Paper Critique - 3

Intro of Data Mining - Fall, 2021

Student

Aissata Diallo (Undergraduate) & Sheng Kai Liao (Graduate)

Date:10/28/2021

Paper:

Independent Component Analysis For Trustworthy Cyberspace During High Impact Events: An Application To Covid-19

Critique: (Eddy)

Nowaday, social media is full of malicious news which attempts to misguide the trend of truth from what really happened in the world. In response to such problems, machine learning has made a significant contribution to detecting fake news, misinformation, and suspicious sources. Based on the fact, according to Boukouvalas et. al, the machine learning models before this paper were basically using the handcrafted features of the data which usually offers incomplete information of all scenarios or using deep learning models whose results can not be easily interpreted. Thus, the authors purpose a data-driven solution employing an independent component analysis (ICA) model which is able to find out the misinformation and detect it at the same time while the training.

For the dataset, they created a new labeled Twitter COVID-19 dataset based on randomly collected tweets from Canada with Conditional Independence Coupling (CIC) method which is able to help authors collect the balance data between gender, race and ages. Furthermore, authors down sampled the data into few amounts and manually determined which data is reliable or unreliable according to socio-linguistic standards. Regarding the experiment, the authors compared ICA with different classifications and evaluated their performance with Accuracy, Sensitivity, Precision and F-1. The authors considered the most general deep learning models as the baselines, such as Long Short Term Memory (LSTM), Bidirectional Long Short Term Memory (BiLSTM) and Bidirectional Encoder Representations from Transformers (BERT), with 70:30 training and testing dataset ratio. For ICA itself, they implemented it with three different SVM kernel models and choosed the best one as the major classification model. In addition, to find out how sparsity affects the classification performance in ICA, they modify the original ICA methods called Sparsel-ICA which could directly consider the influence of statistical independence and sparsity in data and test its performance over the number of sparsity.

In observation, by comparing the result of SVM models with different kernels, the best option for ICA is the SVM model which employs the Gaussian kernel. Also, even though the best performance of model is BERT, the ICA still obtains competitive results with Accuracy, Precision and F-1score. For time consuming-wise, the BERT requires more than 6 times the time cost compared to ICA with SVM/Guassian (BERT are trained with 2 min 6 sec and ICA with SVM/Guassian is trained with 18.56 sec) which illustrates that the remarkable computation efficacy of authors' model. Furthermore, according to the experiment, while the number of valuables increased, the accuracy of SparseICA-EBM with SVM/Gaussian dropped rapidly which they believed that considering the sparsity features did not benefit the classification model. Finally, by considering the weight of each entity in the model, they are able to observe which entities could mostly affect their model's decision which improves the model's explainability.

To conclude, the author proposes a data-driven solution that relies on the ICA computational method to achieve misinformation detecting and discovering simultaneously. Their model has comparative results compared to the most general deep learning models yet has a surprise advantage on time consuming aspect. Also, the model has great explainability due to its simplicity. In my opinion, the model could be helpful as an early research tool of misinformation detection in specific topics.

Paper:

Label-Assisted Memory Autoencoder for Unsupervised Out-of-Distribution Detection

Critique: (Eddy)

In the machine learning domain, Out-of-distribution (OoD) data has not often been considered while developing the neural network model which may affect the robustness of the model in real-world application. For this paper, Zhang et. al propose a novel unsupervised OoD detector model (LAMAE) which has outstanding performance over the state-of-the-art relevant models and not required to be trained with OoD samples. Furthermore, the authors discuss some issues which all OoD detection relevant models suffer from and propose several methodologies to tackle those kinds of issues and integrate them with their own model as an upgrade version (LAMAE+).

Before this paper, mostly OoD detector models required tons of OoD data for training. Yet, the OoD data is not easy to access in the real world. To solve this issue, Generative models such as AutoEncoder (AE) are frequently used in OoD detectors to leverage the reconstruction error between the ID data and unseen OoD data by only training with the ID data. There are several ways to implement generative models into OoD detectors. In this paper, the authors mainly refer to MemAE to build their model since the MemAE added a memory component into the traditional AE in order to force the input's representations stored in memory as similar as the ID sample's representations. Thus, their model ,the Label-Assisted Memory AutoEncoder (LAMAE), contains four parts. An encoder to compress the input data and extract its features. A classifier (CLF) for feature regularization. A label-assisted memory (LA-M) to analyze the features and distribute the calculated weight in different chunks of memory. And finally, A decoder for reconstructing the image based on the information in memory.

In addition, the reconstruction error is highly sensitive with the complexity of images which degrades the performance of OoD detection directly. The author uses an entropy-based metric, called Complexity Normalizer (CN), to adjust the reconstruction error and integrate it into their own model as LAMAE+.

