Tensor-based Method for Temporal Geopolitical Event Forecasting

Mehrnoosh Mirtaheri [1] **Sami Abu-El-Haija** [1] **Tozammel Hossain** [1] **Fred Morstatter** [1] **Aram Galstyan** [1]

Abstract

Many political and societal factors can affect the course of geopolitical events, making them extremely challenging to forecast.In this work, we seek to develop a computational approach for forecasting geopolitical events. We show that one can leverage past geopolitical interactions to predict future related geopolitical interactions that have not necessarily occurred before. We propose to leverage ICEWS, a geopolitical interaction event dataset, to develop a predictive modeling framework for geopolitical events using both supervised and unsupervised machine learning approaches.

1. Introduction

Many phenomena can be represented as dyadic relationships between actors in a dynamic network, varying from link prediction in social networks to relationships between countries on a global stage. Because of the wide applicability of this representation, developing approaches that can predict how these networks evolve is an important problem to focus on.

One domain where this problem is of utmost significance would be in geopolitical forecasting, where actors are countries (e.g. USA, China, etc) and links represent the interactions between them (e.g. signing treaties, declaring war, etc). Within this context, often the most interesting relationships to predict are those that happen with the least frequency. For example, the meeting between North Korea's leader and a member of the G7 was unprecedented prior to June, 2018. However, in the months leading up to this event, there were tertiary events indicating a meeting might happen in the near future.[1] Predicting these kinds of events would be of enormous value to policy makers, governments, and journalists, among others.

Fortunately, during the last couple of years geopolitical events have been coded in datasets such as GDELT[2] and Integrated Crisis Early Warning System (ICEWS[3]). The advent of these large geopolitical event datasets, automatically extracted and coded from internet news archives, gives us the chance of monitoring these events over time. In both datasets, CAMEO coding scheme is used to represent the events. The benefit of the CAMEO coding scheme is that it produces a series of dyadic events where each event falls under a "CAMEO code", mapping the event into one of a several predefined geopolitical interactions (Gerner et al., 2002), including 20 high level CAMEO actions, e.g., the code '19' corresponds to 'Fight'. Following from this, a CAMEO-coded dyadic event consists of four pieces of information: a sender, a receiver, an action type, and a timestamp. We can map a geopolitical question into a CAMEO coded event. Here is an example of those questions:

Will Kim Jong-un meet the head of government from the United States in-person before 15 June 2018?

Which can be extracted to a (North Korea, Meet, US, 06-15-2018), where "Meet" is associated with a CAMEO code in ICEWS dataset.

In this work, we present a Tensor Forecasting model that combines a Bayesian Poisson tensor factorization and a convolutional autoregression model to leverage the evolving and multi-relational nature of a Temporal Knowledge Graph for temporal reasoning. We test the performance of the model over ICEWS dataset and compare our method with the other state-of-the art statics and dynamics approaches in sequence modeling and graph representation learning. Our approach outperforms all the baselines in most of the cases in forecasting geopolitical events.

2. Related Work

The literature of this work can be divided into three complementary threads: (i) Traditional machine learning (ii) Tensor factorization (iii) Knowledge Graph embeddings.
The first category of approaches (Keneshloo et al., 2014; Korkmaz et al., 2015; Parrish et al., 2018) are mostly supervised, designed to specifically target the prediction of geopo-

[1]USC Information Sciences Institute, Marina Del Rey, California, USA. Correspondence to: Mehrnoosh Mirtaheri <mehrnoom@usc.edu>.

Presented at the ICML 2019 Workshop on Learning and Reasoning with Graph-Structured Data

[1]https://www.armscontrol.org/factsheets/dprkchron

[2]https://www.gdeltproject.org/

[3]https://dataverse.harvard.edu/dataverse/icews

litical events by extracting features from different datasets such as GDELT, ICEWS, social media, etc. Some among them use ground truth data manually curated by experts, such as GSR provided by (Ramakrishnan et al., 2014) and ICEWS Event of Interests ground truth data (Lustick et al., 2015), which limits the task to specific countries or events. The methods applied to this problems include: discriminant analysis, HMM (Qiao et al., 2017), and Bayesian time series forecasting (Montgomery et al., 2012). (Keneshloo et al., 2014) is the first that considers the graph nature of the problem. They detect and predict Domestic Political Crisis by mining frequent sub-graphs of the interaction graph for both negative and positive class.

