Responses must be**no more than 900 words in total**

We thank Reviewer #1 for the detailed comments. We appreciate the reviewer’s positive feedback on xxx

We thank Reviewer #1’s encouraging remarks on our work.

We thank Reviewer #2 for the encouraging remarks on our work, especially on the novelty of our approach

We thank Reviewer #3 for the constructive and detailed comments. We are glad to know that the reviewer thinks our approach is effective and the paper is well-written. The reviewer has also raised interesting questions about the design and analysis of our model.

entity typing is a long standing task for IE, and

[Reviewer #1]

The goal of the new task doesn't seem to be quite clear, compared to previous similar tasks (which either has different label set for different context, or aims to complete a KB instead). The new task, fine-grained chemical entity typing, doesn't seem to have a clear goal. The usefulness of entity typing appears when either (1) the mention types depends on the context, or (2) it's used to complete a KB. This paper doesn't seem to mention anything about completing a KB, so it seems to be directed towards the first case. In CHEMET, however, I noticed that each mention has the exact same set of labels assigned to them (based on the training data given). So the mention types don't depend on the context, which makes the task a bit moot, since it's essentially classifying existing entries in the PubChem based on the novel ontology derived from Wikipedia. But it is not presented as such, and so this raises doubt whether this is the intention of the authors.

We thank Reviewer #1 for the detailed and constructive comments.

To the best of our knowledge, the output of the entity typing task does not only presents necessary information for a full-fledged relational graph (or knowledge graph), but also used to improve performance on downstream models (relation extraction, event extraction, KB completion as you mentioned, etc.)

Given the importance of the function, the severe lack of IE models in Chemistry domain (as we mentioned in section 1), and the difficulty of understanding chemistry text, we contribute this work. In short, the goal is to facilitate research in chemistry IE, and therefore we propose a carefully annotated dataset, and meanwhile, based on the characteristics (multimodal information) of chemistry, we designed an effective model to tackle issues in chemistry text.

[mention types don't depend on the context] Reviewer raises a very nice point. A characteristics in chemistry text is that there is no ambiguity in chemical names (i.e., mention name only belongs to a single entities), and this is different from news domain (e.g., “Jack, the bar located behind house”, and “Jack plays in League 10 as a professional player” ). In fact, however, context of sentence is a highly needed feature in chemistry text because of the unnatural forms of most chemical names (discussed in section 1&3), which are difficult to be understand by current contextual word embeddings.

[same as classifying existing entries in the PubChem] Moreover, the classification task has little to do with PubChem entries, as our mention are from research literature and they may or may not be present in PubChem chemical records; it just happened that we use research papers from PubChem database

[linkability] The fact that the model shows clear improvement even when so many mention are unlinkable proves the usefulness of our model in chemistry FET. This shows the potential of our model when an entity linking module can be incorporated, which also corresponds to our goal of facilitating this research area. In fact, as mentioned in section 6, our next step is to develop entity linking algorithm into our model.

[Reviewer #2]

[Reviewer #3]

**[ === References ==== ]**

References.

[1] Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior knowledge and the comprehension of compatible and incompatible text. Reading Research Quarterly.

[2] Braasch, J. L. G., & Goldman, S. R. (2010). The role of prior knowledge in learning from analogies in science texts. Discourse Processes.

[General Response]

We thank the reviewers for taking the time to review our paper and for their insightful comments. We are happy that all three reviewers agree on the effectiveness of our approach. As we have elaborated in each individual response, if given more space, we can fully address all the reviewers’ comments.

We thank the reviewers for their insightful comments. We are happy that all three reviewers agree on the novelty and effectiveness of our proposed approach. As we have elaborated in each individual response, if more space is given, we will be able to address all the reviewers’ comments and improve our current draft. Due to the space constraints for short papers, we tried to include most of the key analysis and experiments in the initial paper submission.

reviewer 1

\* The new task, fine-grained chemical entity typing, doesn't seem to have a clear goal. The usefulness of entity typing appears when either (1) the mention types depends on the context, or (2) it's used to complete a KB. This paper doesn't seem to mention anything about completing a KB, so it seems to be directed towards the first case. In CHEMET, however, I noticed that each mention has the exact same set of labels assigned to them (based on the training data given). So the mention types don't depend on the context, which makes the task a bit moot, since it's essentially classifying existing entries in the PubChem based on the novel ontology derived from Wikipedia. But it is not presented as such, and so this raises doubt whether this is the intention of the authors.

\* The proposed contribution on using the graph structure of the chemical compound is also quite limited in the sense that it can only be used if the model can link the mention to an existing entry in the database. So this component cannot be used in other tasks such as entity recognition, which is arguably also quite important.

mention types

The

reviewer 2

What is this paper about, what contributions does it make, and what are the main strengths and weaknesses?

