NoUniteD-SRL: Enhancing Multilingual Semantic Role Labeling with Nominal Synsets

Marco Lo Pinto

1 Introduction

In the field of Natural Language Processing, a linguistic inventory is used for a variety of tasks, such as Semantic Role Labeling. The choice of a particular inventory with respect to another could depend on the coverage of word meanings, the type and number of clusters used to classify, the number of semantic roles and so on. One of them, called VerbAtlas (Di Fabio et al., 2019) offers, with respect to other inventories such as FrameNet (Baker, 2014) and PropBank (Palmer et al., 2005), cross-domain roles, semantically consistent frames and informative role labels.
One main problem concerning VerbAtlas is that it's only verb-specific. In order to improve it further, a carefully-selected dataset containing samples for both verbal and nominal synsets is needed. The aim of this paper is to show how (and to which extent) we can use different strategies in order to delegate all (or at least most of) the difficulties, that can be found in generating such dataset, from manual to automatic annotation. The general pipeline followed for this project is in figure 1. It must be noted that tests will be done only on the English language; nonetheless, all the steps can be easily extended to any language.

2 Starting Dataset

The first step is to gather a corpus of sentences that will be used as as basis for the starting dataset of this approach. As an example, we can use part of big unannotated corpus from Wikimedia dumps (Wikimedia) for the targeting language. However, for this task, the sentences used are extracted from the MAVEN dataset (Wang et al., 2020): this is because we need phrases that contain nominal events. Using the train dataset split, 32 431 unique phrases were collected.

3 SynsetExplorer: finding Nominal Synsets candidates

3.1 Approach

In VerbAtlas, each frame corresponds to a cluster of BabelNet synsets and each BabelNet synset can be associated with a WordNet synset. Using these informations, a connection between these three linguistic resources can be generated. Then, by exploiting NLTK (Loper and Bird, 2002), we can move from verbal to nominal synsets using their derivationally related forms (e.g. eat (v) $\rightarrow$ eating (n), eater (n)), thus obtaining nominal synset candidates. In order to go up in the graph using hypernyms and be sure to have explored every crucial part of it (a synset can have multiple hypernyms), a classic Depth First Search (Wikipedia, 2022) is applied, categorizing each encountered synset by the number of times that is explored and its definition. A visual representation of a VerbAtlas frame and its connection with WordNet is reported in figure 2.

3.2 Selection of best hypernyms

From the obtained list of synsets, its definiton and number of visits, the next step is to identify which of the derivationally related forms (and their hypernyms) are events. To do so, by ordering the explored synsets in descending order (i.e. from the most visited to the least visited) I classified the most predominant and pertinent synsets that represent events. The steps to do so are as follows:

1. Select the most predominant synsets that represent events (e.g. event.n.01 or act.n.02).
2. Explore again the graph and remove from the output all synsets that are hyponyms of the selected synsets.
3. Do again these steps until there are no more synsets that represent events.

3.3 Selection of candidate synsets

The graph is explored again using DFS so to obtain candidate synsets that are connected with a derivationally related form present in a VerbAtlas frame. The logic behind it is that if via recursive hypernymy a synset is in the predominant synsets (obtained in section 3.2), then that synset is a possible candidate for a VerbAtlas frame.

3.4 Results

Around 3974 nominal synsets were found and saved, with 42 VerbAtlas frames that are not connected with nominal events out of 432. The number of ambiguous candidates is around 1075. It must be noted that the definition of "ambiguous" is that of a nominal synset that can be in multiple frames. For example, the synset articulation.n.03 is connected via derivationally related forms with synsets in frames EXPLAIN, SPEAK and PRONOUNCE (as we can see in figure 3).

