COMP61332 Text Mining Coursework Specifications

Introduction

Relation Extraction (RE) is a task that is concerned with identifying the type of relationship, if any, that holds between two given named entities, based on the sentence in which they were mentioned.

In this coursework, your group (with 3-4 members) will develop your own RE methods and evaluate them using a benchmark dataset of your choice.

Intended Learning Outcomes

- to develop two different relation extraction methods that will be applied to an existing labelled dataset of your own choice
- to evaluate and compare the said RE methods
- to describe the said RE methods and analyse their performance, in the form of a short paper
- to act as a responsible member of a team, communicate with team mates, and contribute to the team's self-organisation, planning and conflict resolution for the duration of the group work

Instructions

Carefully read each of the sections below and take note of the corresponding deadlines.

A. Selection of a suitable dataset

Together with your group mates, look for a dataset that contains RE annotations, that you can use to develop and evaluate two RE approaches of your own choice. Below are some factors that you should take into consideration when choosing a dataset:

1. Availability: The dataset should be publicly available and should be easily accessible to you. Some datasets, though suitable, might be available only upon request; these are not ideal as the time spent on waiting for your request to be approved will likely lead to delays in your method development.
2. Size: Although we do not require that the dataset should contain a certain minimum number of examples, you might want to consider whether there are enough to support the development of your intended approaches. For example, if you wish to train a traditional machine learning-based model, you might need at least thousands of training examples.
3. Comparability: Later on, when you analyse the performance of your RE methods, you might want to be able to draw comparisons with the performance obtained in previously reported work. In this case, it would be useful to check whether the dataset that you are considering includes an evaluation set (often referred to as "test set" or "validation set") that other researchers have used to evaluate their work.
4. Reliability: You might want to check if the RE annotations included in the dataset are of sufficient quality. Look for papers that describe the dataset and find out how the developers/authors of the dataset ensured reliability (e.g., by measuring inter-annotator agreement).

Examples of suitable RE datasets include FewRel and the SemEval 2010 Task 8 shared task dataset. Each of these are publicly available, contain thousands of annotated examples,

include an annotated evaluation set based on which model performance is typically reported, and their respective papers[1] report how the dataset creators ensured annotation reliability.

There are online resources that can help you in finding datasets, examples of which are Papers with Code,[2] Kaggle and Huggingface.

IMPORTANT NOTE: To make sure that your proposed dataset is suitable, we require you to complete a questionnaire before the Week 3 lab sessions. This will NOT be marked but it will allow us to give you feedback in the Week 3 lab session in case we find any issues.

Deliverable: A completed questionnaire regarding the dataset that you propose to use. This is due on the **10th February 18:00**. Only one submission per group is required; choose a group representative who will be responsible for submitting the completed questionnaire on behalf of the whole group.

B. Development and evaluation of RE methods

Using your chosen dataset as your training and evaluation corpus, develop **two** different approaches to RE. These approaches should come from any two of the following categories:

- Symbolic or unsupervised approach (e.g., using dictionaries/rules, bootstrapping)
- Traditional machine learning-based approach (e.g., support vector machines)
- Deep learning-based approach that does not use transformer models (e.g., LSTMs)
- Deep learning-based approach based on transformer models (e.g., BERT)
- A hybrid of any of the above

The goal is NOT to beat the state-of-the-art performance reported in literature but to: (1) enhance and/or experiment with different approaches, (2) compare the performance of those approaches on the evaluation dataset; (3) analyse, based on your results, the strengths and weaknesses of the approaches.

The last section of this document presents a list of resources that you can refer to or build upon.

You are allowed to reuse existing codebases as long as proper attribution is given to those resources. There are different ways of doing this, including, but not limited to, carefully documenting reused resources in a README file, or providing attribution within Python notebooks.

IMPORTANT NOTE: Failure to explicitly acknowledge that you reused (parts of) someone else's code will be treated as academic malpractice.

Deliverable: A codebase consisting of readily runnable Python notebooks containing your code for developing and evaluating your two RE methods. If your methods require certain packages, make sure that the install statements for those are included in the notebooks.

In your code, there should be a way to run your RE methods in real time (i.e., in inference mode):

- A user should be able to supply some input, formatted in a way that you expect (this should be clearly documented/indicated); and

[1] See https://aclanthology.org/D18-1514.pdf for FewRel and https://aclanthology.org/S10-1006.pdf for the SemEval 2010 Task 8 shared task.
[2] For example, you can try looking for datasets containing RE-annotated English texts via https://paperswithcode.com/datasets?task=relation-extraction&lang=english&mod=texts

- Each of your methods should be able to take that input and then output any extracted relations.

