UIUC TAC 2020 RUFES System Description

Revanth Gangi Reddy, Manling Li, Haoyang Wen, Heng Ji
University of Illinois at Urbana-Champaign
{revanth3,hengji}@illinois.edu

1 Introduction

Our system is based on GAIA (Li et al., 2020), which has the ability to do knowledge extraction from both text and images and then perform cross-media knowledge fusion. Text knowledge extraction involves detection, co-reference and fine-grained typing for both entities and events in addition to fine-grained relation extraction. Visual knowledge extraction involves detection, linking and co-reference of entities in images. These are then combined to get a coherent structured multi-media KB, indexing entities, relations, and events, following a rich, fine-grained ontology. In our submission to RUFES, we use only a part of the GAIA, specifically the mention detection, entity linking, co-reference and fine-grained entity typing components, and adapt them to the RUFES ontology.

2 Task Description

Recognizing Ultra Fine-grained EntitieS (RUFES) is a shared task that extends entity extraction to a new fine-grained entity ontology that consists of approximately 200 fine-grained entity types that are representative of news data. Each entity type in the ontology has a one sentence definition along with some examples. Some sample types can be seen in Table 1.

Given an input document, the system is required to automatically identify an entity as a cluster of name, nominal, and/or pronominal mentions, and classify the entity into one or more of the types defined in the ontology. While the current task involves document-level entity discovery and within-document co-reference of entity mentions, future versions might extend to corpus-level co-reference. The development and evaluation source documents are drawn from a collection of Washington Post news articles from Jan 2012 to Dec 2019, which comprises approximately 100,000 articles.

3 System Description

The text knowledge extraction pipeline is as follows: First, entity mentions are extracted from the input documents, then the identical mentions are clustered together through entity linking and coreference resolution. Finally, the fine-grained typing model assigns the type to the mentions in the cluster. Each individual component is explained in brief below.

3.1 Mention Extraction

We apply a state-of-the-art joint entity, relation and event extraction system OneIE (Lin et al., 2020) to extract coarse-grained mentions. The OneIE system extracts the information network from a given sentence in four steps: encoding, identification, classification, and decoding. First, the input sentence is encoded using a pre-trained BERT encoder (Devlin et al., 2019). Next, entity mentions and event triggers are identified using a feedforward network which computes a score for each word. After that, type label scores are computed for all nodes and pairwise edges among them, with a conditional random field (CRF) layer capturing the dependencies between the predicted tags (e.g., an I-PER tag should not follow a B-GPE tag). Further, a beam search based decoder is used to explore possible information networks for the input sentence and return the one with the highest global score.

We compare the OneIE system against DyGIE++ (Wadden et al., 2019), a state-of-the-art end-to-end IE model that utilizes multi-sentence BERT encodings and span graph propagation. Table 2 reports numbers on two datasets, ACE05-R that includes named entity and relation annotations, and ACE05-E that includes entity, relation, and event annotations.

Type	Definition	Examples
App.CommunicationSoftware.SocialMedia	An application or website that facilitates the sharing of ideas, thoughts, and information ...	Facebook, Twitter, ...
Document.LegalDocument.Certificate	A document attesting to the truth of certain stated facts	Birth certificate, ...
FAC.Building.StoreShop	A facility where retail products are sold	CVS, Walmart, ...
IllHealth.Disease.CommunicableDisease	A disease that can be communicated from one person to another	Flu, Common cold, ...
ORG.CommercialOrganization.Firm	A business that provides professional services such as legal service, accounting, consulting, etc	Capitol CPA LLP, ...
Pathogen.Virus.InfluenzaVirus	RNA virus causing influenza in humans, some other mammals and birds	H1N1, H2N2, ...
Publication.Magazine.NewsMagazine	A magazine devoted to reports of current events	Newsweek, Time, ...

Table 1: Some sample fine-grained types in the ontology with their corresponding examples.