4 Custom SRL Model

Another main component of the pipeline is the custom model used for Semantic Role Labeling. It was designed with modularity in mind: each part can be used separately or jointly with the others, depending on the specific task to be done. It is divided into two main blocks:

1. Predicate Identification and Disambiguation: a XLM-RoBERTa based model (Conneau et al., 2019) is used in combination with linear layers. The input words are represented as [[CLS] sentence [SEP]]. Then, in order to improve the disambiguation part, the model was modified so to receive as optional input the result from the identification part.
2. Argument Identification and Classification: for the sake of modularity, this part implements another XLM-RoBERTa based model in combination with linear layers, but the way the text was encoded and passed onto it was inspired from another paper (Shi and Lin, 2019): the input is represented as [[CLS] sentence [SEP] predicate [SEP]], in order to focus the attention mechanism on the target predicate in the sentence.

The flow of the data for each possible input is reported in figure 5. This model will be used both for evaluating the final generated dataset and to disambiguate nominal synsets.

5 Generation of NoUniteD-SRL

5.1 Approach

For the purpose of creating the NoUniteD-SRL dataset (without the roles for the nominal synsets), the list of sentences computed at section 2 are passed as input for VerbAtlas and AMuSE-WSD (Orlando et al., 2021): the former is used with the purpose of obtaining the predicates and roles for the verbal synsets, whereas the latter is utilized so to find and identify the nominal synsets in each phrase. Then, for each nominal synset, if it is one of the possible unambiguous candidates obtained in section 3.4, we can associate it with the correct VerbAtlas frame. Finally, PoS tags, lemmas and dependency trees are added using Stanza (Qi et al., 2020) so to have more information to train any SRL model. The format of the generated dataset is reported in figure 4.

5.2 Comparisons with UniteD-SRL

The English split of the UniteD-SRL dataset (Tripodi et al., 2021) is used as comparison with the new generated dataset. As we can see from table 1, the number of predicates that are present in the latter are higher, even if we remove the nominal synset part and we limit the number of samples to be the same number as the former. Moreover, the custom SRL model was trained with both datasets, showing amazing results with the newly generated one, as we can see from figures 6 and 7 and in table 2.

6 Nominal Sysnset Disambiguation

As explained in 3.4, a synset can be in multiple frames.

To disambiguate nominal synsets, one possible solution is to select the frame from the synset explorer that exhibits the strongest connection to its derivationally related form (or its hypernyms in case of uncertainty). However, this approach is challenged by the unbalanced nature of frames, where some contain ten times more synsets than others. To address this issue, several approaches were tested, with some evaluated qualitatively due to the lack of an objective metric to measure the disambiguation performance.

One strategy involved using the predicate disambiguation component of a custom SRL model trained on the dataset generated in section 5. Firstly, a new dataset of sentences containing ambiguous nominal synsets was created. Then, by using the said part of the SRL model on the predicates in the sentences, the corresponding frame names are extracted. A first evaluation of this procedure using only the nominal part of the validation dataset showed very good scores (95.9% of f1-score). Nevertheless, it sometimes gives better results than the statistical approach (e.g. with the statistical method, destruction.n.02 is put in CANCEL_ELIMINATE, whereas the model chose DESTROY), but in various cases it wrongly predicts the frame. The result improves if the method is combined with the statistical approach described at the beginning of this section, but there is another main problem concerning this approach: in order to classify a synset, at least one phrase with that synset must be generated. Using the MAVEN dataset, only 559 ambiguous synsets were found out of 1075.

A different approach tested is the use of a zero-shot classification pipeline (Yin et al., 2019): the (pretrained) model used is BART (Lewis et al., 2019) after being trained on the MultiNLI (MNLI) dataset. In essence, given as input the definition of the synset and as labels the possible connected frames with that synset (extracted using the synset explorer), the model can construct an hypothesis and determine in which class the sentence should fall. This approach works only if the labels are few, so it must be used in combination with the synset explorer. Positive factors of using this strategy include the removal of the use of an intermediate dataset for the target synsets.