In this research, they compared their two OoD detector models to the state-of-the-art models, including TraditionalAE,VAE, SSVAE, MemAE and the latest non-reconstruction based detector GODIN over 10 handwritten digits image datasets. To evaluate the performance, this paper chose to use the AUROC score since it considered the true positive rate of ID data against the false positive rate of OoD data. During the experiment, they pick one digit class, such as "7" , as OoD target and the rest of the nine digits class, "0" to "9" but except "7" ,as ID data. According the result, the LAMAE+ achieve the best performance over all cases. Especially in the case that OoD target is "1", the authors observe that all the state-of-art-model are suffering low performance of detecting the OoD in this case but their own model tackles the issue with 80.6% increase compared to lowest score. Furthermore, this paper also evaluate the performance regarding semantic OoD and Non-semantic OoD data, the LAMAE+ get the top score in 5 over 8 datasets which is pretty impressive.

The paper also analyzes the effect of the Label-Assisted Memory and the CN-Adjustment in their model. By observing the performances of models, such as AE, MemAE and LAMAE, implemented with LA-M, the authors find out all models are benefitted yet struggled in detecting the OoD digits “1”, “4”, “7” and “9” cases. In addition, they compared the performance of between w/o using CN-adjustment methods into AE, MemAE and LAMAE models and observed that the performances of all the case over models are improved and especially the OoD digits “1”, “4”, “7” and “9” cases are improved significantly.

In conclusion, this paper offers two novel OoD detector models which not only have outstanding performance over the state-of-the-art models and tackle the issue that many OoD models are suffering but also are not required to train with OoD data. For the future, I believe this model could be considered as an important tool for testing the robustness of machine learning performance for comprehensive evaluation.

Paper:

DeFraudNet: An End-to-End Weak Supervision Framework to Detect Fraud in Online Food Delivery

Critique (Aissata Diallo)

It is hard to detect fraudulent and abusive claims in online food delivery due to the inability to perform "reverse-logistics" like that which is done for e-commerce platforms. This makes it difficult to harvest labels for fraud because the legitimacy of an item can't be confirmed through inspection. Also, manually analyzing transactions to generate labels for items can be expensive and tedious. According to Mathew et al. there is not enough reliable information out there about what constitutes fraud in online food delivery, particularly in the form of customer service interactions. This is why Mathew et al. have taken it upon themselves to present a novel "end-to-end"framework for detecting fraudulent transactions that are based on large-scale label generations using weak supervision. The authors aim to create strong prevention systems that can identify and alert other systems and humans of fraudulent activities pertaining to online food delivery.

Mathew et al. do their fraud detection of online food delivery through the use of Stanford's AI Lab's (SAIL) Snorkel and tree based methods. Through which they are able to generate weak labels through implementing a manual and automated discovery of labeling functions. Next, they follow this up with a "auto-encoder reconstruction-error" based method, which functions as a noise reduction label. Lastly, the researchers use a model called discriminator model, which is a combination of an MLP and an LSTM. Moreover, they rely on customer history such as transactions related to cross sectional and longitudinal features. They use information provided by a graph known as Convolution Network (GCN), which provides details on customer-customer relationship to capture deceptive behavior.

In the experiment the authors work upon SAIL's Snorkel data programming approach. Snorkel is useful to their research because it uses an ensemble of weak "heuristics" and rules based on domain knowledge to create a generative algorithm, which can be applied to an unlabelled dataset. As an extension of this work, the authors adopt a tree based modeling technique where they can automate the end-to-end process of weak label generation. This automation is useful because it removes bias from the process of creating LFs. They further use the class-specific autoencoder method to reduce noise from the weak labels. I believe "noise" in this research is the labels that are susceptible to fraud even though the authors do not actually define what the term means in their research. They finally use a fraud detection pipeline which involves four main steps. The first step has to do with the data and feature pipeline, which is responsible for building the features that go into the training, validation and "golden" datasets. The second step is the label generation pipeline, which contains the components to generate weak labels for all data points involved in the training and validation dataset. The third step is the discriminator pipeline, this trains the final discriminator models on features and labels from the previous stages. The final step is the evaluation step four, which assists in the ongoing evaluation of the fraud detection system. It achieves this by sending a sample of claims to human evaluators and using their judgment to compute precision, recall and related health-check metrics.

Overall, Mathew et al. demonstrate how the use of a pipeline consisting of handcrafted and auto-generated LFs followed by class-specific denoising autoencoders can be effective in building productive supervised models especially when there are limitations to strongly labeled data. Through their experiment they show how each step of the pipeline improves the evaluation metrics and propose an effective method for fraud detection in online food delivery systems. Their achievement through this research is useful for future fraud detection scenarios because their experiment can be used towards any cases where label data is sparse. The only critique I have for this paper is that the authors rely their research on the history of customer transactions and depend on customer-customer relationships to capture collusive behavior, which can not only be biased, but also limit the final predictions for fraud detection.