The second and third category are more general and applicable on different relational dataset. (Schein et al., 2015) and (Schein et al., 2016) specifically model ICEWS dataset a 4-D tensor, of size $N \times N \times E \times T$ where N is the number of country actors and E is the number of action types and T is the time period. They propose Bayesian Poisson Tensor Factorization and Tucker Decomposition to extract the underlying multilateral relationship between countries and evaluate their methods on inference tasks, while we are interested in forecasting unseen events. (Dunlavy et al., 2011) is one of the first works that combine a tensor factorization and timeseries prediction to forecast the future tensor.

There are many relational representation learning algorithms (Socher et al., 2013; Bordes et al., 2013; Yang et al., 2014; Schlichtkrull et al., 2018) for reasoning over large static Knowledge Graphs, all of them fail to capture the temporal dynamics of a Temporal Knowledge Graph. In (Sadeghian et al., 2016) Sadeghian et. al extend the rule mining approach presented in (Yang et al., 2014) for temporal Knowledge graphs. (Trivedi et al., 2017) and (Jin et al., 2019) use recurrent neural network to captures the evolving representation of entities and relations over time and evaluate their method on ICEWS and GDELT dataset. (Jin et al., 2019) is used as one of our baselines as it shows promising performance in beating all the other static and dynamic approaches.

3. Problem Definition

Given an ordered sequence of events between countries, within a time window $[t - h + 1, t]$, we aim to predict whether an event will happen in time interval $[t + 1, t + w]$, where h is the history length and w is the lead time window.

In this study, we are interested in forecasting geopolitical events encoded as CAMEO coded interactions. We precisely address the following question:

Given two countries s and o, will a geopolitical event r happen between them in the future?

4. Our Model

We represent the interaction data as a 4-dimensional Tensor M of size $N \times N \times T \times R$, where N is the number of countries, T is the number of time steps, and R is the number of events, where each time step is the aggregation of w days. m_{sotr} corresponds to the number of interactions of type r from country s to o at time step t.

Given Tensor M, we want to extrapolate entries along the third (i.e. time) dimension. Specifically, we want to output a tensor with dimensions $N \times N \times C \times R$, where C is the number future time steps we would like to predict. This output tensor is an estimate of the number of different interactions, between all countries, that will happen in the upcoming S time steps according what our model expects. Our proposed algorithm includes following steps:

1. **Tensor Factorization.** Tensor factorization methods identify the underlying hidden structure of the data Our 4-dimensional M, it can be factorized into four low-rank (k-dimensional) factor matrices $\theta^S \in \mathbb{R}^{N\times k}, \theta^O \in \mathbb{R}^{N\times k}, \theta^T \in \mathbb{R}^{T\times k}$, and $\theta^R \in \mathbb{R}^{R\times k}$, and their outer tensor product should recover M. Specifically, each element of M can be estimated as:

$$\widetilde{m}_{sotr} = \sum_{f=1}^{k} \left(\theta_i^S \circ \theta_j^O \circ \theta_t^T \circ \theta_e^R\right)_f \qquad (1)$$

   where $\circ$ is Hadamard product, the subscripts on θ select row vectors, and the subscript f selects scalars from the row vector. Schein et al. (2015) proposed Bayesian Poisson Tensor Factorization (BPTF) which is a probabilistic approach for identifying the latent structures. BPTF assumes that each element of M is coming from a Poisson distribution (as it is suitable for count data) with mean $\widetilde{m}_{sotr}$. They also impose four sparsity-inducing Gamma priors over the latent factors. Each entry in the factor matrices e.g. θ_{sk}^S:

$$\theta_{sk}^I \sim \text{Gamma}(\alpha, \beta^S),$$

   and similarly for θ^O, θ^T, and θ^R, where all θ's share the same shape parameter $\alpha \in \mathbb{R}_{>0}$ but each has its own rate parameter $\beta \in \mathbb{R}_{>0}$

2. **Forecasting**. We extrapolate θ^T producing θ^C through a simple autoregressive convolutional model. In particular, we train convolutional filter $W \in \mathbb{R}^{\tau\times k\times k}$, Filter height τ allows us to process τ timesteps in the past for predicting a single timestep. The convolution $\theta^T * W$ produces response $\in \mathbb{R}^{(T-\tau+1)\times k}$. Finally, we train W to minimize:

$$\min_W ||\theta^T_{\tau:} - \theta^T * W||_F^2, \qquad (2)$$

where $||.||_F$ is the Frobenius norm. Other autoregressive alternatives such as RNN are plausible, but we were able to get slightly better performance using the simple convolutional architecture. Finally, we use W to append rows to θ^T, one row at a time. We refer to these new (extrapolated) rows as θ^S.