Lastly, another custom model was developed that utilizes the definitions of each synset as input and outputs the corresponding frame. Two experiments were conducted to evaluate the model. In the first one, the model was trained and tested only on nominal unambiguous synsets, with both the definition and the synset name provided as input. In the second experiment, it was trained with also the verbal synsets added in the training split and received only the definitions as input. The latter produced better results, as shown in figure 8, with 63% of f1-score. This approach circumvents the need for an intermediary dataset and avoids the use of the synset explorer, thereby exhibiting promising results.

7 Annotation of the roles for the nominal event synsets

Up until now the generated dataset in section 5 lacks of roles for the nominal part: this is the last step to complete it. The first attempt involved the use of the argument identification and classification part of the custom model trained on the verbal part of the generated dataset. Unfortunately, this strategy didn't work, showing that the trained model isn't capable of producing roles for the nominal part. A different approach implemented is based on the idea that, in order to generalize between verbal and nominal samples, the target predicate is substituted with the VerbAtlas frame name. A confusion matrix is reported in figure 9. As we can see from table 3, the performances are slightly worse w.r.t the custom SRL model (around 1%). Nonetheless, it showed that it can sometimes generate correct roles for the input phrase.

8 Conclusions

As we have seen from each section of this paper, the creation of silver data from unannotated corpus is not an easy task. We have proved that it is possible to automatically generate an incomplete dataset (i.e. without the roles for the nominal synsets) using a combination of different models. Then, we have seen that there are several strategies in order to disambiguate nominal synsets, each of them with their advantages and disadvantages. Finally, we have seen that it is somewhat possible to add automatically roles for the nominal synsets samples. It must be noted that this last approach is still sub-optimal: this is because sometimes the model predicts incorrectly (or doesn't find) the targeting roles. To further improve the dataset, a group of experts could be employed to manually correct the samples using a dedicated web interface, as done for the generation of the UniteD-SRL dataset. This approach offers a more efficient and feasible alternative to creating the dataset from scratch.

References

Collin F. Baker. 2014. FrameNet: A knowledge base for natural language processing. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929-2014)*, pages 1–5, Baltimore, MD, USA. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. VerbAtlas: a novel large-scale verbal semantic resource and its application to semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 627–637, Hong Kong, China. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2019. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *CoRR*, abs/1910.13461.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *CoRR*, cs.CL/0205028.

Riccardo Orlando, Simone Conia, Fabrizio Brignone, Francesco Cecconi, and Roberto Navigli. 2021. AMuSE-WSD: An all-in-one multilingual system for easy Word Sense Disambiguation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 298–307, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. *CoRR*, abs/2003.07082.

Peng Shi and Jimmy Lin. 2019. Simple BERT models for relation extraction and semantic role labeling. *CoRR*, abs/1904.05255.

Rocco Tripodi, Simone Conia, and Roberto Navigli. 2021. UniteD-SRL: A unified dataset for span- and dependency-based multilingual and cross-lingual Semantic Role Labeling. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2293–2305, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. 2020. MAVEN: A Massive General Domain Event Detection Dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1652–1671, Online. Association for Computational Linguistics.

Wikimedia. Wikimedia Downloads. Accessed 2022-07-05.

Wikipedia. 2022. Depth-first search — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Depth-first%20search&oldid=1117448546`.

Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. *CoRR*, abs/1909.00161.

