DEIM Forum 2019 P1-142

Towards Future Event Prediction using Graph-LSTM

Peng CHEN†, Adam JATOWT†, and Masatoshi YOSHIKAWA†

† Department of Social Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
E-mail: chenpeng.acmer@yahoo.com, adam@dl.kuis.kyoto-u.ac.jp, yoshikawa@i.kyoto-u.ac.jp

Abstract Predicting what happens next between countries of interest is meaningful for social scientists and policy makers. By using auto-coded events database GDELT (Global Data on Events, Location, and Tone), which records what happened between all the countries in the past few decades, we can predict events in a future time period. However, in the predicting process, we should not ignore the interaction of events between other countries. Therefore, we first build a graph composed of several closely related countries, and use a Graph-LSTM to predict future events between a specific pair of countries on a graph composed of several countries, also taking into account events happened between them and their neighbors.

Key words event prediction, GDELT, graph structure, LSTM

1 Introduction

Predicting what happens next between two countries of interest is meaningful for social scientists, journalists and policy makers, which helps us understand trends in international relations. In recent years, many attempts have been made to predict future events [1,2,3]. In the prediction tasks, one of the most challenging steps is to collect data and build an applicable dataset. Fortunately, with the emergence of more and more open-source global event databases, we are able to access records of events happening around the world. It is worth mentioning that the GDELT (Global Data on Events, Location, and Tone) dataset, one of the largest global event dataset, is freely available and records a tremendous amount of events from a variety of international news sources with daily updates

In the past few years, Recurrent Neural Networks (RNNs) have proved their superior ability to perform predictions with sequential data on a number of sequence-based learning problems[4]. Besides, Long Short-Term Memory (LSTMs), as a special kind of RNN, are capable of learning long-term dependencies and are now widely used because of their outstanding performance on a large variety of problems[.

In this paper, for data, we exploit the GDELT dataset to build our data collection that is a subset of the GDELT, focusing on events happened between several closely related countries. For prediction, we develop a Graph-LSTM based predictive framework to predict future events between a pair of countries based on events which happened, and also leveraging history event records of several countries that are closely related to them. Our proposal is more advanced compared to vanilla RNN/LSTM that only learns knowledge from event records between a specific pair of countries.

2 Related Work

A number of existing works have researched how to make use of the GDELT dataset. Some traditional machine-learning methods have been used. For example, Phua et al. developed decision trees to predict the Singapore stock market's Straits Time Index using the GDELT dataset [6], Qiao et al. used the GDELT for predicting social unrest events across five major nations in Southeast Asia with Hidden Markov Model (HMM) [2], and Yonamine utilized a statistical method, named Autoregressive Fractionally Integrated Moving Average (ARFIMA) model, to predict violence levels in Afghanistan. More recently, Smith et al. explored the use of Neural Networks (RNNs and LSTMs) for predicting the number of conflict events in Afghanistan, compared with ARFIMA model[3]. On the other hand, many previous works have used other data sets for event prediction [6,7,8].

3 Long Short-Term Memory Networks

Recurrent Neural Networks (RNNs) are capable of processing sequential data with variable length by applying a single transition function on hidden states recursively.

LSTMs are explicitly designed to avoid the long-term dependency problem that exists in standard RNNs [5]. For the structure of LSTMs, there are five components in a LSTM unit: an input gate i_t, a forget gate f_t, an output gate o_t, a memory cell c_t, and a hidden state h_t. They are all vectos in $\mathbb{R}^k$, where k is the dimension of hidden state.

Each step in LSTM's recursive process can be defined as a collection of transition functions as follows:

$$i_t = \sigma(W^t x_t + U^i h_{t-1} + b^i) \tag{1}$$
$$f_t = \sigma(W^f x_t + U^f h_{t-1} + b^f) \tag{2}$$
$$o_t = \sigma(W^o x_t + U^o h_{t-1} + b^o) \tag{3}$$
$$u_t = \tanh(W^u x_t + U^u h_{t-1} + b^u) \tag{4}$$
$$c_t = i_t \odot u_t + f_t \odot c_{t-1} \tag{5}$$
$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

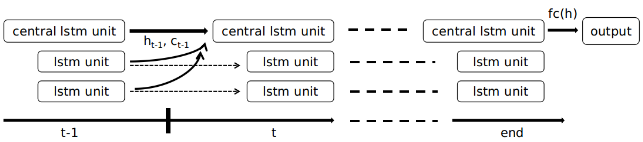
where x_t is the current input at time step t, σ refers to the sigmoid function, $\odot$ denotes elementwise multiplicationm, and tanh represents hyperbolic tangent. As can be seen, when a input vector x_t comes in, the input gate decides which values in the input will be updated, then the forget gate controls how much previous information from c_{t-1} is forgotten, and the output gate filters the memory state c_t, creating the next hidden state h_t.

