ClearTAC: Verb Tense, Aspect, and Form Classification Using Neural Nets

Skatje Myers
University of Colorado at Boulder
Boulder, CO, USA
skatje.myers@colorado.edu

Martha Palmer
University of Colorado at Boulder
Boulder, CO, USA
martha.palmer@colorado.edu

Abstract

This paper proposes using a Bidirectional LSTM-CRF model in order to identify the tense and aspect of verbs. The information that this classifier outputs can be useful for ordering events and can provide a pre-processing step to improve efficiency of annotating this type of information. This neural network architecture has been successfully employed for other sequential labeling tasks, and we show that it significantly outperforms the rule-based tool TMV-annotator on the Propbank I dataset.

1 Introduction

Identifying the tense and aspect of predicates can provide important clues to the sequencing and structure of events, which is a vital part of numerous down-stream natural language processing applications.

Our long term goal is to augment Abstract Meaning Representations (Banarescu et al., 2013) with tense and aspect information. With the assumption that an automatic pre-processing step could greatly reduce the annotation effort involved, we have been exploring different options for English tense and aspect annotation.

In this paper we compare two approaches to automatically classifying tense (present, past, etc.), aspect (progressive, perfect, etc.), and the form of verb (finite, participle, etc.). Our own work trains a BiLSTM-CRF NN, ClearTAC, on the PropBank annotations (Palmer et al., 2005) for the form, tense, and aspect of verbs. We compare the results to TMV-annotator, a rule-based system developed by (Ramm et al., 2017). Not surprisingly, we find our NN system significantly outperforms the rule-based system on the Propbank test data. In Section 2 we discuss related work and provide background information on TMV-annotator. Section 3 reviews the PropBank annotation and our modifications to the test data aimed at ensuring an apples to apples comparison with TMV-annotator. Section 4 describes the system architecture for ClearTAC, and Section 5 presents the experimental results for both systems, a comparison, and error analysis. We conclude in Section 6 and outline our plans for further development.

2 Background

Abstract Meaning Representations (AMRs) (Banarescu et al., 2013) are a graph-based representation of the semantics of sentences. They aim to strip away syntactic idiosyncrasies of text into a standardized representation of the meaning. The initial work on AMRs left out tense and aspect as being more syntactic features than semantic, but the absence of this feature makes generation from AMRs and temporal reasoning much more difficult. Very recently there have been efforts underway to extend AMRs to incorporate this type of temporal information (Donatelli et al., 2018). Since existing AMR corpora will need to be revised with annotations of this type of information, automatically classifying the tense and aspect of verbs could provide a shortcut. Annotators can work much more efficiently by only checking the accuracy of the automatic labels instead of annotating from scratch. Availability of automatic tense and aspect tagging could also prove useful for any system interested in extracting temporal sequences of events, and has been a long-standing research goal.

Much of the previous work on tense classification has been for the purpose of improving machine translation, including (Ye and Zhang, 2005) and (Ye et al., 2006), which explored tense classification of Chinese as a sequential classification task, using conditional random fields and a combination of surface and latent features, such as verb

Proceedings of the First International Workshop on Designing Meaning Representations, pages 136–140
Florence, Italy, August 1st, 2019

telicity, verb punctuality, and temporal ordering between adjacent events.

The NLPWin pipeline (Vanderwende, 2015) consists of components spanning from lexical analysis to construction of logical form representations to collecting these representations into a knowledge database. Tense is included as one of the attributes of the declension of a verb. This system is a rule-based approach, as is TMV-annotator described below.

Other recent work on tense classification includes (Reichart and Rappoport, 2010) attempting to distinguish between the different word senses within a tense/aspect. (Ferreira and Pereira, 2018) performed tense classification with the end goal of transposing verb tenses in a sentence for language study.

2.1 TMV-annotator

TMV-annotator (Ramm et al., 2017) is a rule-based tool for annotating verbs with tense, mood, and voice in English, German, and French. In the case of English, it also identifies whether the verb is progressive.

