

Project Proposal: Veridicality Assessment

Roger Que
query@jhu.edu

Pushpendre Rastogi
pushpendre@jhu.edu

We propose two projects concerning veridicality assessment; i.e., the determination of whether a potentially counterfactual event referenced in some text actually occurred.

Our first task will be to build an automatic tagger for veridicality judgments following the method presented in de Marneffe et al. (?), using a regularized log-linear model on selected features in a supervised setting. We note here that de Marneffe et al. annotated only a selected subset of the sentences in FactBank, and used a different annotation scheme; instead of asking annotators for the believed veridicality judgment of the source, as in (?), they asked annotators to provide their own veridicality judgment, presuming that it lined up with the judgment of the sentence author (? , 306).

One potential diagnostic of our tagger's performance is to check our results against those published in de Marneffe et al., part of which is available online.¹ Although we would like to attempt a direct reproduction of these results with our software, some important information, such as a comprehensive list of the features that were used, has not been published. Thus, without direct contact with the paper's authors, which may not be feasible within the remainder of the semester, it is unlikely that we will be able to perform an exact reproduction.

After building a tagger, we will next turn to the creation of a second corpus of veridicality information similar to FactBank, based on sample tweets retrieved from the Twitter Streaming API.² We anticipate that the construction process will proceed as

follows:

1. A cursory reading is performed to determine whether the tweet in fact contains a mention of an event. Tweets that do not contain events (for example, value judgments like *linguistics is awesome*³) are set aside and not used in future steps, although for the sake of completeness they are not removed from the data set. This step may be partially automatable, for instance by using the presence of copular verbs as a diagnostic for non-action sentences like the above, but all such automated judgments will be checked by a human before continuing.
2. Events are annotated out-of-frame in the TimeML format, using the standard Callisto and Tango tools, according to the TimeML annotation guidelines. This corresponds to the level of annotation in TimeBank.
3. Veridicality information is added, according to the FactBank annotation guidelines. We have not yet been able to find a publicly available annotator, but can construct one if necessary.

We will need to address questions of size, scope, and audience. The tweets we decide to use will change depending on these factors. For example, using unnormalized, ungrammatical tweets may require us to perform large amounts of preprocessing. Tweets by news agencies, on the other hand, will have more regularity and a higher likelihood of actually referring to some event (as opposed to a value

¹<http://christopherpotts.net/ling/data/factbank/>

²<https://dev.twitter.com/docs/api/1.1/get/statuses/sample>

³This in itself may be a difficult judgment; for example, a sentence like *Romney killed the Republicans' chances at regaining the White House* appears at first glance to contain an event, but may be better categorized as a value judgment.

judgment), meaning that we will be able to devote more of our time to the actual task of annotation. We may also choose to restrict ourselves to a certain topic, such as bombings, riots, strikes, etc. Although it would be ideal to stratify our samples across the various veridicality categories, we note that FactBank is heavily skewed towards CT+ and Uu judgments, and so our corpus may follow suit.

References

- Marie-Catherine de Marneffe, Christopher D Manning, and Christopher Potts. 2012. Did it happen? The pragmatic complexity of veridicality assessment. *Computational linguistics*, 38(2):301–333.
- Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227–268.