Research Summary

Hieu Vu

Monday 4th March, 2024

My interests lie in the broad topic of improving the modelling capability of Deep Learning models, which refers to anything related to: *information and knowledge representation, novel architectures and modelling paradigms, learnable causality and reasoning, effective and efficient training methods, etc.* This includes but not limited to topics such as *Modularity, Multimodality, Representation Learning, and Causal Deep Learning.*

However, working in the industry means that my research is heavily influenced by the requirements of my employer. Furthermore, my employer is not a research organization, thus guidance for research was limited which required me to develop my own methodology. The works presented here are my attempts to align my interests with my responsibility as an employee, along with my attempts to explore new areas to broaden my knowledge and refine my interests.

This document briefly describes some highlights of my research works. These came from my full-time job, my undergraduate thesis, and my personal projects. The works are categorized into two main sections: *Information Extraction from business documents*, which is a focus of my full-time job and my undergraduate thesis, and *Personal exploration*, which contains other topics that I dabbled in during my personal time.

A list of my arXiv and peer-reviewed publications can be found in my Google Scholar profile.

Contents

Information Extraction from business documents

At work, my team and I helped clients extract key-value information from their business documents, thus, most of my research focused on this area.

A Span Extraction Approach for Information Extraction on Visually-Rich Documents

Accepted to DIL@ICDAR 2021 - Best Paper Award
https://arxiv.org/pdf/2106.00978.pdf

This work formulated the task of key-value information extraction from document images as a span extraction task and presented two following ideas:

- A cross-lingual transfer learning method for adapting LayoutLM to a low-resource language.
  - LayoutLM (an extension of BERT for document images by taking into account visual and spatial information) was the new state-of-the-art method at the time, but it was pre-trained in English only.
  - But our data is in Japanese, and we did not have the data nor the computation needed to pre-train the model from scratch.
  - Thus, my idea was to swap the text embeddings of the English LayoutLM with those of a Japanese BERT (Fig. 1). Then continue to pretrain the model for a short amount of time on a much smaller dataset (17k compared with over 1M).

  The hypothesis was that:

  - Positional features are language-independent and can be shared across languages with alike reading order.
  - Furthermore, the encoder layers are capable of capturing attention from both semantic and positional inputs.

  This contribution was my original idea.

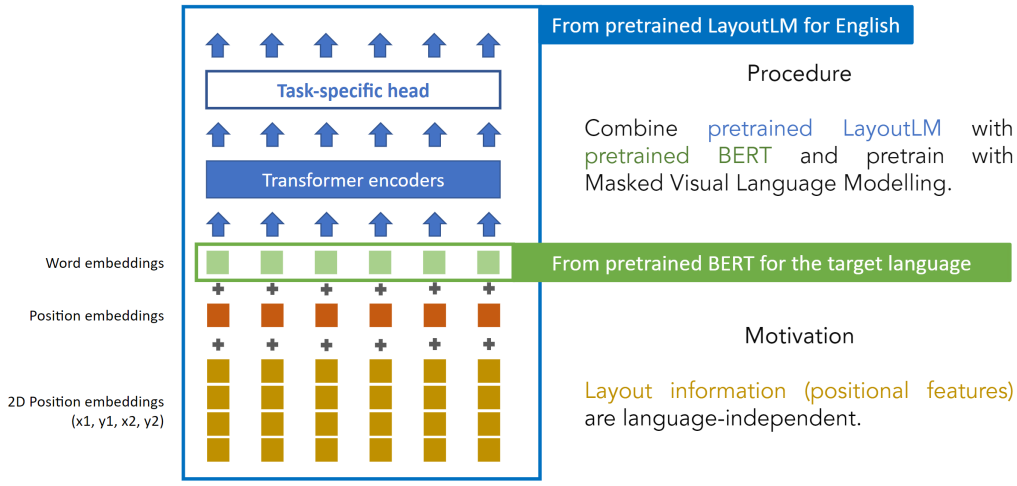


Figure 1: Swapping LayoutLM's word embeddings with BERT's embeddings for cross-lingual transfer on a low-resource language.

- A recursive relation predicting scheme for multi-span extraction.

One problem with span extraction is it only predicts a fixed number of spans for each query. However, in our case, there can be multiple different values for a single key information.

This method addressed it by:

- Extracting the spans as a chain in a recursive manner.
- After the first span is extracted, the embeddings of its first token will be used as the query to extract the second span.
- The process continues until a stopping condition is reached (Fig. 2).

This can be used both as a downstream task for multi-span extraction and also as a second-stage pretraining task.

