An Ensemble-of-Experts Framework for Rehearsal-free Continual Relation Extraction

📄 **(/pdf?id=T_RLJdl4Ni)**

Anonymous

17 Feb 2024   ACL ARR 2024 February Blind Submission   Readers: February, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers, Paper2376 Authors   Show Revisions (/revisions?id=T_RLJdl4Ni)

Abstract: Continual relation extraction (CRE) aims to continuously learn relations in new tasks without forgetting old relations in previous tasks.

Current CRE methods are all rehearsal-based which need to store samples and thus may encounter privacy and security issues. This paper targets rehearsal-free continual relation extraction for the first time and decomposes it into task identification and within-task prediction sub-problems.

Existing rehearsal-free methods focus on training a model (expert) for within-task prediction yet neglect to enhance models' capability of task identification.

In this paper, we propose an Ensemble-of-Experts (EoE) framework for rehearsal-free continual relation extraction. Specifically, we first discriminatively train each expert by augmenting analogous relations across tasks to enhance the expert's task identification ability. We then propose a cascade voting mechanism to form an ensemble of experts for effectively aggregating their abilities.

Extensive experiments demonstrate that our method outperforms current rehearsal-free methods and is even better than rehearsal-based CRE methods.

Paper Type: long

Research Area: Information Extraction

Contribution Types: Model analysis & interpretability

Languages Studied: English

Revealed to Shen Zhou, Yongqi Li, Xin Miao, Tieyun Qian

16 Feb 2024 (modified: 16 Feb 2024)   ACL ARR 2024 February Submission

Authors: *Shen Zhou (/profile?id=~Shen_Zhou2), Yongqi Li (/profile?id=~Yongqi_Li3), Xin Miao (/profile?id=~Xin_Miao4), Tieyun Qian (/profile?id=~Tieyun_Qian1)*

Reassignment Request Action Editor: This is not a resubmission

Reassignment Request Reviewers: This is not a resubmission

Preprint: no

Preprint Status: There is no non-anonymous preprint and we do not intend to release one.

Preferred Venue: ACL 2024

Consent To Share Data: yes

Consent To Review: yes

Consent To Share Submission Details: On behalf of all authors, we agree to the terms above to share our submission details.

A1: yes

A1 Elaboration For Yes Or No: Section Limitations

A2: yes

A2 Elaboration For Yes Or No: Section Ethics Statement

A3: yes

A3 Elaboration For Yes Or No: Abstract; Section Introduction

B: yes

B1: yes

B1 Elaboration For Yes Or No: Section 5.1

B2: n/a

B3: n/a

B4: n/a

B5: n/a

B6: yes

B6 Elaboration For Yes Or No: Appendix C.1

C: yes

C1: yes

C1 Elaboration For Yes Or No: Appendix C.3

C2: yes

C2 Elaboration For Yes Or No: Section 5.1

C3: yes

C3 Elaboration For Yes Or No: Appendix C.3

C4: n/a

D: no

D1: n/a

D2: n/a

D3: n/a

D4: n/a

D5: n/a

E: no

E1: n/a

Reply Type: [all]   Author: [everybody]   Visible To: [all readers]   **14 Replies**

Hidden From: [nobody]

[−] **Meta Review of Paper2376 by Area Chair u4V6**

ACL ARR 2024 February Paper2376 Area Chair u4V6

08 Apr 2024, 01:00   ACL ARR 2024 February Paper2376 Meta Review   Readers:
Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Authors, Paper2376
Reviewers Submitted, Program Chairs   Show Revisions (/revisions?id=G-6y2Opg-5S)

Paper Summary:

1. This paper proposes a novel Ensemble-of-Experts framework for Rehearsal-Free Continual Relation Extraction. It decomposes the problem into task identification and within-task prediction, emphasizing discriminative training and a cascade voting mechanism.

2. EoE aggregates knowledge from different continual learning stages for both task identification and within-task prediction.

3. Experimental results show that the outperforms the existing methods in both rehearsal-free and rehearsal-required settings.