3. **Reconstruction**. Given $\theta^S, \theta^O, \theta^C$ and θ^R, we predict the future tensor using PARAFAC method i.e. using Equation 1.

Convert Regression to Probability The output of this method $\hat{m}_{sotr}$, is the expected value for the number of events of type r between s and o. In (Schein et al., 2015) it is assumed that $\hat{m}_{sotr}$ is the mean of a Poisson distribution. The probability of an event happening can then be calculated by:

$$P(X > 0) = 1 - P(X = 0) = 1 - e^{-\lambda} = 1 - e^{-\hat{m}_{sotr}}$$

5. Baselines

We design the problem, in various ways, to address different aspects of our dataset, including binary classification task (Section 5.1), static (Section 5.2) and temporal (Section 5.3) link prediction. This section explains the baseline methods and the task designs.

5.1. LSTM

We model the proposed problem as a binary classification task: given a triple (s, r, o), and a timestamp t, we aim to to predict the event as positive if the action r happens between s and o in $[t, t + w]$.

Labels. For every pair (s, o) in our dataset, we choose k random non overlapping timestamps from $[t_{start}, t_{end}]$ where t_{start} and t_{end} are the starting and ending date of the ICEWS dataset. An instance (s, o, r) has positive label if action r occur between s and o within $[\tau, \tau + w]$ otherwise it has negative label.

Features. We denote $x_{so,t} \in \mathbb{R}^{20}$ as the vector of CAMEO interactions between country s and o at timestamp t, and $F_{h,\tau} = \{x_{so,t}$ for $t \in [\tau - h, \tau]\}$.

Model: The proposed model is based on a single layer of LSTMs, a type of recurrent neural network designed to capture long term dependencies in sequential data (Hochreiter & Schmidhuber, 1997). For a given instance (s, o, r, t), vectors in $F_{h,\tau}$ are used as an input sequence for the LSTM model. The output of the last units produces the probability of the event.

5.2. RGCN

We can also represent the country-coutnry interactions as a multi-relational temporal graph in which, nodes represents the countries, and country s and o are connected with an edge with label r, t if an event of type r happens at time step t between them. Relation Graph Convolution Network (R-GCN) proposed by (Schlichtkrull et al., 2018) defines the hidden representation of each node as an aggregation of the representation of its neighbors:

$$h_i^{(l+1)} = \sigma(W_0 h_i^{(l)} + \sum_{r \in R} \sum_{j \in \mathcal{M}_i} \frac{1}{n_{r,j}} W_r h_j^{(l)})$$

where $\mathcal{M}_i$ denotes the set of neighbor indices of node i under relation $r \in R$ and $n_{i,r}$ is a normalization constant. For link prediction, DistMult (Yang et al., 2014) factorization is used, in which every relation r is a diagonal matrix $R \in \mathbf{R^{d \times d}}$. Given e_s and e_o as the latent factor representations of s and o, the scoring function is defined as:

$$f(s, r, o) = e_s^T R_r e_o$$

5.3. RE-Net

RE-Net, proposed by (Jin et al., 2019) comprised of an event sequence encoder and a neighborhood aggregation module. Given a sequence of events $\{(s_i, r_i, o_i, t_i)\}_i$, the goal is to predict the probability of unseen events $(s, r, ?, t)$ and $(?, r, o, t)$ by predicting the object or subject entity. They define the conditional probability of object o happening at time t as follows:

$$P(o_t | s, r, O^r_{t-k-1}, \ldots O^r_{t-1} = f(e_s, e_r, h_{t-1}(s, r))$$

Where O^r_t is the history of events of type r between o as an objects and other entities. $h_t(s, r)$ is defined as:

