank the anonymous reviewers for their helpful comments.

⁴In MPQA 2.0, an attitude is marked with an intensity (low, medium, or high) representing the intensity.

References

- Yejin Choi and Claire Cardie. 2008. Learning with compositional semantics as structural inference for sub-sentential sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 793–801. Association for Computational Linguistics.
- Lingjia Deng, Yoonjung Choi, and Janyce Wiebe. 2013. Benefactive/malefactive event and writer attitude annotation. In *ACL (2)*, pages 120–125.
- Minqing Hu and Bing Liu. 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.
- Richard Johansson and Alessandro Moschitti. 2013. Relational features in fine-grained opinion analysis. *Computational Linguistics*, 39(3):473–509.
- Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. In *Annual Meeting-Association For Computational Linguistics*, volume 45, page 432. Citeseer.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.
- Lev Ratinov, Dan Roth, Doug Downey, and Mike Anderson. 2011. Local and global algorithms for disambiguation to wikipedia. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1375–1384. Association for Computational Linguistics.
- Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1631–1642. Citeseer.
- Swapna Somasundaran, Theresa Wilson, Janyce Wiebe, and Veselin Stoyanov. 2007. Qa with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news. In *International Conference on Weblogs and Social Media*, Boulder, CO.
- Veselin Stoyanov, Claire Cardie, and Janyce Wiebe. 2005. Multi-Perspective Question Answering using the OpQA corpus. In *Proceedings of the Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, pages 923–930, Vancouver, Canada.
- Ivan Titov and Ryan T McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *ACL*, volume 8, pages 308–316. Citeseer.
- Peter D Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language ann. *Language Resources and Evaluation*, 39(2/3):164–210.
- Theresa Wilson and Janyce Wiebe. 2003. Annotating opinions in the world press. In *Proceedings of the 4th ACL SIGdial Workshop on Discourse and Dialogue (SIGdial-03)*, pages 13–22.
- Theresa Wilson. 2007. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of private states*. Ph.D. thesis, Intelligent Systems Program, University of Pittsburgh.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136. Association for Computational Linguistics.