Dataset	Task	DyGIE++	OneIE
ACE05-R	Entity	88.6	**88.8**
	Relation	63.4	**67.5**
ACE05-E	Entity	89.7	**90.2**
	Trig-I	-	**78.2**
	Trig-C	69.7	**74.7**
	Arg-I	53.0	**59.2**
	Arg-C	48.8	**56.8**

Table 2: Comparison of OneIE with DyGIE++ on the ACE2005 datasets (F-score, %).

3.2 Entity Coreference Resolution

We first follow (Pan et al., 2017) to link entities to background KB and Freebase. The entities that are linked to the same KB entity are considered as coreferential. Then for entity coreference resolution, we implement a neural model similar to the `BERT-Coref` model (Joshi et al., 2019). However, there are several important differences. First, we remove the higher-order inference (HOI) layer (Lee et al., 2018) from the original architecture. Our preliminary results suggest that HOI typically does not improve the coreference resolution performance while incurring additional computational complexity. This observation agrees with a recent analysis of Xu and Choi (2020). Second, we also apply a simple heuristic rule based on the entity linking results to refine the predictions of the neural models. We prevent two entity mentions from being merged together if they are linked to different entities with high confidence. For English, we use SpanBERT (large) (Joshi et al., 2020) as the Transformer encoder and train the system on ACE 2005 (Walker et al., 2006), EDL 2016[1], EDL 2017[2], and OntoNotes (English) (Pradhan et al., 2012). Table 3 compares the performance of our model against a state-of-the-art method on the OntoNotes dataset. Figure 1 shows the overall architecture of our entity coreference resolution model.

[1]LDC2017E03

[2]LDC2017E52

Model	CoNLL
Previous SOTA (Lu and Ng, 2020)	91.9
Our model	**92.4**

Table 3: Coreference resolution performance on the OntoNotes dataset (using gold entity mentions).

3.3 Fine-Grained Entity Typing

We use a fine-grained type classification model based on Lin and Ji (2019) that uses a latent type representation. The model consists of a novel attention mechanism and a hybrid type classifier. This model advances existing methods in two aspects: feature extraction and type prediction. To capture richer contextual information, contextualized word representations are used instead of fixed word embeddings as in previous work. In addition, a two-step mention-aware attention mechanism is proposed to enable the model to focus on important words in mentions and contexts. Also, the model uses a hybrid classification method beyond binary relevance to exploit type inter-dependency with latent type representation. Instead of independently predicting each type, a low-dimensional vector that encodes latent type features is predicted and the type vector is reconstructed from this latent representation. The training data is automatically generated from Wikipedia. Tables 4 and 5 compare the performance of this model against state-of-the-art methods on OntoNotes and FIGER evaluation sets respectively.

To adapt this model to the RUFES task, we start by mapping the YAGO fine-grained types to RUFES ontology. For each entity, we extract the