Summary Of Strengths:

1. Introduction of rehearsal-free Continual Relation Extraction (RFCRE) problem and the Ensemble-of-Experts (EoE) framework, which is novel.

2. The proposed Ensemble-of-Experts framework is deemed reasonable. EoE framework innovates in task-specific model management and integrates expertise through a cascade voting mechanism.

3. Experimental results demonstrate that EoE outperforms baseline methods.

Summary Of Weaknesses:

1. The paper should elaborate on the mechanisms behind Undetermined relation in Analogous Relation Augmentation.

2. The explanation of the expert voting mechanism is not clear.

3. The performance of EoE heavily relies on accurate task identification.

4. More ablation experiments is desired to provide more clear results.

Overall Assessment: 4 = There are minor points that may be revised

Best Paper Ae: No

Ethical Concerns:

N/A

Needs Ethics Review: No

Information Regarding The New ACL Policy On Deanonymized Preprints: I confirm I have read the information above about changes to the anonymity policy.

[−] **Thanks for recognition of our proposed new problem and the corresponding framework!**

ACL ARR 2024 February Paper2376 Authors Shen Zhou (/profile?id=~Shen_Zhou2) (privately revealed to you)

13 Apr 2024, 14:16   ACL ARR 2024 February Paper2376 Author-Editors Confidential Comment   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Authors   Show Revisions (/revisions?id=D7wbQcawcPn)

Comment:

Dear Area Chair,

Many thanks for your recognition of our proposed rehearsal-free Continual Relation Extraction (RFCRE) problem and the corresponding Ensemble-of-Experts (EoE) framework! We also wish our study can inspire more future work towards the RFCRE problem.

Regarding the detailed illustration and more ablation experiments, we have finished them during the rebuttal phase, and surely we will include them in the next version.

Thanks again for your time, expertise, and thoughtful engagement with our work.

Best regards,

Paper 2376 Authors

[−] **Official Review of Paper2376 by Reviewer aVU4**

ACL ARR 2024 February Paper2376 Reviewer aVU4

20 Mar 2024, 13:08 (modified: 02 Apr 2024, 12:39)   ACL ARR 2024 February Paper2376 Official Review   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=QiY0PGAhLC)

Recommended Process Of Reviewing: I have read the instructions above

Paper Summary:

This paper presents a novel Ensemble-of-Experts (EoE) framework for rehearsal-free Continual Relation Extraction (RFCRE). The framework is designed to facilitate the continuous learning of new relations from tasks while retaining previously acquired knowledge, all without the necessity of storing samples from past tasks.

EoE focuses on decomposing the problem into task identification and within-task prediction .The authors advocate a discriminative training strategy to improve each expert's task identification capability through the creation of augmented relations. Additionally, they introduce a cascade voting mechanism to efficiently combine the expertise of multiple experts.

Summary Of Strengths:

S1. The pioneering proposal of the rehearsal-free Continual Relation Extraction (RFCRE) problem S2. The Ensemble-of-Experts (EoE) framework introduces a discriminative training process to enhance experts' task identification capabilities by augmenting analogous relations across tasks. This approach is complemented by a cascade voting mechanism to effectively aggregate expert knowledge. S3. The experiments show that the Ensemble-of-Experts (EoE) outperforms various baseline methods on standard datasets.

Summary Of Weaknesses:

W1. The authors should elaborate on the specific mechanisms through why the Undetermined relation in Analogous Relation Augmentation enhances the task identification capabilities of expert models. Furthermore, it is advisable to incorporate more comprehensive ablation experiments for the DT in the experimental section to provide a clearer illustration. W2. The potential outcomes of utilizing solely the final expert during the Inference phase to determine the voting result are of interest for examination.

Comments, Suggestions And Typos:

Add a more detailed ablation experiment 'w/o UndeterminedRelation' under the ablation experiment 'w/o DT' and a more detailed description of the contribution of adding Undetermined Relation to the expert identification task.

Soundness: 3.5

Overall Assessment: 3 = Good: This paper makes a reasonable contribution, and might be of interest for some (broad or narrow) sub-communities, possibly with minor revisions.

