Towards Personalized Descriptions of Scientific Concepts

Sonia K. Murthy, Daniel King, Tom Hope, Daniel Weld, Doug Downey
Allen Institute for Artificial Intelligence
{soniam,daniel,tomh,danw,dougd}@allenai.org

Abstract

A single scientific concept can be described in many different ways, and the most informative description depends on the audience. In this paper, we propose generating personalized scientific concept descriptions that are tailored to the user's expertise and context. We outline a complete architecture for the task and release an expert-annotated resource, ACCoRD, which includes 2,360 labeled extractions and 1,309 hand-authored concept descriptions for the key first step of extracting and generating multiple distinct descriptions of a concept in terms of different reference concepts. Our results show that existing models are not suitable for our task and that our extractive model substantially outperforms these baselines.

1 Introduction

Theories of learning suggest that an effective way to describe a new concept to someone is to ground its description within the network of concepts they are already familiar with (NRC, 2000). For example, while many would appreciate a description of MultiRC (Khashabi et al., 2018) stated in terms of a widely-accessible reference concept (e.g., "dataset"), a reader familiar with other machine comprehension benchmarks (e.g., SQuAD (Rajpurkar et al., 2016)) might find a comparison that highlights the differences between the benchmarks to be more helpful (see Fig. 1).

This paper investigates the new challenge of producing multiple descriptions for a single concept (hereafter "target") in terms of distinct reference concepts, and identifying the description that is most helpful for a given user. Given text from scientific papers, our system **extracts** the sentences that describe one scientific concept in terms of another and **generates** succinct, self-contained descriptions of the concepts' relationship. For **personalization**, we propose a description-ranking scheme that incorporates an estimate of user expertise.

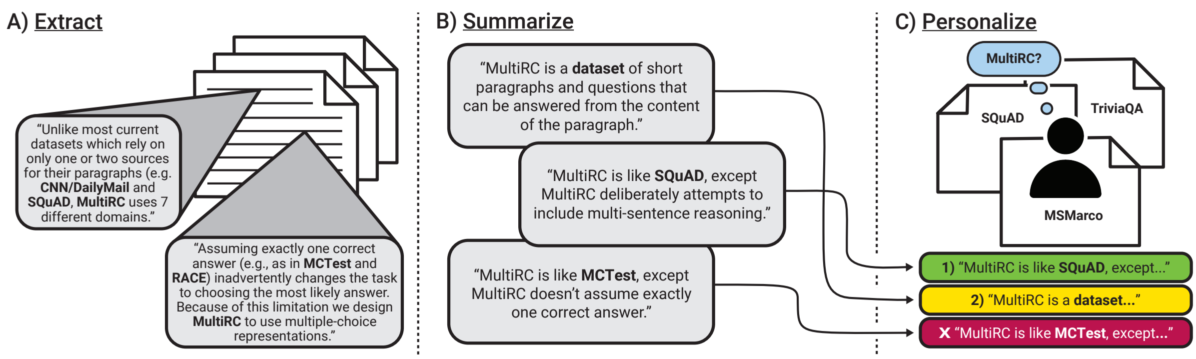

Figure 1: Overview of our proposed system: (A) extract sentences describing one concept in terms of another, (B) summarize the extracted sentences into succinct, self-contained concept descriptions, and (C) rank the most suitable descriptions of a target concept for the user.

Model	HEDDEx	HEDDEx$_{is\text{-}a}$	DyGIE++	SciBERT	Positive
Train set	W00	W00	SciERC	ACCoRD	baseline
F1	0.329	0.449	0.532	**0.624**	0.484

Table 1: Results for our extractive model and relevant baselines on the ACCoRD test set ($n = 674$). Our model trained on ACCoRD outperforms models that target related tasks, even when they beat a baseline that always assigns positive labels, suggesting that our data set addresses an importantly different task.

While previous work has investigated extracting and generating definitions of scientific concepts (Espinosa-Anke and Schockaert, 2018; Vanetik et al., 2020; Veyseh et al., 2020; Kang et al., 2020), they focus on producing a *single* canonical description for each concept, whereas we aim to preserve multiple distinct descriptions. Likewise, extracting concept comparisons has been targeted in the context of relation extraction (Luan et al., 2018; Luu et al., 2021), but this setting does not require that the comparison be *descriptive* of a concept.