Although the rules were hand-crafted for each language, they operate on dependency parses. The authors specifically use the Mate parser (Bohnet and Nivre, 2012) for their reported results, although the tool could be used on any dependency parses that use the same part of speech and dependency labels as Mate. The first step of their tool is to identify verbal complexes (VCs), which consist of a main verb and verbal particles and negating words. Subsequent rules based on the words in the VC and their dependencies make binary decisions about whether the VC is finite, progressive, active or passive voice, subjunctive or indicative, as well as assign a tense. A subset of output for an example sentence is shown in Table 1.

For tense tagging, the authors report an accuracy of 81.5 on randomly selected English sentences from Europarl. In Section 5.2, we evaluate TMV-annotator on the Propbank I data and compare it to ClearTAC.

3 Data

3.1 Propbank I

The first version of Propbank, PropBank I, (Palmer et al., 2005) annotated the original Penn Treebank with semantic roles, roleset IDs, and inflection of each verb.

Sentence	The finger-pointing has already begun.
Verbal complex	has begun
Main	begun
Finite?	yes
Tense	present perfect
Progressive?	no

Table 1: Partial output of TMV-annotator for an example verbal complex, showing the fields relevant to this work.

The information in the inflection field consists of form, tense, aspect, person, and voice. We trained our model to predict form, tense, and aspect, which were labeled in the dataset with the following possible values:

- Form:
  - infinitive (i)
  - finite (v)
  - gerund (g)
  - participle (p)
  - none (verbs that occur with modal verbs)
- Tense:
  - present (n)
  - past (p)
  - future (f)
  - none
- Aspect:
  - perfect (p)
  - progressive (o)
  - both (b)
  - none

Not all combinations of these fields are valid. For instance, gerunds, participles that do not occur with an auxiliary verb, and verbs that occur with a modal verb are always tenseless and aspectless. Table 2 shows example Propbank I annotations.

We removed 13 files from our training/development sets, which seem to have been overlooked during original annotation. In total, the data contains 112,570 annotated verb tokens, of which the test set consists of 5,273 verb tokens.

Roleset ID	Form	Tense	Aspect
come.01	finite	past	-
halt.01	participle	past	progressive
trade.01	gerund	-	-

Table 2: Example Propbank I annotation for the sentence: At 2:43 p.m. EDT, came the sickening news: The Big Board was halting trading in UAL, “pending news.”

3.2 Reduced Propbank I

The goals of the TMV-annotator tool (described in Section 2) do not perfectly match with the annotation goals of Propbank I. Therefore, we created a reduced version of the Propbank I data to avoid penalizing the tool for using a different annotation schema. The changes are as follows:

- Remove gerunds.
- Ignore tense for participles that occur with an auxiliary verb. TMV-annotator assigns only aspect, whereas Propbank assigns both.
- Remove standalone participles that occur without an auxiliary verb. For example: ”Some circuit breakers **installed** after the October 1987 crash failed their first test.”

This reduces the number of verbs in the dataset to 92,686, of which 4,486 are in the test set.

4 ClearTAC System Architecture

Bidirectional LSTM-CRF models have been shown to be useful for numerous sequence labeling tasks, such as part of speech tagging, named entity recognition, and chunking (Huang et al., 2015). Based on these results, we expected good performance on classification of tense and aspect. Our neural network consists of a Bi-LSTM layer with 150 hidden units followed by a CRF layer. The inputs to the NN were sentence-length sequences, with each token represented by pre-trained 300-dimension GloVe embeddings (Pennington et al., 2014). No part-of-speech or syntactic pre-processing was used. Classifying form, tense, and aspect was treated as a joint task.

5 Results

Our model was evaluated on both the full and reduced Propbank I datasets, as described in Section 3. The results are presented in Table 3.