Create a zip file containing your notebooks, README and any other relevant resources. You are welcome to include any resources, even models, as long as the resulting zip file **does not exceed 10MB**. If it does, upload any large files onto an online storage server such as OneDrive or Google Drive, and include links to them in a README, making sure that we have access to all the information and data that we need to mark your deliverable. Your zip file is due on the **7th March 18:00**. Only one submission per group is required; choose a group representative who will be responsible for submitting the zip file on behalf of the whole group. Note that the zip file should also include a short paper (see succeeding section).

C. Paper writing

Your group should prepare a short paper written for an academic audience, similar to papers published in NLP conferences or arXiv. In this paper, you will: (1) provide an introduction to the RE task and describe your chosen RE dataset; (2) review previously reported work that is related to your approaches; (3) describe the two RE approaches that you developed, justifying why they were chosen and detailing how they were implemented; (4) describe how you evaluated the approaches (e.g., using which metrics); and (5) discuss the results of your evaluation, comparing the approaches and highlighting their strengths and weaknesses.

We require you to use a two-column, standard format used by members of the NLP community, that is available in a Word, LaTeX or Overleaf template[3] (choose one to your liking). The paper content should be **at most four pages long**, plus any number of pages for references and any appendices.

IMPORTANT NOTE: Do not modify the template. For example, we expect that you will use the font, font size and margins that are pre-specified in the template linked above.

You are free to structure the content of the paper as you wish. For inspiration on paper sections that are typically included, you can refer to the following two papers:

> Yang, S., Zhang, Y., Niu, G., Zhao, Q., & Pu, S. (2021). *Entity Concept-enhanced Few-shot Relation Extraction.* In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (pp. 987-991).
>
> Miller, C., & Vosoughi, S. (2020). *Big Green at WNUT 2020 Shared Task-1: Relation Extraction as Contextualized Sequence Classification.* In Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020) (pp. 281-285).

Note that the examples above are meant to show you ways on how you can structure your short paper, and are not meant to imply that the approaches reported in those papers are favoured over others.

IMPORTANT NOTE: All group members should be named as authors of the short paper.

Deliverable: A short paper describing your group's work. This should be included in the zip file containing your codebase, that should be submitted on the **7th March 18:00.**

[3] If you are using the LaTeX or Overleaf template, make sure that (before submission) you change the line that says **\usepackage[review]{ACL2023}** to **\usepackage{ACL2023}**, so that your names are shown in the generated PDF file.

Timeline

10th Feb	Submission of questionnaire regarding your proposed dataset (by 18:00) Only one response required per group: https://forms.office.com/e/Xu1HZp50Ef
12th Feb	Feedback (on selected datasets) given to students during lab sessions
07th Mar	Submission of codebase and short paper (by 18:00, via **Blackboard**) Only one submission required per group

Marking Scheme

This coursework accounts for 50% of your final mark for COMP61332. The following rubric will be used in marking your group project, where the first column specifies the various criteria and the second column indicates the maximum number marks your group can possibly be given (for a total of 40 possible marks).

Implementation		
Method 1 Functionality	0	The method is not functional and does not produce any output (extracted relations).
	2	The method is functional but is practically unusable, e.g., it takes too long to run or it requires too much memory, making it difficult to obtain any output.
	4	The method is functional and is able to produce extracted relations in reasonable time.
Method 1 Adaptations	0	There was no attempt to incorporate any enhancements or novelties to known implementations. The submitted method is very similar to an existing implementation or codebase, or is simply relying on APIs (e.g., OpenAI).
	2	Some adaptations and enhancements were attempted but more effort or creativity could have been put into these.
	4	It is evident that novel adaptations and enhancements were implemented.
Method 2 Functionality	0	The method is not functional and does not produce any output (extracted relations).
	2	The method is functional but is practically unusable, e.g., it takes too long to produce extracted relations, or it requires too much memory, making it difficult to obtain any output.
	4	The method is functional and is able to produce extracted relations in reasonable time.
Method 2 Adaptations	0	There was no attempt to incorporate any enhancements or novelties to known implementations. The submitted method is very similar to an existing implementation or codebase, or is simply relying on APIs (e.g., OpenAI).
	2	Some adaptations and enhancements were attempted but more effort or creativity could have been put into these.
	4	It is evident that novel adaptations and enhancements were implemented.
Evaluation	0	There was no attempt to evaluate the methods.
	2	The methods were evaluated but without using metrics that are suitable for the task or dataset.