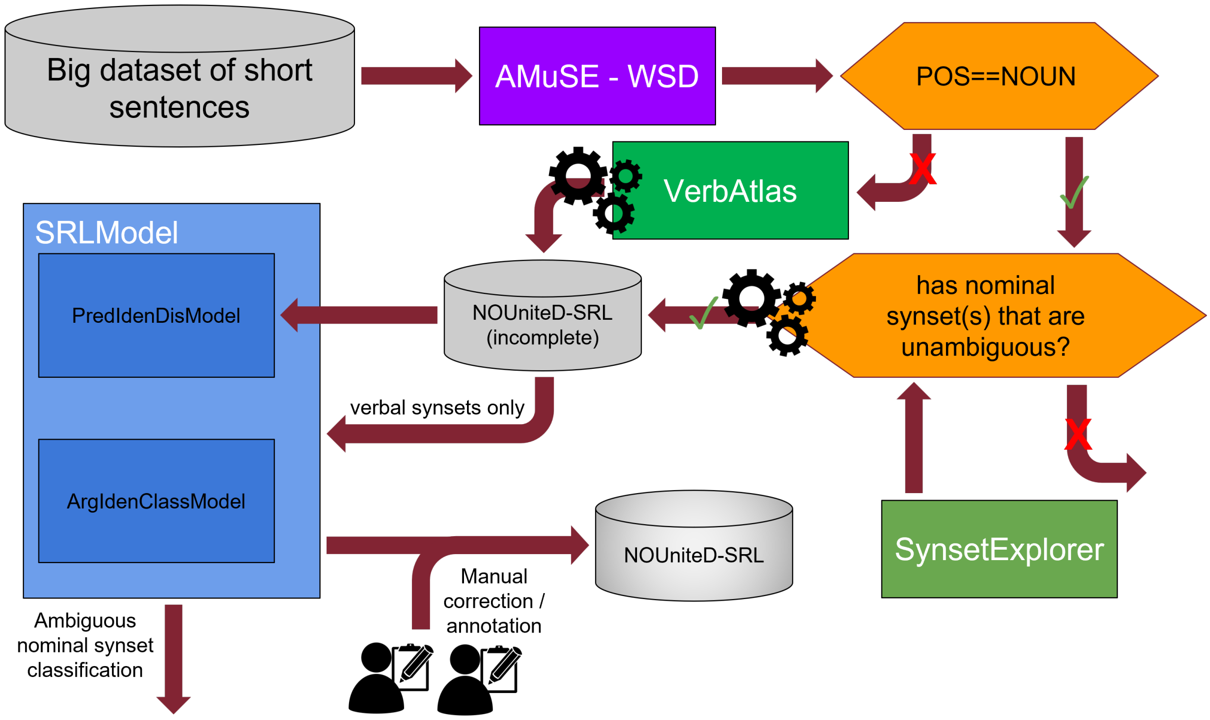

Figure 1: The general pipeline adopted for this study.

Dataset	# predicates		
	train	**development**	**total**
UniteD	302	214	306
NoUniteD	405	350	405
NoUniteD (only verbal)	397	335	397
NoUniteD (only verbal + limited)	370	242	374

Table 1: Comparisons between the UniteD-SRL dataset and the one generated in section 5.

Dataset	predicate		argument	
	F1-iden	**F1-disamb**	**F1-iden**	**F1-class**
UniteD	0.9440	0.8371	0.8973	0.8407
NoUniteD (only verbal)	0.9878	0.9487	0.9498	0.9139
NoUniteD	0.9749	0.9406	-	-

Table 2: Evaluations on the datasets using the custom SRL model.

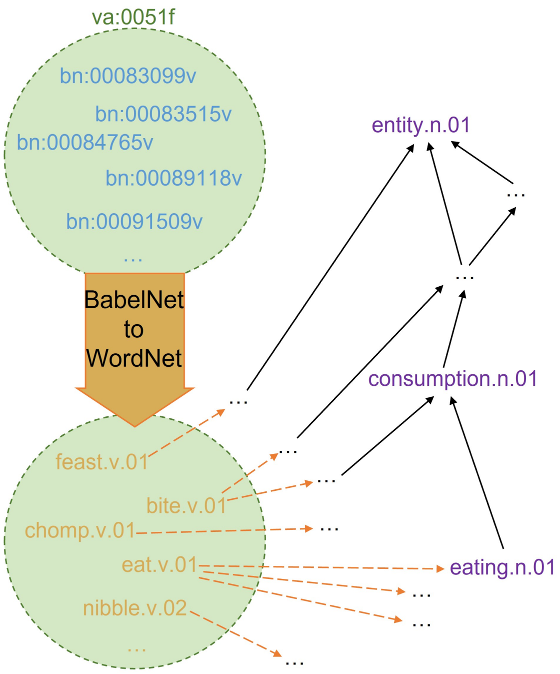

Figure 2: Visual representation of the mapping and exploring part for VerbAtlas (green), BabelNet (blue) and WordNet (brown), va:0051f is EAT-BITE VerbAtlas frame. Dotted arrows represent the semantically related forms connections, whereas uniform arrows represent hypernyms connections.