---------------------------------------------------------------------------

2898 Fine-Grained Chemical Entity Typing with Multimodal Knowledge

Representation

This paper makes two contributions. The first contribution is a new benchmark dataset called CHEMET. First, a chemical

ontology was constructed based on the graph structure of Wikipedia categories, resulting in 30 fine-grained

types. Labels for each entity were generated systematically with the following procedure. At a certain node,

perform a depth-limited depth-first search starting from that node, and use a spell-checker dictionary of

chemistry terms to determine relevant subcategories. The titles of all subcategories encountered in this DFS

are considered entities of the starting node. This process is repeated for all nodes. Human annotators are

then used to annotate types as they occur in the PubChem documents, resulting in a datasetof100 documents

containing 1269 sentences. The second contribution of this paper is a novel FET model architecture. The

first component of the architecture uses SciBERT to obtain a contextual embedding for a chemical name

in the document, as well as an embedding that only represents the surrounding contextÍ¾ these are concate-

nated into one embedding representing local information. The second component is a multimodal encoder

that uses the PubChem search API to obtain a text description and the graph molecular structure of a given

chemical. It uses SciBERT on the description to obtain another contextual embedding and uses Graph Iso-

morphism Network to obtain node embeddings of the molecular structure. A Transformer is used to learn

dependencies between the description and node representations to obtain cross-modal features. These fea-

tures are concatenated with the unimodal node embeddings and the entire description embedding to obtain

the complete multimodal representation. Both the local text and multimodal representations are used to train

a feed-forward neural network that predicts the entity type. This model is evaluated on a multi-label classi-

fication using entity types as labels and micro-F1 and accuracy as metrics. The baselines used are SciBERT,

SciBERT variants that also use individual components from the proposed model, and Latent Type Repre-

sentation. The proposed model shows better performance compared to all baselines. Analysis of the results

shows the majority of errors are due to limited scope of the PubChem database, extraneous general labels in

the dataset, and misunderstandings of the chemical name.

This paper considers the valid and important task in information extraction, as well as NLP in general.

New benchmark datasets especially are extremely valuable for the field and can be used in the evaluation of

future models, and the domain of chemistry is common enough for this dataset to be useful for models that

are domain-specific to chemistry as well as a good challenge for general entity-typing models. Overall, the

methodology for constructing the dataset is good and well-explained, and it appears that adequate measures

were taken to ensure the cleanliness of the final data. Regarding the proposed model,the motivation and ideas are clearly explained. Regarding the

results discussion, the error analysis was insightful about the main difficulties in the task, considering the

relatively low scores on a difficult task.

I had several concerns about the paper. First, the dataset, while valuable, is quite small. It is not clear how useful it will be or if there are plans to expand it. Second, the use of wikipedia categories instead of scientific ontologies such as ChEBI was puzzling to me. It was not clear why Wikipedia would be valuable for typing chemical entities. A third issue is the lack of citations to other work on chemical information extraction and typing, such as

ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature

Matthew C. Swain and Jacqueline M. Cole

ChemEx: information extraction system for chemical data curation

Atima Tharatipyakul, Somrak Numnark, Duangdao Wichadakul, Supawadee Ingsriswang

Krallinger, M., Rabal, O., Leitner, F. et al. The CHEMDNER corpus of chemicals and drugs and its annotation principles. J Cheminform 7, S2 (2015).

NLM-Chem, a new resource for chemical entity recognition in PubMed full text literature. Rezarta Islamaj et al Sci Data 8, 91 (2021)

It was not clear to me what the exact differences were between this dataset/approach and other work, and why the differences were important.

Related to the above, I found the baselines used in the experiments too limited.

The majority of the baselines

are simply variants of the proposed model, and there is really only one alternative baseline model that is

used for results comparison. For comparison, the paper that proposes that alternative baseline (Latent Type

Representation) evaluates 10 other baselines in addition to variants of the proposed model. In addition,

the evaluation of model variants are also considered part of the ablation study, which gives more reason to

include other baseline models. It would be good to include more baselines if possible to show a more accurate

picture of this modelâ€™s performance compared to state-of-the-art models, especially considering that this

paper also introduces a new benchmark dataset. In addition, the experiment results do not show much

improvement in performance, as the difference in performance between the variants of the proposed model is

quite small. Based on the ablation study results, it appears that most components appear to only marginally

improve performance, with most variants performing worse than Latent Type Representation. The only

component of the model that appears to significantly improve performance is the contextual embedding

from the chemical description.

I also did not understand how the hierarchical prediction problem was solved exactly by the proposed approach---were the labels "flattened" and then predicted?

The paper also has issues with numerous language issues and typos, such as repeated references to "multimedia" rather than multimodal. It needs to be carefully proofread.

The paper should cite the Suzuki reaction paper rather than Wikipedia.

To summarize, the paper is tacking an important problem, but has a number of issues that, if resolved, could improve the paper.

first of all thanks so much for pointing our deficiency of our paper positioning