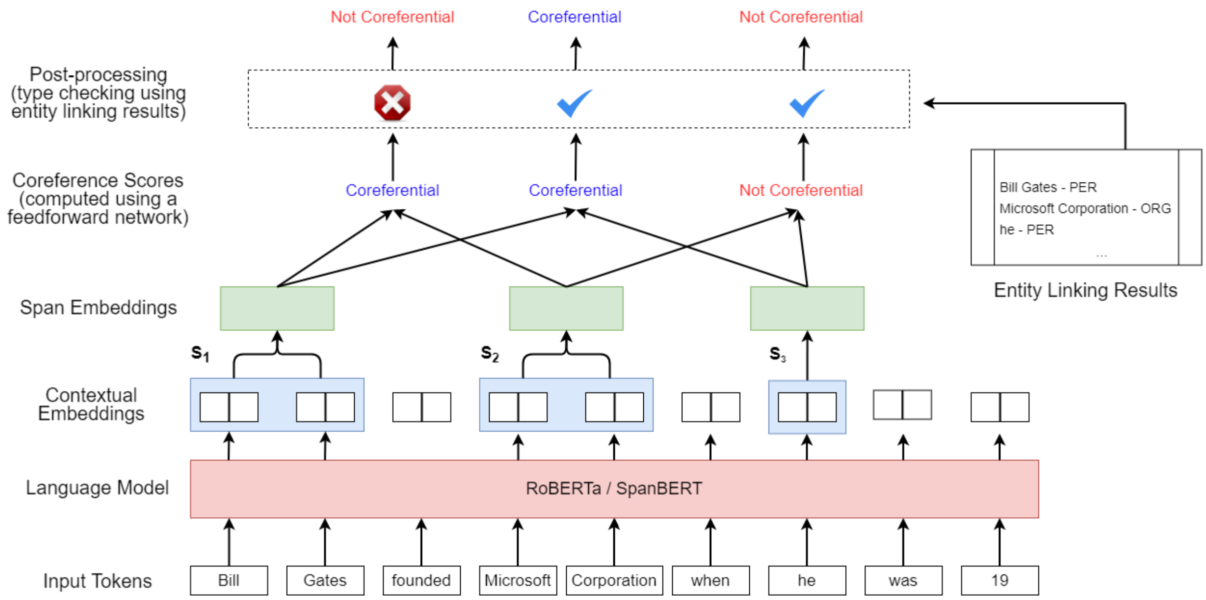

Figure 1: Overall architecture of our entity coreference resolution model.

Model	Acc	Macro F1	Micro F1
(Shimaoka et al., 2016)	51.7	70.9	64.9
(Ren et al., 2016)	57.2	71.5	66.1
(Choi et al., 2018)	59.5	76.8	71.8
Lin and Ji (2019)	**63.8**	**82.9**	**77.3**

Table 4: Results on the OntoNotes test set.

Model	Acc	Macro F1	Micro F1
(Ling and Weld, 2012)	53.2	69.9	69.3
(Yogatama et al., 2015)	-	-	72.3
(Shimaoka et al., 2017)	54.5	74.8	71.6
+ hand-crafted	59.7	79.0	75.4
Lin and Ji (2019)	**62.9**	**83.0**	**79.8**

Table 5: Results on the FIGER (Gold) test set.

sentences mentioning it, and use the YAGO fine-grained types of this entity as labels to construct noisy training data.

Furthermore, we obtain the YAGO fine-grained types by linking entities to the Freebase (LDC2015E42), and map them to RUFES entity types. Besides, for Geo-political and location entities, we link them to GeoNames [3] and determine their fine-grained types using GeoNames attributes *feature_class* and *feature_code*. Considering that most nominal mentions can not be linked to Freebase or GeoNames and the lack of training data, we develop a nominal keyword list to identify nominal mentions for each type. We then compute a weighted score for these typing results and normalize the scores as typing confidence values. We also exploit the example keywords provided in the RUFES ontology for each fine-grained type.

4 Evaluation Results

4.1 Results

Table 6 shows our results on the development set and from the preliminary evaluation. The development set consists of 50 fully annotated sample documents which were taken from the Washington Post News corpus.

For the evaluation round, we submitted our system's output on 100,000 articles from Washington Post. We present our numbers from preliminary evaluation, which was done on a 106 article subset of our submission.

Evaluation	Precision	Recall	F1
Development set	44.0	36.0	37.8
Prelim. evaluation	44.8	36.3	38.2

Table 6: Performance of our system on the evaluation sets.

The scorer[4] is a new entity-level type metric to evaluate fine-grained entity typing. The system entity IDs and gold entity IDs for the mentions are aligned to compute the type precision, type recall, and type F1 on the types for each pair of aligned system and gold entities, whereas the unaligned system entities and gold entities each have F1=0. The final score is a macro-averaged type F1, which

[3] http://geonames.org/

[4] https://github.com/shahraj81/rufes

is the mean F1 over the unaligned entities and the pairs of aligned entities.

4.2 Analysis

We analyze our system output for 10 articles that were randomly selected from the evaluation corpus and pick the first 40 errors in each article. We categorize the system errors as follows: **Wrong type**: the system and gold mention-level entity types don't match; **Missing mention:** the gold mention span does not match any system mention span; **Extraneous mention:** the system mention does not match or overlap with any gold mention; **Wrong extent:** the system mention span and gold mention span overlap but have different extents; **Wrong coreference:** errors relating to document-level coreference.