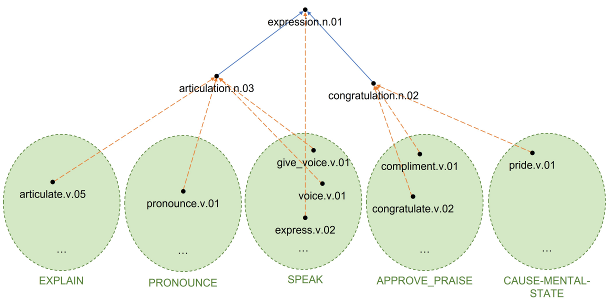

Figure 3: Visual representation of the articulation.n.03 synset ambiguity. Dotted arrows represent the derivationally related forms connections, whereas uniform arrows represent hypernyms connections.

	argument	
Dataset	**F1-iden**	**F1-class**
NoUniteD (only verbal)	0.9552	0.9066

Table 3: Results for the trained model for the task of adding roles for the nominal part.

```
[
    ...,
    {
        "words": ["Three","coaches","derailed","as","a","result","of","the", ...],
        "predicates": ["_","_","STOP","_","_","_","_","_", ...],
        "predicates_v": ["_","_","STOP","_","_","_","_", ...],
        "predicates_n": ["_","_","_","_","_","_","_","_", ...],
        "roles": {
            "2": ["_","theme","_","_","_","_","_","_", ...],
            "11": ["_","_","_","_","_","_","_","_", ...]
        },
        "roles_v": {
            "2": ["_","theme","_","_","_","_","_","_", ...],
            "11": ["_","_","_","_","_","_","_","_", ...]
        },
        "roles_n": {},
        "num_v": 2, # number of verbal synsets found (can be used for balancing the dataset)
        "num_n": 2, # number of nominal synsets found (can be used for balancing the dataset)
        "lemmas": ["three","coach","derail","as","a","result","of","the", ...],
        "pos_tags": ["NUM","NOUN","VERB","ADP","DET","NOUN","ADP","DET", ...],
        "dependency_heads": [2,3,0,6,6,3,9,9, ...],
        "dependency_relations": ["nummod","nsubj","root","case","det","obl","case","det", ...]
    },
    ...,
]
```

Figure 4: An example on how the final generated dataset is formatted.