Confidence: 3 = Pretty sure, but there's a chance I missed something. Although I have a good feel for this area in general, I did not carefully check the paper's details, e.g., the math or experimental design.

Best Paper: No

Limitations And Societal Impact:
The authors cite the shortcomings of EoE as the unavoidable time and space required to implement multi-expert models.

Ethical Concerns:
None

Needs Ethics Review: No

Reproducibility: 4 = They could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method.

Datasets: 1 = No usable datasets submitted.

Software: 1 = No usable software released.

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources

Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Reviewer Certification: aVU4

[-] ## Kindly reminder to Reviewer aVU4

ACL ARR 2024 February Paper2376 Authors   Shen Zhou (/profile?id=~Shen_Zhou2) (privately revealed to you)

01 Apr 2024, 18:49 (modified: 01 Apr 2024, 18:50)   ACL ARR 2024 February Paper2376 Official Comment   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=NF900j86rY)

Comment:
Dear Reviewer aVU4,

We greatly appreciate your efforts in reviewing our manuscript. We have tried our best to address your concerns. We kindly want to know whether our responses have satisfactorily resolved your concerns and if you are considering raising our score.

Thank you very much for your hard work, and we look forward to hearing from you.

Best regards!

[-] ## Response to insightful comments of Reviewer aVU4 (1 / 2)

ACL ARR 2024 February Paper2376 Authors   Shen Zhou (/profile?id=~Shen_Zhou2) (privately revealed to you)

29 Mar 2024, 20:46   ACL ARR 2024 February Paper2376 Official Comment   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=hZvDwqEHQa)

Comment:
Dear Review aVU4,

Thank you for your valuable comments. We wish we can address your concern with the following response.

Suggestion 1: The authors should elaborate on the specific mechanisms of why the undetermined relation enhances the task identification capabilities of expert models and incorporate more comprehensive ablation experiments for the DT. Add a more detailed ablation experiment 'w/o UndeterminedRelation' under the ablation experiment 'w/o DT.'

Response 1: Thank you for your insightful suggestions. In addition to two standard datasets in the paper, including FewRel and TACRED, we add a Chinese domain dataset, DuIE2.0 [3], to validate the cross-domain generalization performance as Reviewer aw2T's request. We conduct more fine-grained ablation experiments on DT to validate the

effectiveness of Analogous Relation Augmentation on all three datasets. The experimental results are as follows.

Table 1: Fine-grained ablation study on TACRED regarding task identification accuracy (IA). ✓ and ✗ denote w and w/o the component, respectively.

Undetermined Relation	Reverse Relation	T6	T7	T8	T9	T10
✗	✗	86.1	84.6	83.2	82.5	81.6
✗	✓	86.8	85.4	84.5	83.5	82.6
✓	✗	86.4	84.9	83.9	82.9	81.7
✓	✓	87.5	85.9	84.8	84.1	83.1

Table 2: Fine-grained ablation study on FewRel regarding task identification accuracy (IA). ✓ and ✗ denote w and w/o the component, respectively.

Undetermined Relation	Reverse Relation	T6	T7	T8	T9	T10
✗	✗	91.7	90.5	89.5	88.1	86.6
✗	✓	91.9	90.8	89.8	88.6	87.2
✓	✗	91.8	90.7	89.6	88.3	86.9
✓	✓	91.8	90.8	89.7	88.6	87.2

Table 3: Fine-grained ablation study on DuIE2.0 regarding task identification accuracy (IA). ✓ and ✗ denote w and w/o the component, respectively.

Undetermined Relation	Reverse Relation	T6	T7	T8	T9	T10
✗	✗	88.4	87.5	86.9	85.9	85.4
✗	✓	90.0	89.0	88.4	87.7	87.6
✓	✗	89.1	87.7	86.9	86.4	86.0
✓	✓	90.4	89.4	88.8	88.1	87.6

From Table 1, Table 2, and Table 3, we can conclude that:

(1) Adding either reverse relations (RR) or undetermined relations (UR) can significantly enhance the experts' task identification capability.