	4	It is evident that the methods were evaluated using suitable metrics.
Documentation	0	The submitted code does not contain any form of documentation (e.g., in-line comments, descriptions).
	2	The submitted code contains some documentation but not enough to clearly explain any adaptations/enhancements made.
	4	The submitted code contains documentation that clearly explains any adaptations/enhancements made.
Short paper		
Academic writing and presentation	0	The short paper looks like a technical report/user manual rather than an academic paper. References are lacking or are not properly cited/presented.
	2	The short paper was written for an academic audience. However there are some ideas which were not clearly presented, or the discussion lacked originality/argumentation.
	4	The short paper was written for an academic audience and can potentially be published in a research workshop or symposium. Ideas were presented in a clear and well-argued manner.
Introduction and Related Work	0	The paper does not provide an introduction to the task/dataset (e.g., why it is interesting) and the review of related work is very minimal.
	2	The paper provides an introduction to the task/dataset. It also provides a review of related work although there are some shortcomings, e.g., it is not clear how these relate to the group's own work.
	4	The paper provides a convincing introduction to the task/dataset. It also provides a good review of related work, which shows how previous work relates to the group's own work.
Methodology	0	The paper does not provide sufficient details on the group's methods.
	2	The paper provides details on the group's methods, although some parts need further elaboration or justification.
	4	The paper provides sufficient and clear details on the group's methods. Design decisions relating to the two chosen RE methods have been justified well.
Results and Discussion	0	Results of the group's methods were not discussed and interpreted.
	2	Results of the group's methods were discussed and interpreted. However, in some parts the interpretation seems exaggerated, or a deeper analysis could have been carried out.
	4	Results of the group's methods were discussed and interpreted in a convincing way.

Your deliverables will be assessed based on the marking scheme above, which will lead to one overall mark (out of 40). Everyone in your group will be given the same mark: one of the Learning Outcomes of this coursework is focussed on acting as a responsible team member (see Intended Learning Outcomes), hence it is everyone's duty to ensure that tasks are delegated fairly, that there are equal contributions, and that integration goes smoothly.

In the exceptional case where one or more group members have not put in an acceptable contribution, despite the team's effort to bring them back into the team and suitable discussions of the issues arising, we implement a grievance procedure. A group can bring forward a "case of grievance" to the COMP61332 teaching staff by providing the following pieces of evidence:

- minutes of team meetings
- a written description of the events that led to the break-down of the team work, including a description of the actions that were taken to get the team back on track

The case should be brought forward to the teaching staff no later than one working week after the coursework deadline. The COMP61332 teaching staff will then decide whether the situation is indeed exceptional and warrants a mark re-distribution.

Potentially helpful resources

A. Relation Extraction Survey Papers

Bassignana, E., & Plank, B. (2022). What Do You Mean by Relation Extraction? A Survey on Datasets and Study on Scientific Relation Classification. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop (pp. 67-83).

Detroja, K., Bhensdadia, C. K., & Bhatt, B. S. (2023). A Survey on Relation Extraction. Intelligent Systems with Applications, 200244.

B. Traditional machine learning-based approaches to RE

Hong, G. (2005). Relation Extraction Using Support Vector Machine. In Second International Joint Conference on Natural Language Processing: Full Papers.

Zhou, G., Su, J., Zhang, J., & Zhang, M. (2005). Exploring various knowledge in relation extraction. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics. ACL 05 (pp. 427-434).

Rink, B., & Harabagiu, S. (2010). UTD: Classifying Semantic Relations by Combining Lexical and Semantic Resources. ACL 2010 (pp. 256-259).

C. Deep learning-based approaches that do not use transformer models

Zhang, S., Zheng, D., Hu, X., & Yang, M. (2015). Bidirectional long short-term memory networks for relation classification. In Proceedings of the 29th Pacific Asia conference on language, information and computation (pp. 73-78).

Valette, Marion (2019). Simple Relation Extraction with a Bi-LSTM Model. Available online: https://medium.com/southpigalle/simple-relation-extraction-with-a-bi-lstm-model-part-1-682b670d5e11

D. Deep learning-based approaches based on transformer models

Wu, S., & He, Y. (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 (pp. 2361-2364).

Thillaisundaram, A., & Togia, T. (2019). Biomedical relation extraction with pre-trained language representations and minimal task-specific architecture. In Proceedings of the 5th Workshop on BioNLP Open Shared Tasks (pp. 84-89).