

Data Science

Natural Language Processing

NLP Papers Summary

Day 143: NLP Papers Summary – Unsupervised Pseudo-Labeling For Extractive Summarization On Electronic Health Records

By Ryan 22nd May 2020 No Comments

Objective and Contribution

Proposed an effective unsupervised method of generating pseudo-labels for Electronic Health Records (EHRs) by taking advantage of the intrinsic correlation between multiple EHRs. We then used this pseudo-labels to train an extractive summariser to help physicians better digest EHRs. Specifically this paper answers the following three research questions:



1. How to measure the quality of disease-specific summarisation for the same patient:

2. How can we use research question 1 to generate effective and accurate pseudo-labels? ^
3. Given the generated pseudo-labels, what model architecture should be used for summarisation in a medical setting?

Methodology

PROBLEM DEFINITION

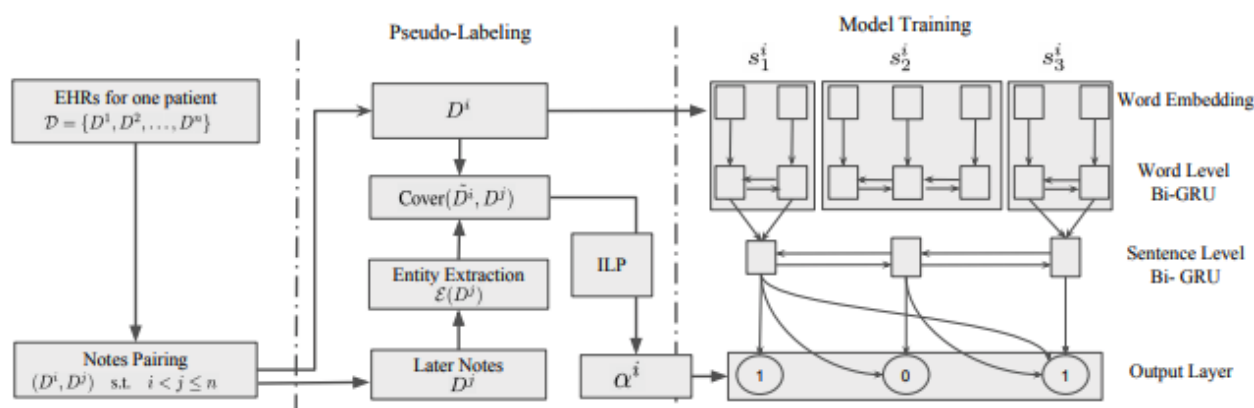


Figure 1: The workflow of our unsupervised summarization method for EHRs. Left part is the notes pairing process. Middle part generates pseudo-labels, which are fed to train the neural model on right part.

For most patients, there exist many EHRs recorded over time. Our objective is to find a subset of these EHRs that best summarises patient's information for a specific disease. The figure above showcase our overall architecture. For the first research question, we observed that physicians tend to focus on medical entities when reading and summarising clinical notes and so we proposed to summarise clinical notes to cover more relevant entities.

Each EHR has an entity set that contains all the medical entities. As you can imagine, each EHR could contains more than hundreds of entities and capturing all of them is a very challenging problem. To capture these entities, we observed that information in earlier health record usually persist through to later records, reminding physicians to pay attention for future treatments. Inspired by this, we use a coverage score to evaluate the quality of an EHR based on one of its later records. We used inverse document frequency to measure the importance of an entity in the entire corpus and measures semantic similarity between the entity and sentences in the EHR by encoding them using word embeddings trained on PubMed.