(2) Adding both RR and UR can further improve the task identification capability of the experts on TACRED and DuIE2.0, but the results on FewRel do not meet our expectations. We believe the main reason lies in the characteristics of the datasets. Commonly used relation extraction samples can be generally divided into the following two categories:

- The samples that do not require the model have a deep understanding of the context, such as typical samples in FewRel. FewRel is collected from Wikipedia, where many samples involve commonsense relations between entities and do not rely too much on the textual context. For example, we can determine the relation between "Obama" and "person" is "P31:instance of" without specific textual context.
- The samples that require the model have a deep understanding of the context, such as typical samples in TACRED and DuIE2.0. For example, for the entity pair "American Bankers Association" and "Bob Davis," it is hard to determine whether the relation between these two entities is "org:founded_by" or "org:top_members/employees." without the specific context. Given the sentence "American Bankers Association was founded by Bob Davis," the relation is "org:founded_by," while given the sentence "Bob Davis is the executive vice president of American Bankers Association," the relation is "org:top_members/employees."

Overall, the UR we proposed mainly targets the second category. It aims to enhance the CRE model's attention to the context in the samples. Although the UR strategy does not bring significant performance improvements for the first category of RE tasks when combined with RR, it can play an important role for the second category of RE tasks, which might be more common in real continual learning scenarios.

We sincerely thank you again for providing us with the chance to have such an in-depth discussion to clarify this issue.

[−] **Response to insightful comments of Reviewer aVU4 (2 / 2)**

29 Mar 2024, 20:53 (modified: 31 Mar 2024, 16:10)   ACL ARR 2024 February Paper2376 Official Comment   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=NCw68ASBcy)

Comment:
Suggestion 2: A more detailed description of the contribution of Undetermined Relation to the expert identification task should be added.

Response 2: Thank you for your instructive suggestion. The core motivation of undetermined relations is to avoid the expert model extracting insufficient relation representations, i.e., ignoring context information related to relation, which is harmful to the performance of task identification. Below is a detailed description of the contribution of Undetermined Relation to the expert identification task.

[1] and [2] have pointed out that existing relation extraction models may learn shadow representations from entity mentions, lacking an in-depth understanding of context and entity information. Hence, we propose undetermined relation augmentation to generate a new relation by removing the textual context, the goal of which is to prevent the expert from only extracting shadow representations from entity mentions while ignoring the relation-describing textual context.

Overall, if we decompose the problem into within-task prediction and task identification and involve the experts in task identification, the experts will naturally suffer from the inter-task analogous relations. The undetermined relation augmentation and reverse relation augmentation are both proposed to alleviate the effect of inter-task analogous relations. We believe this will inspire more future work in this direction.

Suggestion 3: The potential outcomes of utilizing solely the final expert during the Inference phase to determine the voting result are of interest for examination.

Response 3: Thanks for your comments. However, there might be some misunderstanding about the expert's voting mechanism. Please note that an expert can only vote for a task in which it has stored task-specific knowledge based on task samples. Our work aims to protect potential privacy and security issues in CRE and thus proposes rehearsal-free CRE, which does not save any training samples after its training. Therefore, the last expert model cannot get task-specific knowledge for previous tasks, i.e., we cannot only use solely the expert from the last stage for task identification. The first expert can vote for all tasks, the second expert can vote for tasks 2 to n, ..., and the final expert can only vote for the final task. This imbalanced voting issue is the motivation for proposing our cascade voting mechanism.

Suggestion 4: The authors cite the shortcomings of EoE as the unavoidable time and space required to implement multi-expert models.

Response 4: Despite the extra time cost in our solution to our proposed RFCRE problem, our main contribution lies more in the problem itself, which is the first one to target the potential privacy issues in continual relation extraction. We believe our proposed problem will inspire more research along this line, as well as the potentially more efficient solutions to this problem.