2 Data set

We release Automatic Comparison of Concepts with Relational Descriptions (ACCoRD)[1], a high-quality (Cohen's $\kappa = 0.658$) data set from the computer science (CS) domain that includes 2,360 labeled extractions and 1,309 hand-authored concept descriptions. The concept descriptions in our data set describe a target scientific concept in terms of another concept using one of four relation types: *is-a*, *compare*, *part-of*, and *used-for*. We include both 1-sentence and 2-sentence source text settings to allow for experiments on the effects of richer context on our task.

To create this data set, we considered the abstract, introduction, and related works sections of 741 CS papers from S2ORC (Lo et al., 2020), a large corpus of academic papers. We automatically identified candidate sentences with at least one significant CS concept by performing simple string matching against a set of high-precision concepts extracted from CS papers (King et al., 2019). These sentences were labeled as positive if they described a target scientific concept in terms of any other concept. Each positive extraction could yield multiple concept descriptions if a target concept was described in terms of multiple other concepts in the source text, or if the extraction contained multiple target concepts.

3 Experiments and results

We build a SciBERT-based (Beltagy et al., 2019) extractor to identify sentences that describe a target concept in terms of another concept. The model is trained on ACCoRD extractions, with concepts demarcated as in Wu et al. (2019). We select optimal hyperparameters using cross-validation on the training set. The output of our extractive model is used as input to our BART$_{\text{LARGE}}$ (Lewis et al., 2020) model fine-tuned on the hand-authored concept descriptions in ACCoRD to generate the final target concept description. This model maps each extraction to a single concept description; multi-sentence summarization is only done in the 2-sentence source text setting.

Extraction We compare our model's results on ACCoRD to those of existing state-of-the-art scientific definition and relation extraction systems. For our definition extraction baseline, we test HEDDEx (Kang et al., 2020) trained on W00 (Yiping et al., 2013), a similarly sized corpus of definition sentences from workshop papers from the 2000 ACL Conference. Since HEDDEx was originally only intended to produce a single canonical definition of a scientific term within a paper, we also evaluate its performance on the subset of ACCoRD that was marked as containing an "is-a" relationship between the reference and target concept, to more faithfully evaluate its potential. For our relation extraction baseline, we test DyGIE++ (Wadden et al., 2019) trained on SciERC (Luan et al., 2018), a scientific relation extraction data set. Table 1 shows these results for the union of the 1- and 2-sentence source text settings in ACCoRD, as our qualitative conclusions remained unchanged across these settings.

[1]Dataset available at `https://github.com/allenai/ACCoRD`

Error type	Example source text	Example BART summary	% of samples
Factuality	in this work, we propose a novel end-to-end structureaware convolutional network (sacn) that takes the benefit of <ENT> gcn <ENT> and conve together.	<ENT> gcn <ENT> is a novel end-to-end structureaware convolutional network (sacn).	35
Reference concept is too general	...a more recent generation including the microsoft academic search 5 and <ENT> arnetminer <ENT> [12] systems has highlighted the importance trend discovery and prediction, and proposed novel features for those purposes.	<ENT> arnetminer <ENT> is a system that has highlighted the importance trend discovery and prediction, and proposed novel features for those purposes.	20
Fluency	the proposed protocol inherits the group motion support of landmark routing (<ENT> lanmar <ENT>). as in lanmar, a landmark node is elected within each group; the election procedure is dynamic and is described in detail in [8, 20] .	<ENT> lanmar <ENT> is a proposed protocol that inherits the group motion support of landmark routing, and as in lanmar, a landmark node is elected within each group; the election procedure...	15

Table 2: Examples of different error types from a small sample ($n = 20$) of our initial summarization results produced by our BART model fine-tuned on ACCoRD.

Generation Preliminary evaluations on a small sample ($n = 20$) of our initial generative results show that our model learned to identify the target concept and description format with high accuracy (100% and 85%, respectively), but that these descriptions were factual in only 65% of samples, and fluent in 85%. We also find that 20% of generated descriptions in our sample were in terms of a reference concept like “tool”, “system”, or “method” that is too general to support useful personalization. Table 2 shows examples of source texts and their predictions for each of these error types. Finally, preliminary explorations into the personalization phase showed that rankings of a random sample of generated descriptions produced by 2 users with similar domain knowledge were highly correlated.