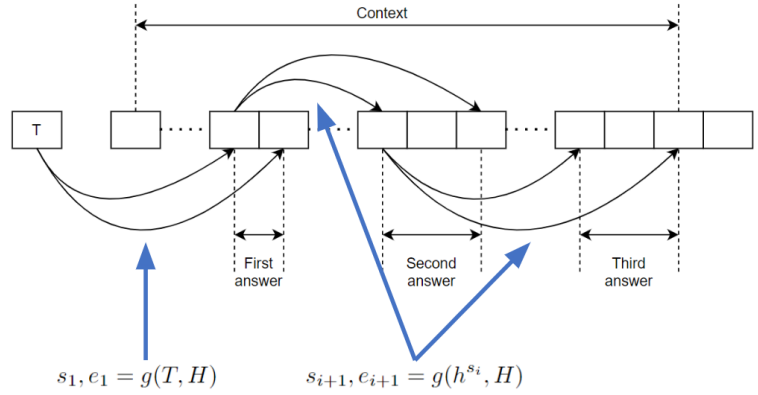


Figure 2: The recursive span extraction mechanism.

Jointly Learning Span Extraction and Sequence Labeling for Information Extraction from Business Documents

Accepted to IJCNN 2022 - Oral presentation
https://arxiv.org/pdf/2205.13434.pdf

This work aimed at the task of extracting multiple short-span key information from a long document. More specifically, it tried to address the following issues:

- Due to the target information only making up a small portion of the document, a sequence labelling approach is affected by data imbalance.
- A span extraction approach is less affected by data imbalance, but it cannot extract multiple spans for a single key information. It also takes a long time to train and inference due to only being able to extract a single information at a time.

To this end, the paper proposed the following:

- A query-based span extraction scheme that can process multiple queries at the same time.
  - The queries refer to key information that needed to be extracted (e.g. *contract date* or *company name*).
  - The queries are represented as learnt embeddings vectors.
  - Then, attention scores are computed between the queries and the document.
  - Finally, the attention scores are used to determine the start and end positions of the span.

  This setup enables processing multiple queries at the same time, which reduces training and inference time significantly.
- Jointly learning span extraction and sequence labelling on two different branches, then combine the predictions.
  - Since the above idea enables span extraction on multiple queries simultaneously just like sequence labelling, the model can perform sequence labelling on a parallel branch to get the best of both worlds.
  - The sequence labelling branch allows the model to extract multiple values for each query.
  - Furthermore, optimizing two downstream tasks simultaneously also improves the gradient, which contributes toward better performance and faster convergence.

  This was my contribution to the methodology of this paper.

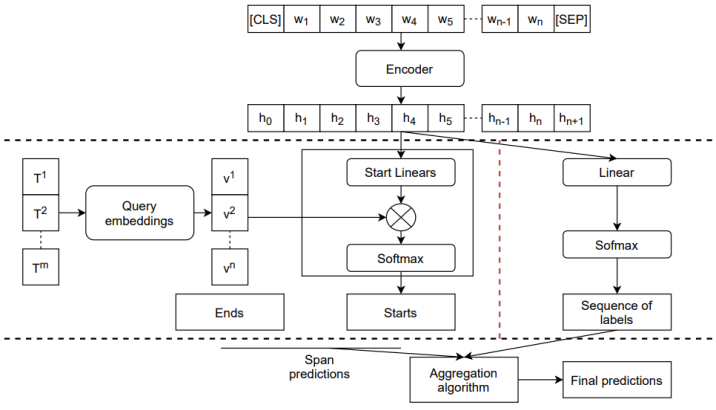


Figure 3: The proposed model, which jointly learns span extraction and sequence labelling.

A layout-aware key-value relation predicting model for document images

Undergraduate thesis
Advised by Dr. Diep Thi-Ngoc Nguyen

My undergraduate thesis focused on detecting key and value regions in document images by formulating it as a segmentation problem. In which I presented the following:

- A revised version of the FUNSD dataset
  - Upon inspecting the FUNSD dataset, I found it to be erroneous and have inconsistent labelling logic (An example is given in Fig. 4).
  - Thus I made the correction and published the revised version along with a report. The report is available at https://arxiv.org/pdf/2010.05322.pdf.