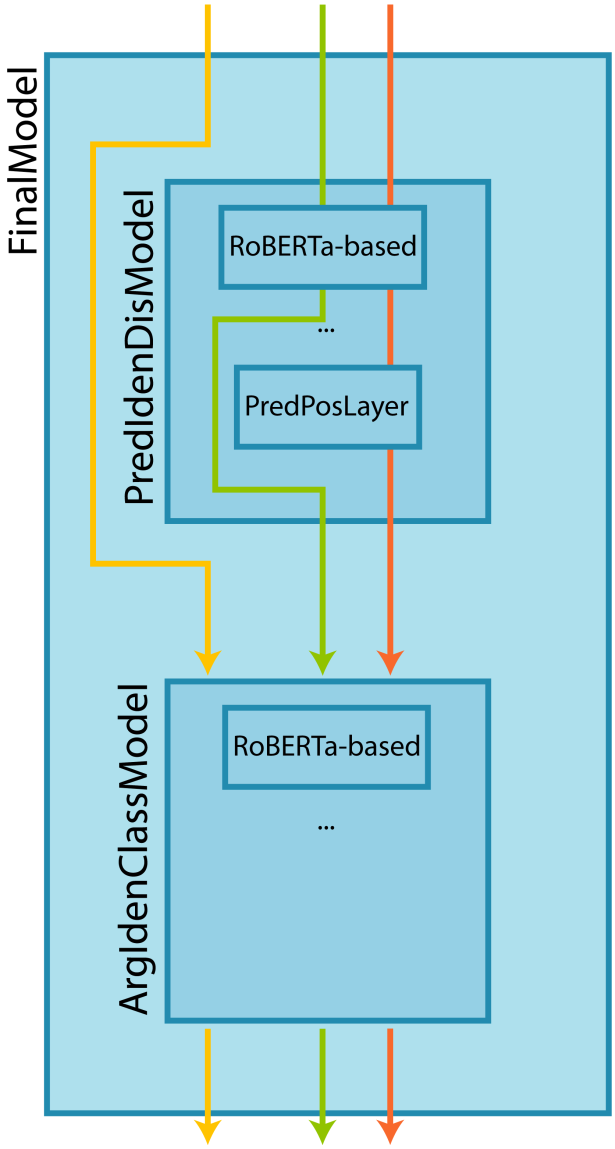

Figure 5: Custom SRL model flow. Orange = argument identification + argument classification; red = predicate disambiguation + argument identification + argument classification; green = predicate identification + predicate disambiguation + argument identification + argument classification

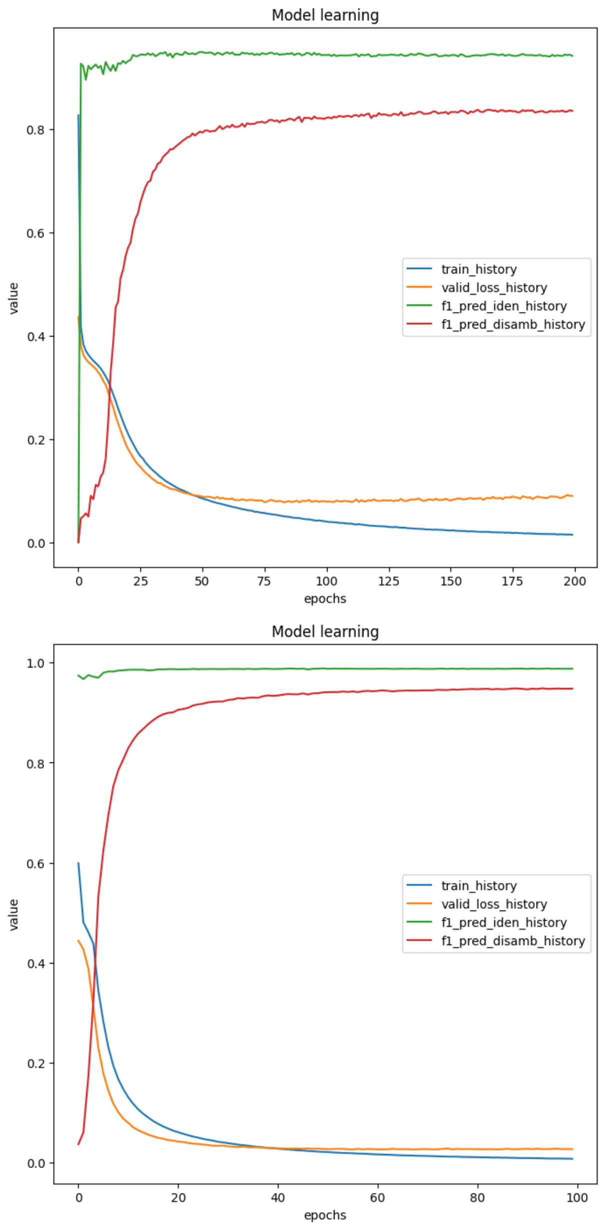

Figure 6: Training history of the custom model for predicate identification and disambiguation with the UniteD dataset (top) and with the new NoUniteD dataset (bottom, verbal split only).