References

[1] Learning from Context or Names? An Empirical Study on Neural Relation Extraction. EMNLP 2020

[2] Should We Rely on Entity Mentions for Relation Extraction? Debiasing Relation Extraction with Counterfactual Analysis. NAACL 2022

[3] DuIE: A Large-Scale Chinese Dataset for Information Extraction. NLPCC 2019

Thanks again for the constructive suggestions. We will continue to improve our paper according to your suggestions.

[−] **Comments after Author Discussion**

ACL ARR 2024 February Paper2376 Reviewer aVU4

02 Apr 2024, 12:42   ACL ARR 2024 February Paper2376 Official Comment   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=fi_MDEt41R)

Comment:
The authors effectively addressed my inquiries regarding fine-grained ablation experiments and substantiated the design's validity through the presentation of experimental results.

Regarding the expert voting mechanism query, the author's response adequately addressed my question. However, for enhanced reader comprehension, a more detailed explanation of the expert voting mechanism within the main body of the text would be beneficial.

Overall, I will raise my score.

[−] ## Response to Reviewer aVU4

ACL ARR 2024 February Paper2376 Authors   Shen Zhou (/profile?id=~Shen_Zhou2) (privately revealed to you)
02 Apr 2024, 13:40   ACL ARR 2024 February Paper2376 Official Comment   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=_yCwP-5NR6)
Comment:
Dear Reviewer aVU4,

Thank you so much for taking the time to read our response and raise the score. We are pleased to know that we have addressed your concerns. Such a discussion is helpful to improve our work. Thanks again.

Best regards!

Paper 2376 Authors.

[−] ## Official Review of Paper2376 by Reviewer zZ9N

ACL ARR 2024 February Paper2376 Reviewer zZ9N
17 Mar 2024, 13:14 (modified: 17 Mar 2024, 13:24)   ACL ARR 2024 February Paper2376 Official Review   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=UUtqwoyElxF)

Recommended Process Of Reviewing: I have read the instructions above

Paper Summary:
This paper proposes a rehearsal-free continual relation extraction problem. To solve the problem, the authors develop an ensemble-of-experts framework that aggregates experts knowledge from different continual learning stages for task identification and within-task prediction. The results from experiments show superior performance over rehearsal-free and rehearsal-based continual relation extraction methods.

Summary Of Strengths:
1. This paper proposed a new problem in continual relation extraction, i.e., under the rehearsal-free setting. In comparison to the rehearsal-based continual learning setting, which requires storing training data and thus is potentially subject to privacy issues, the proposed setting presents a realistic modeling of practical scenarios where new relations are incrementally learnt, and no training data is stored. This extension is worth discussion and exploration.
2. The proposed Ensemble-of-Expert framework to tackle the problem is reasonable and achieved good performance on both settings.

Summary Of Weaknesses:
1. The authors claim that rehearsal-free CRE is a new setting to tackle the privacy issues under application cases where training data are not allowed to be stored. However, there are many other methods to alleviate this, such as through federated learning. With that said, the value of the proposed rehearsal-free setting might be questionable.

Comments, Suggestions And Typos:
None.

Soundness: 3.5
Overall Assessment: 3.5

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Best Paper: No

Ethical Concerns:
None.

Needs Ethics Review: No

Reproducibility: 4 = They could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method.

Datasets: 1 = No usable datasets submitted.

Software: 3 = Potentially useful: Someone might find the new software useful for their work.

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources, other (specify)

Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Reviewer Certification: zZ9N

[−] Response to insightful comments of Reviewer zZ9N

ACL ARR 2024 February Paper2376 Authors ● *Shen Zhou (/profile?id=~Shen_Zhou2) (privately revealed to you)*

29 Mar 2024, 20:59 (modified: 31 Mar 2024, 16:38) ● ACL ARR 2024 February Paper2376 Official Comment ● Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors ● Show Revisions (/revisions?id=U2fWCnV76Q)

Comment:
Dear Review zZ9N,

Thanks for your valuable comments! We are encouraged by your recognition of our proposed RFCRE problem and the corresponding new framework. Below is the response to your suggestion.

Suggestion: The authors claim that rehearsal-free CRE is a new setting to tackle privacy issues under application cases where training data are not allowed to be stored. However, there are many other methods to alleviate this, such as through federated learning. With that said the value of the proposed rehearsal-free setting might be questionable.