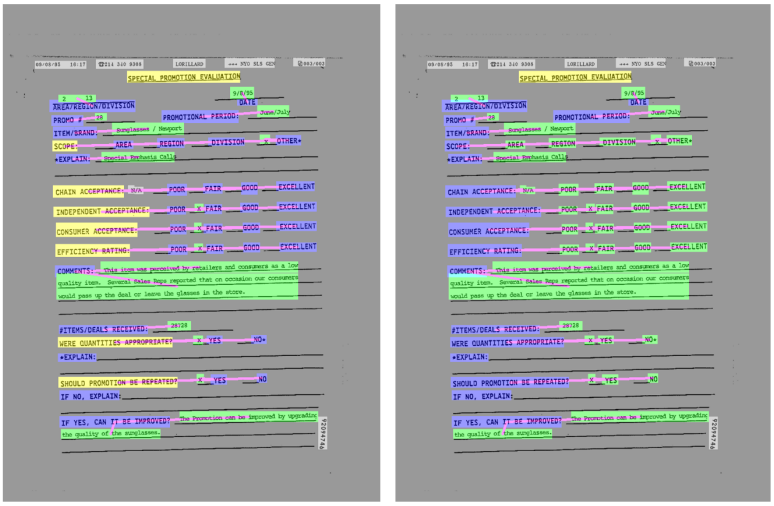
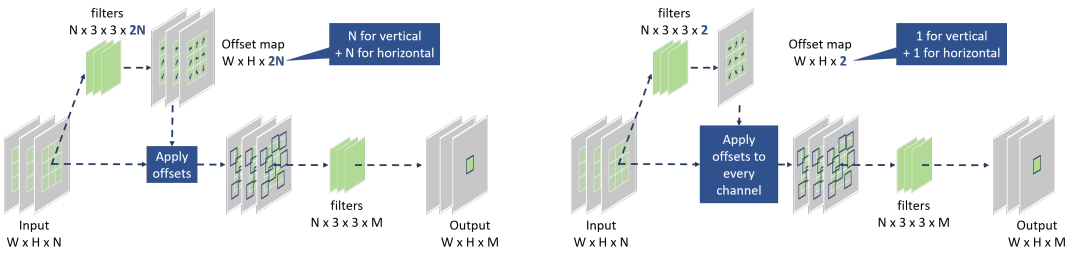


Figure 4: A sample of from the FUNSD dataset (left) and its revised version. ● denotes *headers*, ● denotes *questions*, ● denotes *answers*, and ● denotes the key-value relations.

- A modified version (Fig. 5b) of Deformable Convolution (Fig. 5a) where the convolution offsets stay the same for all channels.
  - Learns only one set of offsets for all channels, keeping the channel-wise relation unchanged.
  - Reduces the number of parameters significantly and improves processing time while still achieving comparable results.



(a) Deformable Convolution (b) Channel Invariant Deformable Convolution

Figure 5: Illustration of two 3x3 Deformable Convolutions.

The motivations were:

- Deformable Convolution is computationally expensive.
- Deformable Convolution learns the offset for each channel separately, thus the features in each channel will be sampled differently. This breaks the depth-wise invariant relation between channels but instead enhances the spatial relations within each channel.
- For classification tasks, this characteristic is preferable, as it enables each channel to embed entirely different information.
- But for segmentation tasks, it introduces unnecessary complexity.

Personal exploration

Apart from work-related topics, I tried to find opportunities to explore other areas for the purpose of learning new things and satisfying my curiosity. This also helps me expand my knowledge and be clearer about my interests.

Learning Causal Inference and Improving DECAF

https://github.com/lone17/DECAF/blob/main/main.pdf

Causal Inference was a completely new topic that I discovered by chance. This field, along with Causal Discovery, gives me the thought that it could potentially help us build models with reasoning capability. I spent two weeks teaching myself the basics of Causal Inference and applied what I learnt to try to improve the paper DECAF: Generating Fair Synthetic Data Using Causally-Aware Generative Networks.

DECAF is a method that aims to generate synthetic (and debiased) tabular data using a GAN-based model with an assumed causal DAG. The method consists of two stages:

- First, given a DAG describing the underlying causal relation of the training data, a GAN-based model is trained to generate synthetic data by following the topology of the given DAG.
- During inference, a relaxed DAG is used to generate the data. The relaxed DAG is the original DAG with some edges removed to eliminate the causal relation between the protected attributes and the target attributes, which increases fairness.

Another aspect that DECAF is concerned with is data utility, which is the measure of how similar the synthetic data is to the original (biased) data. Naturally, there is a trade-off between data utility and fairness. My work attempted to improve the data utility of DECAF while still achieving a similar level of fairness. More specifically, it proposed:

- A a new relaxed DAG (with different relaxation logics) to generate the target attribute (An example for the *income* attribute of the Adult dataset is given in Fig. 6).
- Alternating between the DECAF's relaxed DAG and the new relaxed DAG to maximize the information flow to the target attribute while not violating any d-separation constraints.
- Some metrics to measure the trade-off between data utility and fairness.