Full Propbank I			
	P	R	F1
Verb identification	94.30	97.25	95.75
Form	92.61	95.51	94.03
Tense	92.66	95.56	94.09
Aspect	93.88	96.81	95.32
Form + tense + aspect	92.04	94.92	93.46
Reduced Propbank I			
	P	R	F1
Verb identification	96.20	97.10	96.65
Form	95.36	96.26	95.81
Tense	95.30	96.19	95.74
Aspect	96.05	96.95	96.49
Form + tense + aspect	95.19	96.08	95.63

Table 3: Evaluation of our system on Propbank I.

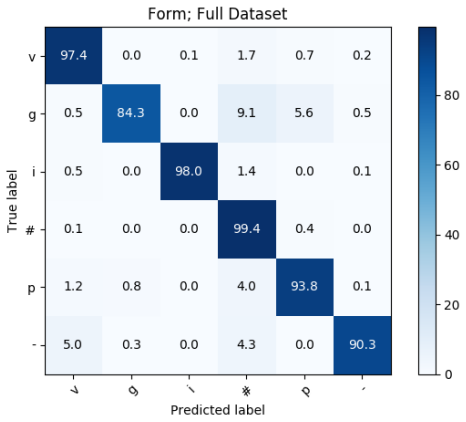

Figure 1: Confusion matrix of our model output for verb **form** on the full Propbank dataset. See Section 3.1 for a legend. ’#’ is the label for a non-verb token.

Performance across the board for the various subtasks on both datasets was consistently in the mid-90’s. The more challenging task of tagging all forms, tenses, and aspects in Propbank I saw a performance decrease of only 2 points compared to the reduced dataset.

5.1 Error Analysis

Overall, the model had the most challenges with gerunds and verbs with modals, often predicting them not to be a verb. With these forms also being tenseless, the effect can also be seen in the high number of gold ”no tense” labels being misclassified as not a verb.

Figures 1, 2, and 3 show confusion matrices for the model’s output for each of the three subtasks on the full Propbank I dataset.

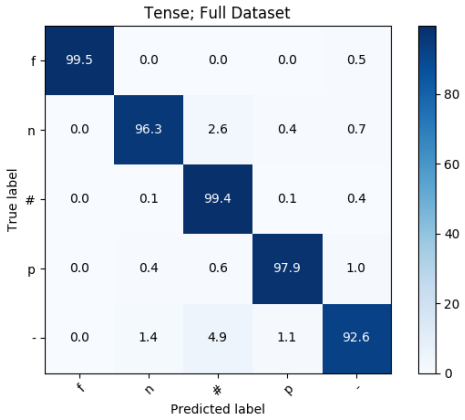

Figure 2: Confusion matrix of our model output for verb **tense** on the full Propbank dataset. See Section 3.1 for a legend. '#' is the label for a non-verb token.

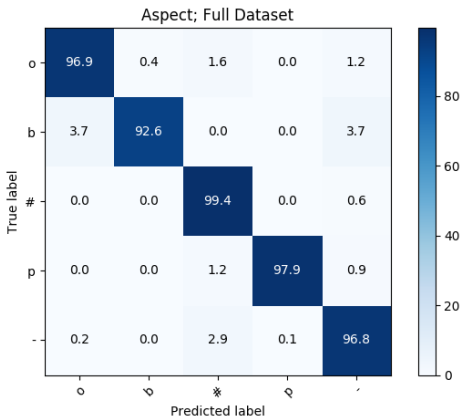

Figure 3: Confusion matrix of our model output for verb **aspect** on the full Propbank dataset. See Section 3.1 for a legend. '#' is the label for a non-verb token.

Full Propbank I			
	P	R	F1
Verb identification	95.68	77.77	85.80
Form	76.46	62.15	68.56
Tense	85.74	69.69	76.89
Aspect	93.56	76.05	83.90
Form + tense + aspect	70.79	57.54	63.48
Reduced Propbank I			
	P	R	F1
Verb identification	94.14	89.95	92.00
Form	76.22	72.83	74.49
Tense	76.43	73.03	74.69
Aspect	92.70	88.56	90.58
Form + tense + aspect	75.64	72.27	73.92

Table 4: Evaluation of TMV-annotator on the complete and reduced Propbank I test sets.