4 Graph-LSTM

Intuitively, one of the limitations of the conventional LSTM architectures is they can only handle single sequential input. Recently, several structural variants of LSTMs have been proposed. Tai et al. introduced the tree-structured LSTM, in which a LSTM unit may have several precedents [10]. Liang et al. first proposed graph LSTM for semantic object parsing in the field of image processing, which extends the traditional LSTMs from sequential data to general graph-structured data [12]. Moreover, Peng et al. extended the use of graph LSTM to cross-sentence relation extraction [11].

In this paper, we use a designed graph LSTM for future event prediction using multiple sequence data. In the graph LSTM we designed, there are two kinds of units, one is traditional LSTM unit, the other is called central LSTM unit. A central LSTM unit has multiple predecessors pre_t, which means each central unit has multiple inputs from the previous time step. There are several traditional LSTM units and a central LSTM unit in one layer of the graph LSTM.

The structure of the graph LSTM is shown below:



The following equations describe the transition process in a central unit:

$$i_t = \sigma(W^t x_t + \sum_{k=1}^{num(pre_t)} (U_k^i h_{k,t-1} + b_k^i)) \tag{7}$$
$$f_t = \sigma(W^f x_t + \sum_{k=1}^{num(pre_t)} (U_k^f h_{k,t-1} + b_l^f)) \tag{8}$$
$$o_t = \sigma(W^o x_t + \sum_{k=1}^{num(pre_t)} (U_k^o h_{k,t-1} + b_k^o)) \tag{9}$$
$$u_t = \tanh(W^u x_t + \sum_{k=1}^{num(pre_t)} (U_k^u h_{k,t-1} + b_k^u)) \tag{10}$$
$$c_t = i_t \odot u_t + \sum_{k=1}^{num(pre_t)} f_t \odot c_{k,t-1} \tag{11}$$
$$h_t = o_t \odot \tanh(c_t) \tag{12}$$

where a central LSTM unit takes current input x_t and multiple hidden states from predecessors as input. This kind of unit is capable of learning knowledge from the hidden states from traditional LSTM units, that is to say, the central LSTM can not only learn from its own sequential input but also several other sequential inputs. Besides, plenty of parameters inside the central LSTM allow the networks learning how to make use of multiple sequential data.

5 Data Representation

GDELT provides auto-coded event records all over the world, which contain various valuable details such as event code, actors involved, country code and so on. By using GDELT, we can collect event records related to some entities we are interested in. In this paper, we focus on four major countries: Japan, India, China and USA, and try to predict events that will happened between two countries of them.

Each event record has an attribute called goldstein scale that is a numeric score from -10 to +10 (negative for conflicts and positive for cooperation), capturing the theoretical potential impact that type of event will have on the stability of a country. Another attribute, quad class, aggregates goldstein scale into 4 classes: 1 for verbal cooperation, 2 for material cooperation, 3 for verbal conflict, and 4 for material conflict.

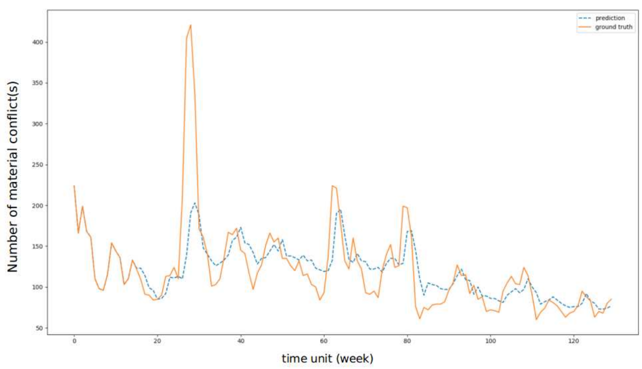
In the experiment, the numbers of occurrences of these four kinds of events per week are counted. Having a sequence of event counts, the model predicts the number of material conflict events (quad class 4) in the next week. Obviously, this is a coarse-grained prediction, which provides future trends in relationship between two countries. Due to time constraints, we leave more suitable data representations and more fine-grained predictions for future work.

6 Experiments

Define all events into 4 classes: 1=Verbal Cooperation, 2=Material Cooperation, 3=Verbal Conflict, 4=Material Conflict. In the preliminary experiment, we focus on the country pair: (USA, Japan), and we count the number of four kinds of events that happened between USA and Japan in each week. For prediction, the number of material conflict events in the 16th week is predicted by the model based on the previous 15 weeks' data.