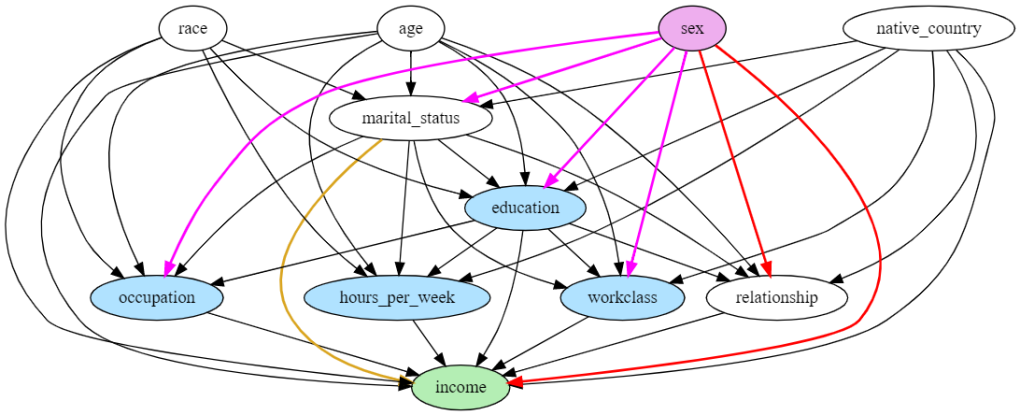


Figure 6: (The same example in Figure 6 of the DECAF paper) DAG for generating *income* in the Adult dataset. The target variable is in green, the protected attribute in pink, and the allowed Conditional Fairness variables in blue. *Demographic Parity* is achieved by removing: 🔴🟡🟣; *Conditional Fairness* is achieved by removing: 🔴🟡.

Inferring Properties of Graph Neural Network and Defending Against Backdoor Attacks

A work of a friend that I contributed to.
https://arxiv.org/abs/2401.03790

My friend, who is a PhD Candidate at The University of Melbourne, needed a hand with his research, so I joined to help him and also to have a chance to gain knowledge on GNNs and backdoor attacks.

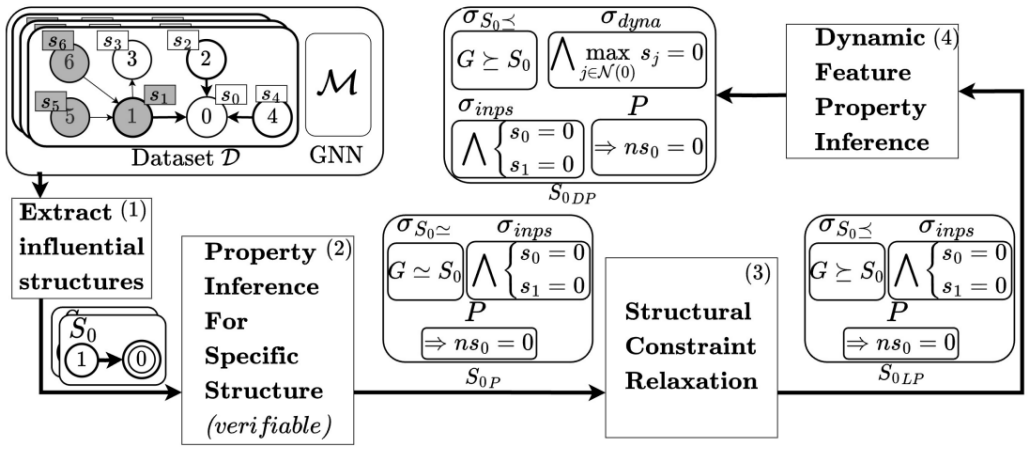


Figure 7: GNN-Infer overview.

The work proposed *GNN-Infer*, a new method for automatic property inference technique of GNNs, and applied it to detecting and defending against backdoor attacks. The method comprises 4 stages (as illustrated in Fig. 7):

- *Extract influential structures*
  - Analyzes the trained GNN and the training data and outputs influential structures.
  - Influential structures are subgraphs that frequently appear in the dataset and significantly impact GNN predictions.
- *Property inference for each specific structure*
  - For each influential structure, turn it into the equivalent feed-forward network (FNN). This was done because there is no known property inference method for GNNs due to its dynamic nature.
  - Infers the properties of the FNN using existing techniques.
  - The properties are rules that describe the future state of the target node based on the current state of connected nodes.
- *Structure constraints relaxation*
  - Generalizes each structure-specific property to a *subgraph isomorphic* (i.e. graphs that contain the same subgraph) property, covering a broader set of input graphs.
  - These properties are more general and useful, but they are just *likely properties* as they cover dynamic structures.
- *Dynamic Property Inference.*
  - Augments the *likely properties* with dynamic feature properties.

  - Those are properties that consider the aggregated states of nodes in the full graph that are connected to the target node (as opposed to only considering connected nodes in the subgraph).

For defending against backdoor attacks, the method is applied as follows:

- Given a (poisoned) GNN and its training data, it infers the properties.
- Decides which properties are likely to be triggers based on their noticeability, stealthiness and effectiveness.
- Then prunes those properties from the GNN.
- The defence mechanism is effective if it reduces the Attack Success Rate of the attack while maintaining the Accuracy on the clean test set.

I contributed to the design and engineering of experiments for backdoor defence.