Natural Language Processing Approaches for Predicting Discrete Event Time Series

Son Do (sqdo) 2021, Advisor: Olga Fink (ETH Zuirch Advisor) and Danqi Chen (Princeton Advisor)

Abstract

This project will revolve around creating a framework that will encode discrete events and their temporal analysis in a way to power predictions of other events in predicative maintenance. Given a sequence of temporal event log codes we hope to be able to create an embedding for these log codes that can capture the temporal structure and relationship of events with the hopes that other models can use them for recurrent predictions. This is analogous to the NLP task of creating embedding for words.

1 The Motivation

The condition of complex engineered systems is tightly monitored with numerous and diverse condition monitoring devices. Many systems are equipped with monitoring devices that record logs of events detected based on a given set of rules. The events can range from simple operational process confirmations to system status changes if a set of predefined conditions occurs. This results in time series of events that is on the one hand linked to a sequence of operational changes and on the other hand linked to a sequence of changes in system states. The events and their occurrences have certain interdependencies and influence the occurrence of operational disturbances and the system changes can require maintenance interventions. The analysis of the time series of events and their interdependencies can become very difficult if the event representation does not contain any information on the relationships between the different events. Similar challenges have been encountered in Natural Language Processing (NLP) where words are characterized by relationships and interdependencies. Encoding implementations that are not considering these interdependencies are typically underperforming.

2 The Goal

The goal of the project is to develop a framework that combines encoding of the discrete events and their temporal analysis. The framework should provide an informative and interpretable encoding of the event relationships. This can be performed in an analogy to Word2Vec encodings for NLP tasks by predicting the occurrence of the events in a temporal neighborhood and using the latent representation of the neural networks as an informative encoding for the subsequent tasks. Subsequently, the framework should be able to take these representations as input for the prediction tasks demonstrating that these encodings improve the temporal analysis of the time series and the predictive power of the recurrent predictions.

3 Related Research

There has been an abundant amount of research done on creating word embeddings, the most notably being Word2Vec (5) and more recently ELMo(4) and BERT embedding (1). These embeddings have all been able to contextualize words and create a way to store word dependencies and relationships. The project's main goal is to apply this research into the realm of predicative maintenance by creating an embedding of log codes that stores information on the relationship between the different codes. There has been little research done in applying these strategies to predicative maintenance. This project will apply the NLP findings and papers about word embeddings to the realm of predicative maintenance. With the different NLP strategies, my goal is to be able to create a framework that can encode events in a way that would be more informative to recurrent predictions in maintenance.

4 Approach

My plan to tackle this project will be split into 2 main stages. The first is to create a synthetic dataset using a deterministic approach and then one with a probabilistic approach. Once the datasets are made I will then try different embeddings schemes such as (but not limited too):

- A simple Sub-Word Embeddings
- Word to Vector algorithms (CBOW and Skip-grams)
- Global Vectors (GloVe) (3)
- And if there is time and computational power, a deep learning approach: ElMo and/or BERT.

I am still waiting for a dataset such that I can apply these embeddings on. I am in the talks with my professor at ETH Zurich and we discussing different options. We are currently looking at a Appliances energy prediction Data Set (2) and using the that data set to see if we can improve energy usage predictions with the embeddings made.

5 Evaluation

If these embeddings can capture the structure of the synthetic dataset that I created, then these same embeddings can be used on other machine fault datasets to capture their relations as well. The main evaluation will thus be very similar to the ways word embeddings are evaluated through a combination of extrinsic and intrinsic evaluation. I will plug these embeddings into a real predicative maintenance tasks and see whether this improves their performance. In addition, I will use my embeddings to recreate my datasets that I used for training and evaluate its ability to capture the rules that I used to create the datasets.

6 References

References

[1] Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[2] Luis M. Candanedo, Veronique Feldheim, Dominique Deramaix, Data driven prediction models of energy use of appliances in a low-energy house, Energy and Buildings, Volume 140, 1 April 2017, Pages 81-97, ISSN 0378-7788, https://www.sciencedirect.com/science/article/pii/S0378778816308970?via%3Dihub.

[3] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. ”Glove: Global vectors for word representation.” Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.

[4] Peter, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

[5] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean, Distributed Representations of Words and Phrases and their Compositionality, CoRR 1310.4546, 2013 http://arxiv.org/abs/1310.4546,