Response: Many thanks for your insightful comments, which help us further clarify the motivation. Indeed, federated learning and continual learning have their unique task scenarios.

As you pointed out, federated learning can also be used to protect privacy, but most existing federated learning frameworks have data categories that remain fixed over time. The continual learning scenario addressed in this paper requires the model to learn new relations (data categories) as time progresses continuously. Moreover, existing CRE methods need to store some samples of old relations, which would lead to potential privacy security issues. Our rehearsal-free continual relation extraction is proposed to continuously learn new categories while protecting the privacy of old data.

Thanks again for the valuable comment. We will further clarify the motivation in the later version.

[−] Official Review of Paper2376 by Reviewer aw2T

ACL ARR 2024 February Paper2376 Reviewer aw2T

07 Mar 2024, 01:33 ● ACL ARR 2024 February Paper2376 Official Review ● Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors ● Show Revisions (/revisions?id=blXMPI5c_Pn)

Recommended Process Of Reviewing: I have read the instructions above

Paper Summary:
The paper introduces an Ensemble-of-Experts (EoE) framework designed for Continual Relation Extraction (CRE) without the need for rehearsal, addressing the challenge of catastrophic forgetting. It proposes a novel method that leverages multiple expert models, each specialized in a different task, combined through a cascade voting mechanism for decision-making. This approach allows for efficient learning of new tasks while preserving knowledge of previously learned tasks. The EoE framework is evaluated against traditional methods, showing superior performance in both rehearsal-free and rehearsal-required settings, demonstrating its effectiveness and efficiency in continual learning environments.

Summary Of Strengths:

The paper's strengths lie in its innovative approach to addressing the challenge of catastrophic forgetting in Continual Relation Extraction (CRE) without requiring rehearsal. It introduces the Ensemble-of-Experts (EoE) framework, which not only innovates in task-specific model management but also in how it integrates the expertise of these models through a cascade voting mechanism. This methodology allows for efficient learning and memory retention across tasks. The empirical evaluation demonstrates the framework's superior performance compared to existing methods, showcasing its practicality and potential for broad application in the field of CRE.

Summary Of Weaknesses:

1. Dependency on Accurate Task Identification: The framework's performance heavily relies on the accurate identification of tasks for each input. Misidentification can lead to the wrong expert being queried, potentially degrading performance.
2. Complexity in Expert Management: As the number of tasks increases, managing an ensemble of experts becomes more complex. This complexity could impact the scalability of the model, making it less efficient for very large numbers of tasks.
3. Generalization across Tasks: The framework might face challenges in generalizing across significantly diverse tasks. While it performs well in scenarios presented in the paper, its effectiveness in broader, more varied contexts is not thoroughly explored.
4. Resource Intensity: Implementing and maintaining multiple expert models for different tasks can be resource-intensive, requiring significant computational power and memory, which may not be feasible for all applications.
5. Adaptability to Dynamic Environments: The paper does not address how the framework adapts to dynamically changing environments where new information might alter the nature of existing tasks or introduce unforeseen tasks.

Comments, Suggestions And Typos:

Enhance Task Identification: Develop more robust techniques for task identification to reduce misclassification and improve the overall performance of the framework.

Optimize Expert Management: Investigate strategies for managing the ensemble of experts more efficiently, perhaps through dynamic model pruning or consolidation, to address scalability and complexity issues.

Broaden Generalization Studies: Conduct additional experiments across a wider variety of tasks and domains to evaluate and improve the framework's generalization capabilities.

Reduce Resource Intensity: Explore methods to decrease the computational and memory requirements of the ensemble model, such as model compression or shared representations, to make it more feasible for practical applications.

Improve Adaptability: Implement mechanisms for the framework to adapt to dynamic environments more effectively, possibly through online learning or incremental model updates.

Soundness: 3 = Acceptable: This study provides sufficient support for its major claims/arguments. Some minor points may need extra support or details.