$$h_t(s, r) = RNN(e_s, e_r, g(O^r_s), h_{(t-1)}(s, r))$$

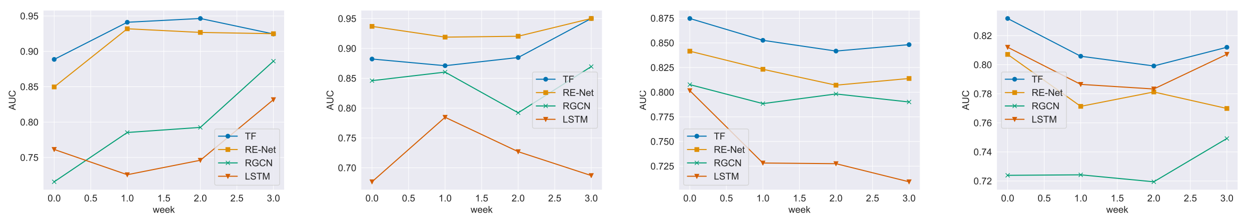
They study different aggregation functions in the paper, including mean, convolution, and attention aggregator. For our experiments, we use mean, as it is shown to give the best results. For a given event (s, r, o, t), this model outputs a score for $(s, r, ?, t)$ and $(?, r, o, t)$. We average them to calculate the score of (s, r, o, t).

6. Experiments

We evaluate the proposed methods against the static and temporal methods. The metric used for evaluating and comparing the models is AUC-ROC. We choose four types of CAMEO codes: Fight, Reduce Relations, Diplomatic Cooperation, and Consult. The first two are among the most rare relations in the ICEWS dataset, while Consult is the most frequent one.

6.1. LSTM Evaluation

Features and labels are selected as explained in Section 5.1. For each pair in the dataset with at least one r interaction,



(a) Fight (b) Reduce Relations (c) Diplomatic Cooperations (d) Consult

Figure 1. Performance comparison on event prediction task. Tensor Forecasting method outperforms other methods except for "Reduce Relation" prediction.

Table 1. Ratio of positive samples in the train and test set for different events, shown by their CAMEO code for brevity (19: Fight, 16: Reduce Relations, 05: Diplomatic Cooperation, 04: Consult)

Event	**Number of Samples**	**Train month**	**Test month**	**Train week**	**Test week**
19	$\approx 18K$	0.083	0.084	0.033	0.033
16	$\approx 15K$	0.054	0.056	0.017	0.018
05	$\approx 37K$	0.209	0.209	0.075	0.076
04	$\approx 29K$	0.416	0.406	0.179	0.175

Table 2. AUC-ROC score reported for LSTM model, labeled chosen by looking at a (a) week (b) month in the future. Features are selected from 90 days back before a timestamp

Model	Fight	Reduce Relations	Diplomatic Cooperation	Consult
Weekly	0.823	0.838	0.795	0.762
Monthly	0.754	0.678	0.768	0.782

10 random non-overlapping timestamps are drawn to make the sample set. For a given random timestamp t, we look at a (i) week and (ii) month ahead to label the instance as positive or negative. For this task we use ICEWS dataset from 01-01-2012 to 07-01-2018. We make a $70\% - 30\%$ train/test split by adding any instance that happened after 07-01-2016 into the test set and all others into the train set. Table 1 shows the ratio of positive samples in the train and test set in both weekly and monthly setting. We evaluate the model performance for $h \in [30, 60, 90, 120]$ days to select our feature set $F_{h,\tau}$. The best performance is achieved for LSTM with $h = 90$ days, shown in Table 2 for the four CAMEO event types.

6.2. Model Comparison

We setup an experiment to compare the performance of our models against each other in the following setting:

Tensor Forecasting. We make a 4-D matrix from ICEWS dataset between 01-01-2012 and 10-01-2017, as explained in Section 4. The Tensor rank convolutional filter size is equal to 37 and 15 respectively.

LSTM. This model is trained on the ICEWS events that occurred in between 01-01-2012 and 10-01-2017. For a given pair starting at week t, the feature vector is extracted from $[t - 90\text{days}, t]$. The model is a single hidden layer lstm, with a drop out rate 0.1 and learning rate 0.001. The hidden layer dimension is 32.

RGCN. We ignore the timestamps on the edges and create a multirelational static graph, by aggregating edges across all the timestamps. The output score $f(s, r, o)$ is proportional to the likelihood of the event happening.

RE-Net. We train the model on country-country sequence interactions, from ICEWS dataset between 10-01-2016 and 10-01-2017. The number of hidden layers used is 100, learning rate 0.001 and drop out rate 0.5.