Two kinds of models are evaluated, vanilla LSTM and the graph LSTM. For vanilla LSTM, we simply use the data of the country pair (USA, Japan) as input that includes events happened between these two countries. For graph LSTM, the data of (USA, Japan) is fed into the central LSTM while the data of other closely related country pairs, (Japan, China), (Japan, India), (USA, China) and (USA, India) are taken as inputs to other parallel LSTMs in the graph LSTM model.

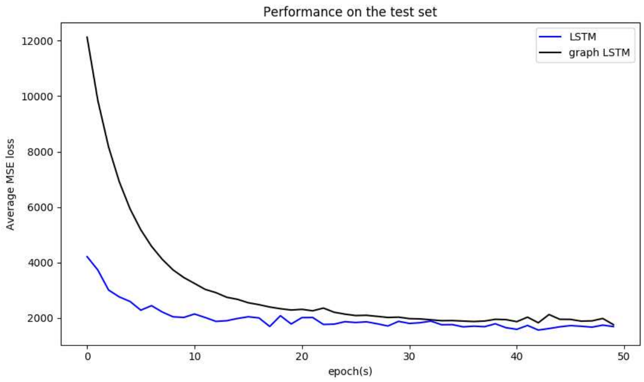
The following figure shows a test result from LSTM:



where the dotted blue line shows the predicted number of material conflicts, and the orange line represents the ground truth.

Training data consists of events that happened from 2005/01/01 to 2016/06/01 (595 weeks), and test set includes data from 2016/06/01 to 2018/12/01 (130 weeks). The size ratio of training set and test set is about 8:2. In evaluation, we use average Mean Squared Error (MSE) to evaluate the loss between prediction results and the ground truth. Traditional LSTM and Graph-LSTM are both used to perform prediction for comparison.

The losses after convergence become 1684.9 for LSTM, and 1718.51 for graph LSTM. The convergence process comparison is shown in the following figure:



As is shown in the figure, the performance of graph LSTM is no better than traditional LSTM, and graph LSTM took more time to converge. However, trying to consider the interaction of events between related countries can be reasonable and helpful. Because in most major events such as material conflict events, there are multiple countries involved, not only two countries. Even some particularly significant events occur in a global context.

7 Future Work

There are still many shortcomings in our model. At present, we manually select closely related countries which are sometimes inappropriate. Ideally, we are supposed to consider as many counties as possible. On the other hand, we need to be more careful when combining hidden states from several LSTMs. Currently, we simply add them together which ignore the weight of each state. Obviously, a more beautiful and reasonable combination need to be designed. For future work, we will build a relational graph of countries, trying to select more suitable neighbors of two countries. For hidden state combination, the attention mechanism can be a potential solution, which transforms the simple summation into an adaptive weighted summation.

8 Acknowledgments

This research has been supported by MIC SCOPE (#171507010).

References

[1] Brandt, Patrick T., John R. Freeman, and Philip A. Schrodt. "Real time, time series forecasting of inter-and intra-state political conflict." Conflict Management and Peace Science 28.1 (2011): 41-64.

[2] Qiao, Fengcai, et al. "Predicting social unrest events with hidden Markov models using GDELT." Discrete Dynamics in Nature and Society 2017 (2017).

[3] Smith, Emmanuel M., et al. "Predicting the occurrence of world news events using recurrent neural networks and auto-regressive moving average models." UK Workshop on Computational Intelligence. Springer, Cham, 2017.

[4] Karpathy, Andrej. "The unreasonable effectiveness of recurrent neural networks." Andrej Karpathy blog (2015). URL

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

[5] Olah, Christopher. "Understanding LSTM networks, Aug 2015." URL http://colah. github. io/posts/2015-08-Understanding-LSTMs.

[6] Weidmann, Nils B., and Michael D. Ward. "Predicting conflict in space and time." Journal of Conflict Resolution 54.6 (2010): 883-901.

[7] Hammond, Jesse, and Nils B. Weidmann. "Using machine-coded event data for the micro-level study of political violence." Research & Politics 1.2 (2014): 2053168014539924.

[8] Perry, Chris. "Machine learning and conflict prediction: a use case." (2013).

[9] Yonamine, James E. "Predicting future levels of violence in afghanistan districts using gdelt." Unpublished Manuscript (2013).

[10] Tai, Kai Sheng, Richard Socher, and Christopher D. Manning. "Improved semantic representations from tree-structured long short-term memory networks." arXiv preprint arXiv:1503.00075 (2015).

[11] Peng, Nanyun, et al. "Cross-sentence n-ary relation extraction with graph lstms." arXiv preprint arXiv:1708.03743 (2017).

[12] Liang, Xiaodan, et al. "Semantic object parsing with graph lstm." European Conference on Computer Vision. Springer, Cham, 2016.