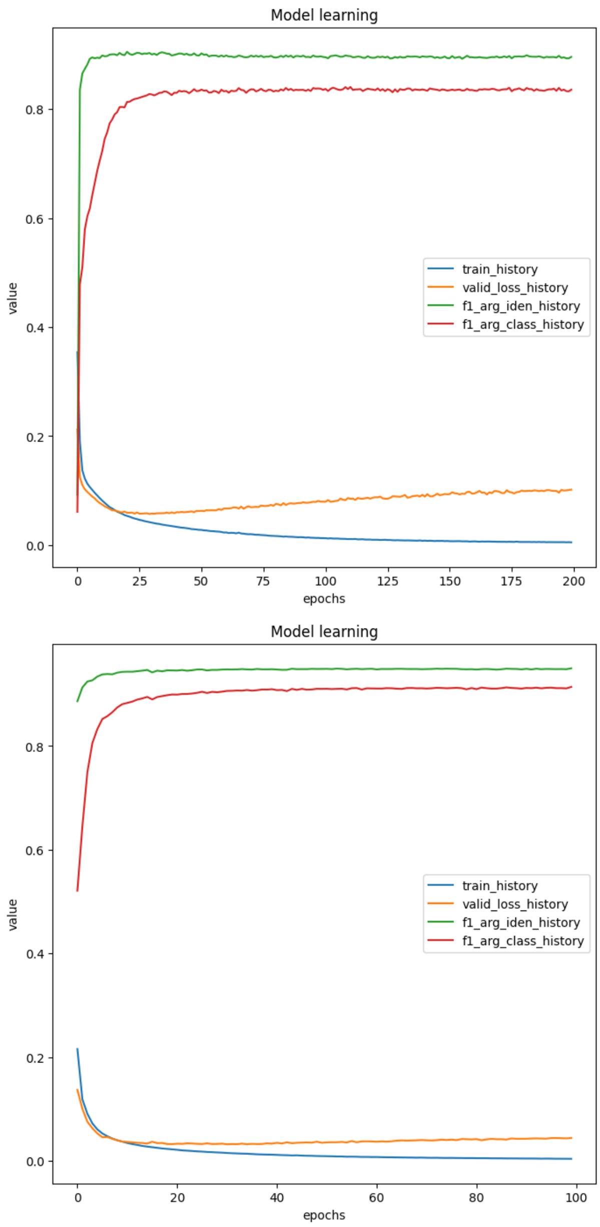

Figure 7: Training history of the custom model for argument identification and classification with the UniteD dataset (top) and with the new NoUniteD dataset (bottom, verbal split only).