4 Conclusions and future work

We presented personalized scientific concept description, the task of describing a target scientific concept in terms of a reference concept most helpful for a given user. Our preliminary results show that the computer science literature often contains multiple distinct descriptions for the same concept, and that our extractive methods can produce more accurate results than relevant baselines.

Exploring personalization is a key item of future work. In particular, we plan to extract the scientific concepts from the papers authored, cited, and read by the user to estimate which concepts the users knows, and then present the generated description whose reference concept is most appropriate and informative for the user’s background knowledge. Our approach requires that a diversity of descriptions for the same concept be available in the corpus. Fortunately, extrapolation from our labeled data and the term frequencies in S2ORC suggests this is the case. Specifically, of the 119k ForeCite concepts strings we consider, we estimate that roughly half of them are described in terms of five or more distinct reference concepts in S2ORC.

Additional future work will be directed at improving the factuality, fluency, and informativeness of our generative methods; implementing and evaluating personalization schemes on top of the methods evaluated in this work; and measuring the benefit of the complete system for users. While our current methods only generate descriptions for computer science concepts found in ForeCite, extending our methods to include concepts from other domains remains an open direction.

Acknowledgements

This project was supported in part by NSF Grant OIA-2033558.

References

Amir Pouran Ben Veyseh, Franck Dernoncourt, Dejing Dou, and Thien Huu Nguyen. 2020. *A joint model for definition extraction with syntactic connection and semantic consistency*. In Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20).

Damien Sileo, T. V. D. Cruys, Camille Pradel, and Philippe Muller. 2019. *Discourse-Based Evaluation of Language Understanding*.

Daniel Khashabi, Snigdha Chaturvedi, Michael Roth, Shyam Upadhyay, and Dan Roth. 2018. *Looking Beyond the Surface: A Challenge Set for Reading Comprehension over Multiple Sentences*. In Proceedings of NAACL-HLT, pages 252–262

Daniel King, Doug Downey, and Daniel Weld. 2020. *High-Precision Extraction of Emerging Concepts from Scientific Literature*. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2020. *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, pages 5784–5789.

Dongyeop Kang, Andrew Head, Risham Sidhu, Kyle Lo, Daniel Weld, and Marti A. Hearst. 2020. *Document-Level Definition Detection in Scholarly Documents: Existing Models, Error Analyses, and Future Directions*. In Proceedings of the First Workshop on Scholarly Document Processing, pages 196–206.

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. *SciBERT: Pretrained Language Model for Scientific Text*. 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 3615–3620.

Kelvin Luu, Xinyi Wu, Rik Koncel-Kedziorski, Kyle Lo, Isabel Cachola, and Noah A. Smith. 2021. *Explaining Relationships Between Scientific Documents*. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pages 2130–2144.

Kyle Lo, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Daniel Weld. 2020. *S2ORC: The semantic scholar open research corpus*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 4969–4983.

Luis Espinosa-Anke and Steven Schockaert. 2018. *Syntactically aware neural architectures for definition extraction*. 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 378–385, New Orleans, Louisiana. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2020. *BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 7871–7880.

Natalia Vanetik, Marina Litvak, Sergey Shevchuk, and Lior Reznik. 2020. *Automated discovery of mathematical definitions in text*. In Proceedings of The 12th Language Resources and Evaluation Conference, pages 2086–2094, Marseille, France. European Language Resources Association.

National Research Council. 2000. *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. The National Academies Press, Washington, DC. https://doi.org/10.17226/9853.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. *SQuAD: 100,000+ Questions for Machine Comprehension of Text*. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 2383-2392.

Shanchan Wu and Yifan He. 2019. *Enriching Pre-trained Language Model with Entity Information for Relation Classification*. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pages 2361–2364.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. *Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction*. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3219–3232.

Yiping Jin, Min-Yen Kan, Jun-Ping Ng, and Xiangnan He. 2013. *Mining Scientific Terms and their Definitions: A Study of the ACL Anthology*. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 780–790.