PSEUDO-LABELLING WITH INTEGER LINEAR PROGRAMMING (ILP)

We used integer linear programming with length constraints to generate binary pseudo-labels for an EHR using later records. The following is the final optimisation problem:

$$\begin{aligned} \max \quad & \sum_{e \in \mathcal{E}(D^j)} \lambda(e) \cdot \log \left(\sum_{t=1}^{|D^i|} \exp(\alpha_t^i \cdot \text{sim}(e, s_t^i)) \right) \\ \text{s.t.} \quad & \sum_{t=1}^{|D^i|} \alpha_t^i \cdot |s_t^i| \leq L, \quad \text{and} \quad L_1 \leq \sum_{t=1}^{|D^i|} \alpha_t^i \leq L_2, \quad \forall i \end{aligned}$$

SUMMARISATION MODEL

Here, we use the generated pseudo-labels to train a supervised extractive summarisation model to summarise medical records. The model consists of two-layer biGRU, where the first one focuses on the word level and generate sentence embedding and the second one focuses on sentence level using the output from the first layer and compute the final representation for each sentence. For the output layer, we have logistic function with several features including content, salience, novelty, and position. The salience feature will helps us identify how important the current sentence is to the entire note and the novelty feature helps us with reducing redundancy.

Experiments and Results

Our evaluation dataset is Medical Information Mart for Intensive Care III (MIMIC-III). We extracted a total of 5875 admissions that contain at least one diagnostic ICD code related to heart disease. We also utilise the clinical notes from the NOTEEVENTS table. We hired an experienced physicians to manually annotated 25 clinical notes and compared the results of our models during inference time. Our evaluation metric is the standard summarisation metric of ROUGE scores.

MODELS COMPARISON

Our approach is unsupervised since we don't required any external annotations and so all our baseline models comparison are unsupervised:

1. *Most-Entity (ME)*. Picks sentences with most medical entities
2. *TF-IDF*.



3. *TF-IDF + MMR (TM)*. Extension of TFIDF which aims to reduce duplication of informat ^ as MMR measures reduces the weight of sentence that has high similarity to the sentences that have already been selected
4. *Ablation variation of our own model*. We have our model without novelty feature, without position feature, and our full model

RESULTS

As shown in the result table above, our model achieved the highest ROUGE scores against all our baseline models. We also observed that redundancy heavily affects our summarisation models. Both MMR and novelty feature improves the performance of TF-IDF and our model significantly. The position feature has also proven to improve performance and this is to be expected as clinical notes are often written with a structured template.

The table below showcase some examples of extracted sentences by our models and compares it to the physician's annotation. From the table, we can see that ME and TM have strong tendency to select long sentences with more entities, which it's to be expected. This means that they both failed to select important short sentences like sentence 1 and 2. Our model also suffers from this problem but it's less severe. The weakness of TM is that by using TFIDF, when sentences contain infrequent terms, it can mislead the model into classifying it as being important but in medical domain, terms are often very specific and so although infrequent terms, they might not be relevant to a particular disease. Our model is shown to select sentences very similar to our physician's annotation.



Conclusion and Future Work

Overall, we explored three research questions in summarising EHRs:

1. Used medical entities to cover intrinsic correlation between multiple EHRs for one patient
2. Developed an optimisation target and used ILP to generate pseudo-labels
3. Used the generated pseudo-labels to train our supervised extractive summarisation model

Potential future work could involve adding new features such as coverage or attention mechanism to our extractive summarisation model to avoid duplications and paying special attentions to parts of the sentence that are important.

Source: <https://arxiv.org/pdf/1811.08040.pdf>

Ryan

Data Scientist

Previous Post

Day 142: NLP
Papers Summary -

Next Post

Day 144: NLP
Papers Summary - 



Measuring
Emotions in the
COVID-19 Real
World Worry
Dataset

Attend to Medical
Ontologies:
Content Selection
for Clinical
Abstractive
Summarization



[Data Science](#) [Natural Language Processing](#) [NLP Papers Summary](#)

Day 365: NLP Papers Summary – A Survey on Knowledge Graph Embedding

Ryan

30th December 2020



[Data Science](#) [Implementation](#) [Natural Language Processing](#)

Day 364: Ryan's PhD Journey – OpenKE-PyTorch Library Analysis + code snippets for 11 KE models

Ryan

29th December 2020





[Data Science](#) [Ryan's PhD Journey](#)

Day 363: Ryan's PhD Journey – Literature Review – Knowledge Acquisition – 1st Passes XIV

Ryan

28th December 2020