Overall Assessment: 3 = Good: This paper makes a reasonable contribution, and might be of interest for some (broad or narrow) sub-communities, possibly with minor revisions.

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Best Paper: No

Ethical Concerns:

None

Needs Ethics Review: No

Reproducibility: 4 = They could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method.

Datasets: 3 = Potentially useful: Someone might find the new datasets useful for their work.

Software: 1 = No usable software released.

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources

Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Reviewer Certification: HNGt

[-] **Response to insightful comments of Reviewer aw2T (1 / 2)**

29 Mar 2024, 21:03 (modified: 29 Mar 2024, 21:25)   ACL ARR 2024 February Paper2376 Official Comment   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=4tjk4EaklV)

Comment:

Dear Review aw2T,

We really appreciate your constructive suggestions on our work, which will certainly deepen future research in this direction.

Suggestion 1: Enhance Task Identification: Develop more robust techniques for task identification to reduce misclassification and improve the overall performance of the framework.

Response 1: As you pointed out, the continual learning framework in this paper would rely on the accuracy of task identification. Existing two-phase methods [1, 2] in other fields also suffer this drawback.

We are also aware of this issue and thus propose a cascade voting mechanism by invoking experts from different stages collectively to improve task identification accuracy. In addition, we introduce discriminative training to enhance the task identification capability of expert models. Overall, as shown in the ablation experiments (Table 2) in the paper, the proposed cascade voting mechanism and discriminative training can both effectively improve the accuracy of task identification. Finally, thank you again for the suggestions, and we will continue to make more efforts to develop more robust techniques for task identification.

Suggestion 2: Optimize Expert Management: Investigate strategies for managing the ensemble of experts more efficiently, perhaps through dynamic model pruning or consolidation, to address scalability and complexity issues.

Response 2: Thank you very much for your constructive suggestions. Dynamic model pruning or consolidation is indeed a feasible research direction that will be considered in our future work. Thanks again!

Suggestion 3: Broaden Generalization Studies: Conduct additional experiments across a wider variety of tasks and domains to evaluate and improve the framework's generalization capabilities.

Response 3: Per your suggestion, we select a Chinese relation extraction dataset, DuIE2.0 [3], to validate the domain generalization of our framework. The corpus for DuIE2.0 comes from Baidu Baike, Baidu Information Flow, and Baidu Tieba texts, covering formal and informal expression corpora. We select 40 relation categories and divide them into ten tasks, each containing 100 training samples and 100 test samples, so as to keep the experimental setup consistent with our submitted paper. The experimental results are shown in Table 4.

Table 4: Additional experiments compared with EPI on the DuIE2.0 dataset regarding task identification accuracy (IA) and classification accuracy (CA).

IA	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
EPI	100.0	92.5	90.1	88.2	87.7	86.7	86.0	85.2	84.5	84.4
EoE	100.0	95.8	93.8	91.4	91.1	90.4	89.4	88.8	88.1	87.6

CA	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
EPI	98.3	89.9	88.6	86.8	86.4	85.3	84.8	84.2	83.5	83.5
EoE	99.4	94.3	92.7	89.9	89.7	89.1	88.2	87.7	87.0	86.6

Table 4 shows that even in other domains, our proposed EoE model performs significantly better than the best baseline EPI, demonstrating that our framework has certain generalization capabilities across different domains.

[-] **Response to insightful comments of Reviewer aw2T (2 / 2)**

29 Mar 2024, 21:08 (modified: 29 Mar 2024, 21:25)   ACL ARR 2024 February Paper2376 Official Comment   Readers: Program Chairs, Paper2376 Senior Area Chairs, Paper2376 Area Chairs, Paper2376 Reviewers Submitted, Paper2376 Authors   Show Revisions (/revisions?id=be7gPcLkjf)

Comment:

Suggestion 4: Reduce Resource Intensity: Explore methods to decrease the computational and memory requirements of the ensemble model, such as model compression or shared representations, to make it more feasible for practical applications.