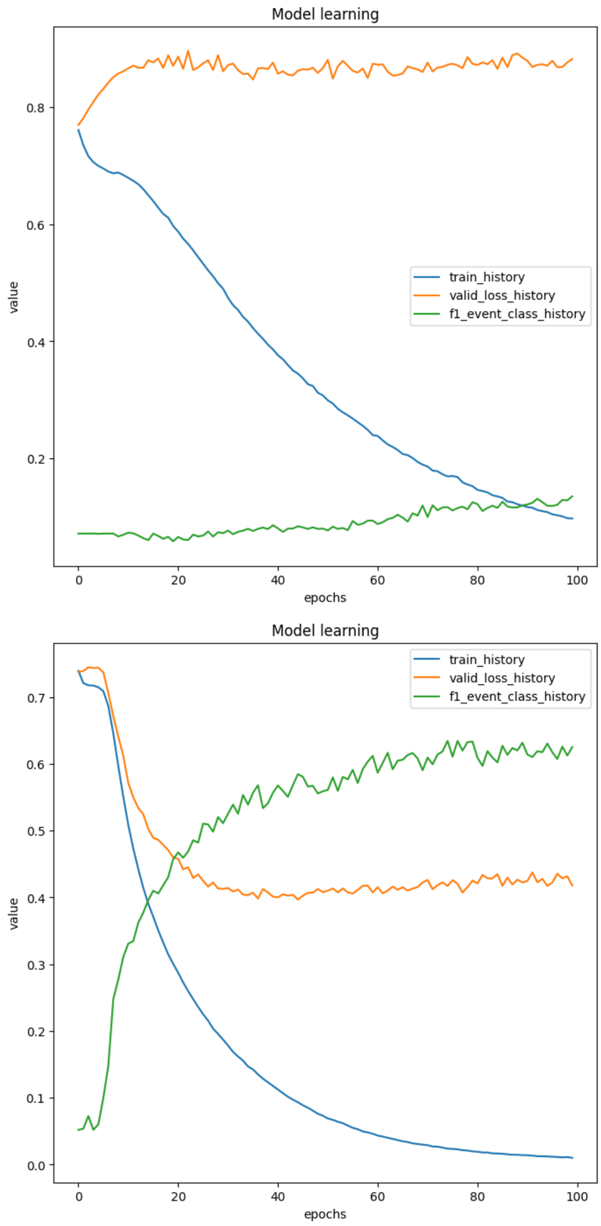

Figure 8: Top: training history of the nominal event classificaion model trained and evaluated with only the nominal synsets. Bottom: training history of the nominal event classificaion model trained with both verbal and nominal synsets and evaluated on nominal synsets.

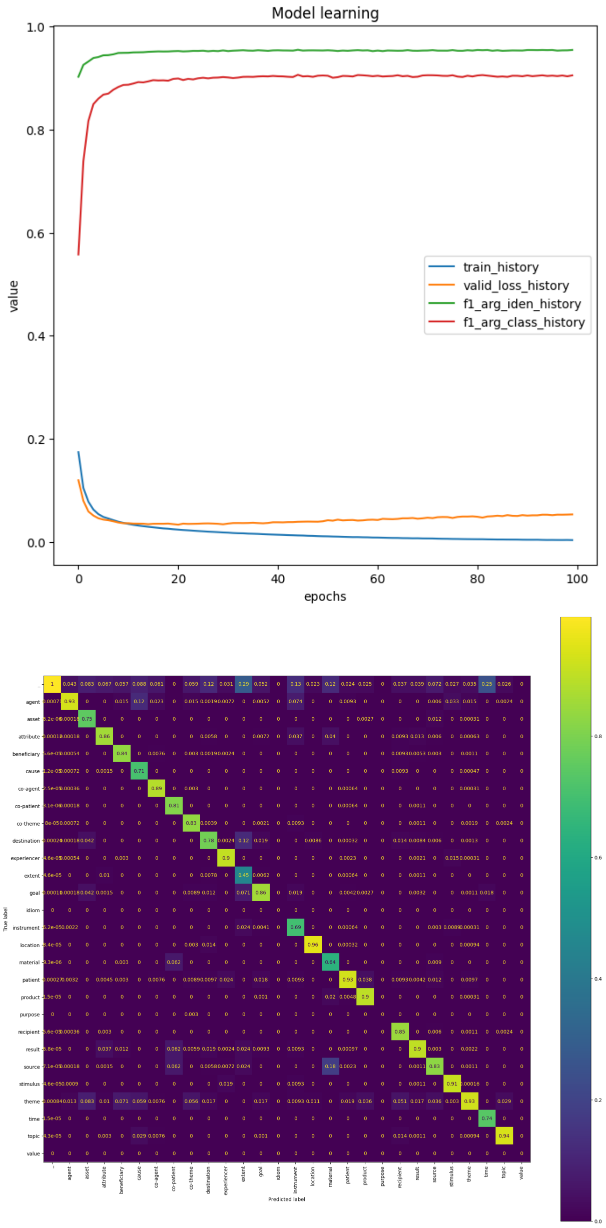

Figure 9: Train history and confusion matrix for the role adder model.