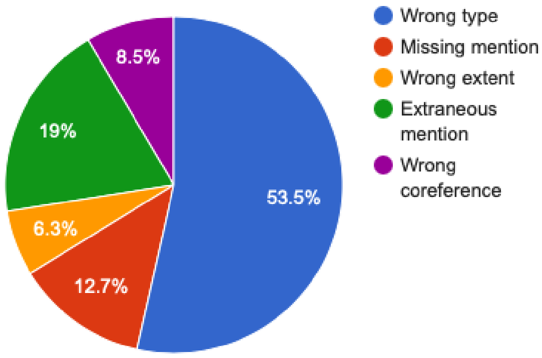

Figure 2: Distribution of different error categories in the system output.

Figure 2 shows the distribution of these error categories in the system output. We see that wrong entity-type is the most common error, with a suprisingly high amount of errors (19%) coming from extraneous mentions.

5 Conclusion

We have described our submission to the RUFES track of TAC 2020. Our system combines three state-of-the-art components in mention extraction, entity coreference resolution and fine-grained entity typing. We also adapt the entity typing model to the task, by training with noisy silver-standard data constructed from Wikipedia using YAGO types in the RUFES ontology.

Acknowledgement

This research is based upon work supported in part by U.S. DARPA KAIROS Program No. FA8750-19-2-1004, U.S. DARPA AIDA Program No. FA8750-18-2-0014, and Air Force No. FA8650-17-C-7715. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of DARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

Eunsol Choi, Omer Levy, Yejin Choi, and Luke Zettlemoyer. 2018. Ultra-fine entity typing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 87–96.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. 2020. Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77.

Mandar Joshi, Omer Levy, Luke Zettlemoyer, and Daniel S Weld. 2019. Bert for coreference resolution: Baselines and analysis. 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 5807–5812.

Kenton Lee, Luheng He, and Luke Zettlemoyer. 2018. Higher-order coreference resolution with coarse-to-fine inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 687–692, New Orleans, Louisiana. Association for Computational Linguistics.

Manling Li, Alireza Zareian, Ying Lin, Xiaoman Pan, Spencer Whitehead, Brian Chen, Bo Wu, Heng Ji, Shih-Fu Chang, Clare Voss, Daniel Napierski, and Marjorie Freedman. 2020. GAIA: A fine-grained multimedia knowledge extraction system. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 77–86, Online. Association for Computational Linguistics.

Ying Lin and Heng Ji. 2019. An attentive fine-grained entity typing model with latent type representation.

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 6198–6203.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7999–8009.

Xiao Ling and Daniel Weld. 2012. Fine-grained entity recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26.

Jing Lu and Vincent Ng. 2020. Conundrums in entity coreference resolution: Making sense of the state of the art. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6620–6631, Online. Association for Computational Linguistics.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proc. the 55th Annual Meeting of the Association for Computational Linguistics*.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. Conll-2012 shared task: Modeling multilingual unrestricted coreference in ontonotes. In *Joint Conference on EMNLP and CoNLL-Shared Task*, pages 1–40.

Xiang Ren, Wenqi He, Meng Qu, Clare R Voss, Heng Ji, and Jiawei Han. 2016. Label noise reduction in entity typing by heterogeneous partial-label embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1825–1834.

Sonse Shimaoka, Pontus Stenetorp, Kentaro Inui, and Sebastian Riedel. 2016. An attentive neural architecture for fine-grained entity type classification. In *Proceedings of the 5th Workshop on Automated Knowledge Base Construction*, pages 69–74.

Sonse Shimaoka, Pontus Stenetorp, Kentaro Inui, and Sebastian Riedel. 2017. Neural architectures for fine-grained entity type classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1271–1280.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. 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 5788–5793.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. Ace 2005 multilingual training corpus. *Linguistic Data Consortium*.

Liyan Xu and Jinho D. Choi. 2020. Revealing the myth of higher-order inference in coreference resolution. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8527–8533, Online. Association for Computational Linguistics.

Dani Yogatama, Dan Gillick, and Nevena Lazic. 2015. Embedding methods for fine grained entity type classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 291–296.