Response 4: It is true that simultaneously maintaining multiple experts for task identification would incur extra cost overhead. To alleviate this problem, we limit the maximum number of experts involved in task identification by setting a threshold m in the cascade voting mechanism. Below, we show the impact of m in Table 5 and Table 6.

Table 5: Impact of the threshold m on FewRel regarding task identification accuracy (IA).

IA	T6	T7	T8	T9	T10
m=1	91.11	89.94	88.86	87.60	86.06
m=2	91.71	90.65	89.60	88.32	86.88
m=3	91.75	90.67	89.60	88.37	86.95
All	91.81	90.76	89.74	88.59	87.22

Table 6: Impact of the threshold m on TACRED regarding task identification accuracy (IA).

IA	T6	T7	T8	T9	T10
m=1	85.53	84.07	82.71	81.88	80.91
m=2	87.31	85.78	84.53	84.11	82.80
m=3	87.26	85.71	84.72	83.86	82.73
All	87.47	85.87	84.80	84.06	83.07

From Tables 5 and 6, we can see that as the number of experts increases, the accuracy of task identification generally improves, demonstrating that increasing the number of experts can improve the framework's ability for task identification. While increasing m from 1 to 2 shows big improvements, increasing m to 3 or more does not yield much enhancement. That is to say, we just need to maintain a few expert models for task identification. But surely, the model compression and shared representations that you proposed are good research directions.

Suggestion 5: Improve Adaptability: Implement mechanisms for the framework to adapt to dynamic environments more effectively, possibly through online learning or incremental model updates.

Response 5: Continual relation extraction addressed in this paper belongs to the class-incremental continual learning scenario. Following previous methods [4], we also assume that the relation of test samples must belong to the learned relations, which indeed cannot handle unseen relations at the inference stage.

Moreover, continual learning is still under development, and the definitions and scenarios of tasks are becoming more and more aligned with real-world scenarios. This paper is the first to consider potential privacy issues in CRE and targets rehearsal-free CRE. Online learning, as you mentioned, can indeed serve as a future direction of improvement for further adaptation to dynamically changing environments. Thank you for your insightful suggestions.

References:

[1] Rehearsal-free Continual Language Learning via Efficient Parameter Isolation. ACL 2023

[2] Isolation and Impartial Aggregation: A Paradigm of Incremental Learning without Interference. AAAI 2023

[3] DuIE: A Large-Scale Chinese Dataset for Information Extraction. NLPCC 2019

[4] Improving Continual Relation Extraction by Distinguishing Analogous Semantics. ACL 2023

Finally, we must thank you again for your suggestions, which have provided us valuable inspiration for our future work. We look forward to your response.

[−] **Supplementary Materials by Program Chairs**

Reassignment Request Action Editor: This is not a resubmission

Reassignment Request Reviewers: This is not a resubmission

A1: yes

A1 Elaboration For Yes Or No: Section Limitations

A2: yes

A2 Elaboration For Yes Or No: Section Ethics Statement

A3: yes

A3 Elaboration For Yes Or No: Abstract; Section Introduction

B: yes

B1: yes

B1 Elaboration For Yes Or No: Section 5.1

B2: n/a

B3: n/a

B4: n/a

B5: n/a

B6: yes

B6 Elaboration For Yes Or No: Appendix C.1

C: yes

C1: yes

C1 Elaboration For Yes Or No: Appendix C.3

C2: yes

C2 Elaboration For Yes Or No: Section 5.1

C3: yes

C3 Elaboration For Yes Or No: Appendix C.3

C4: n/a

D: no

D1: n/a

D2: n/a

D3: n/a

D4: n/a

D5: n/a

E: no

E1: n/a

Note From EiCs: These are the confidential supplementary materials of the submission. If you see no entries in this comment, this means there haven't been submitted any.

About OpenReview (/about)

Hosting a Venue (/group?
id=OpenReview.net/Support)

All Venues (/venues)

Contact (/contact)

Feedback

Sponsors (/sponsors)

Frequently Asked Questions
(https://docs.openreview.net/getting-
started/frequently-asked-
questions)

Terms of Use (/legal/terms)

Privacy Policy (/legal/privacy)