5.2 Comparison with TMV

As described in Section 2, the TMV-annotator tool (Ramm et al., 2017) is a rule-based tool for annotating tense, aspect, and mood in English, French, and German. We ran this tool on the output of the Mate dependency parser (Bohnet and Nivre, 2012) (which the tool was designed in mind of) using a pre-trained model and evaluated on both the complete Propbank I test data, which includes verb forms that TMV-annotator was never intended to annotate, such as gerunds, as well as the reduced Propbank I test set as described in Section 3.2, which only contains the intersection of TMV-annotator and Propbank I annotations. The results of this are presented in Table 4.

Unsurprisingly, TMV-annotator is only able to reach a F-score of 63.48 on the whole task on the full dataset. As would be expected in this circumstance, the recall is much lower than precision.

On the Reduced Propbank I dataset, TMV-annotator performs significantly better, but still falls over 20 points shy of our NN system. Simply the misidentification of verbs in the data, likely due to parsing errors, drops the F-score a full 8 points. Notably, TMV-annotator achieves an F-score in the 90s on the subtask of classifying aspect, while form and tense prove to be more challenging, with F-scores near 75.

6 Conclusions and Future Work

Our NN model outperformed the rule-based TMV-annotator when annotating the same subset of verb

form, tense, and aspect by 21.71 points. Furthermore, this model achieved a F-score of 93.46 on the more challenging task of classifying the full label set of form, tense, and aspect present in Propbank I. The performance of this model makes it a feasible pre-processing step to add tense annotation to Abstract Meaning Representations.

There are a number of architectural or feature improvements left for future work. Embeddings such as ELMo or Bert could possibly help with performance on out-of-vocabulary words as well as help distinguish between identical verb forms, such as gerunds and present-tense verbs, due to incorporating context. Better performance may also be possible by dividing the subtasks of classifying form, tense, and aspect, rather than treating it as a single joint task.

Another dataset which has been annotated with tense and aspect is TimeML (Pustejovsky et al., 2003). Evaluation of our system on this data would be complementary to this work and is planned for future work.

Acknowledgments

We gratefully acknowledge the support of NSF 1764048 RI: Medium: Collaborative Research: Developing a Uniform Meaning Representation for Natural Language Processing. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DTRA or the U.S. government.

References

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.

Bernd Bohnet and Joakim Nivre. 2012. A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1455–1465. Association for Computational Linguistics.

Lucia Donatelli, Michael Regan, William Croft, and Nathan Schneider. 2018. Annotation of tense and aspect semantics for sentential AMR. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 96–108, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Kledilson Ferreira and Jr Álvaro R Pereira. 2018. Verb tense classification and automatic exercise generation. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*, pages 105–108. ACM.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

James Pustejovsky, José M Castano, Robert Ingria, Roser Sauri, Robert J Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R Radev. 2003. Timeml: Robust specification of event and temporal expressions in text. *New directions in question answering*, 3:28–34.

Anita Ramm, Sharid Loáiciga, Annemarie Friedrich, and Alexander Fraser. 2017. Annotating tense, mood and voice for english, french and german. *Proceedings of ACL 2017, System Demonstrations*, pages 1–6.

Roi Reichart and Ari Rappoport. 2010. Tense sense disambiguation: A new syntactic polysemy task. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 325–334, Cambridge, MA. Association for Computational Linguistics.

Lucy Vanderwende. 2015. Nlpwin–an introduction. Technical report, Microsoft Research tech report no. MSR-TR-2015-23.

Yang Ye, Victoria Li Fossum, and Steven Abney. 2006. Latent features in automatic tense translation between chinese and english. In *Proceedings of the fifth SIGHAN workshop on Chinese language processing*, pages 48–55.

Yang Ye and Zhu Zhang. 2005. Tense tagging for verbs in cross-lingual context: A case study. In *International Conference on Natural Language Processing*, pages 885–895. Springer.