Ground Truth. We select the next four weeks starting from 10-01-2017 as the test set. For each week, starting at t, a country pair (s, o) is labeled as positive with respect to event type r if at least one r interaction happened between s and o in $[t, t + 7\text{days}]$. An equal number of random pairs are chosen as negatives. We make the task harder by filtering country pairs with no or few interactions. Without this preprocessing there is a high chance that a lot of the pairs have feature vectors with almost zero for the LSTM model which will end up with high (e.g. 99%) AUC score.

Results Figure 1 shows the performance comparison of different models. Our Tensor Forecasting approach beats other methods except in "Reduce Relation" prediction tasks where RE-Net performs better. Our method captures the temporal evolving dependencies between entities, as well as all event types, unlike LSTM that only uses historical events between the two entities, RGCN that ignores the temporal notion of the data. RE-Net also predicts events based on a history and graph structure, but the graph structure of a target triple (s, r, o) only considers events with type r.

7. Conclusion and Future Work

We presented an efficient Tensor Forecasting approach, that trains a Bayesian Poisson Tensor Factorization and a convolutional autoregressive model. We showed that our approach outperforms state-of-the arts static and temporal methods in our application. An interesting future work would be to train an end-to-end model that optimizes the tensor factorization and regression loss together.

Acknowledgments

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2017-17071900005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pp. 2787–2795, 2013.

Dunlavy, D. M., Kolda, T. G., and Acar, E. Temporal link prediction using matrix and tensor factorizations. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(2):10, 2011.

Gerner, D. J., Schrodt, P. A., Yilmaz, O., and Abu-Jabr, R. Conflict and mediation event observations (cameo): A new event data framework for the analysis of foreign policy interactions. *International Studies Association, New Orleans*, 2002.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Jin, W., Zhang, C., Szekely, P., and Ren, X. Recurrent event network for reasoning over temporal knowledge graphs. *arXiv preprint arXiv:1904.05530*, 2019.

Keneshloo, Y., Cadena, J., Korkmaz, G., and Ramakrishnan, N. Detecting and forecasting domestic political crises: a graph-based approach. In *Proceedings of the 2014 ACM conference on Web science*, pp. 192–196. ACM, 2014.

Korkmaz, G., Cadena, J., Kuhlman, C. J., Marathe, A., Vullikanti, A., and Ramakrishnan, N. Combining heterogeneous data sources for civil unrest forecasting. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 258–265. ACM, 2015.

Lustick, I., O'Brien, S., Shellman, S., Siedlecki, T., and Ward, M. ICEWS Events of Interest Ground Truth Data Set, 2015. URL https://doi.org/10.7910/DVN/28119.

Montgomery, J. M., Hollenbach, F. M., and Ward, M. D. Improving predictions using ensemble bayesian model averaging. *Political Analysis*, 20(3):271–291, 2012.

Parrish, N. H., Buczak, A. L., Zook, J. T., Howard, J. P., Ellison, B. J., and Baugher, B. D. Crystal cube: Multidisciplinary approach to disruptive events prediction. In *International Conference on Applied Human Factors and Ergonomics*, pp. 571–581. Springer, 2018.

Qiao, F., Li, P., Zhang, X., Ding, Z., Cheng, J., and Wang, H. Predicting social unrest events with hidden markov models using gdelt. *Discrete Dynamics in Nature and Society*, 2017, 2017.

Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R., Saraf, P., Wang, W., Cadena, J., Vullikanti, A., Korkmaz, G., et al. 'beating the news' with embers: forecasting civil unrest using open source indicators. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1799–1808. ACM, 2014.

Sadeghian, A., Rodriguez, M., Wang, D. Z., and Colas, A. Temporal reasoning over event knowledge graphs. In ., 2016.

Schein, A., Paisley, J., Blei, D. M., and Wallach, H. Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1045–1054. ACM, 2015.

Schein, A., Zhou, M., Blei, D. M., and Wallach, H. Bayesian poisson tucker decomposition for learning the structure of international relations. *arXiv preprint arXiv:1606.01855*, 2016.

Schlichtkrull, M., Kipf, T. N., Bloem, P., Van Den Berg, R., Titov, I., and Welling, M. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pp. 593–607. Springer, 2018.

Socher, R., Chen, D., Manning, C. D., and Ng, A. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*, pp. 926–934, 2013.

Trivedi, R., Dai, H., Wang, Y., and Song, L. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. *arXiv preprint arXiv:1705.05742*, 2017.

Yang, B., Yih, W.-t., He, X., Gao, J., and Deng, L. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.