Forecasting Of The Rare Geopolitical Events Using Knowledge Graphs

Anonymous ACL submission

Abstract

Abstract goes here.

1 Literature Review

The problem of forecasting political events has gained attraction among the social and political science community, especially with the advent of large Geopolitical datasets, automatically extracted and encoded from Internet news archives, that gives us the chance of monitoring these events over time. Being able to predict domestic or international crisis, rebellions, military conflicts, etc has broad interest among decision makers and leaders. The goal of this paper is to exploit two major Geopolitical datasets: Global Database of Events, Language, and Tone (GDELT\(^1\)) and Integrated Crisis Early Warning System (ICEWS\(^2\)), both extracted from news article and use CAMEO\(^3\) coding schemes to represent the events. A CAMEO-coded dyadic event consists of four pieces of information: a sender, a receiver, an action type, and a time-stamp.

Traditional machine learning techniques (Keneshloo et al., 2014; Korkmaz et al., 2015; Parrish et al., 2018; ?) have been applied to address some aspects of the problem. The methods applied to this problems include discriminant analysis, HMM (Qiao et al., 2017), Bayesian time series forecasting (Montgomery et al., 2012; Shellman, 2004) and vector auto regression methods (VAR) (Goldstein, 1992). We will cover a few of them in this literature but the main focus shall be on the more recent works that cover more general aspects of the problem.

The work in (Keneshloo et al., 2014) to the best of our knowledge is the first that considers the graph nature of the problem. Authors find the frequent sub-graphs of the interaction graph of both negative and positive sample. They then choose a discriminate set of them as features for detecting Domestic Political Crisis (DPC). They use Lasso-regression over event counts and graph properties to forecast the probability of DPC. Authors in (Korkmaz et al., 2015) exploit somewhat similar event features such as daily counts of events, average intensity of the events in ICEWS, average tone of the daily events in GDELT, and Goldstein scale score of daily events in GDELT (a collaboration score assigned to each event; the higher the score between the two actors, the greater their collaboration) as a part of the input for their Lasso regression classifier along with the other data extracted from social media and blog posts to predict the probability of occurrence of civil unrest events in six countries of Latin America. Focus of (Parrish et al., 2018) is on predicting six classes of events including Domestic Political Crisis, Insurgency, International Crisis, Rebellion, Ethnic/Religious Violence, and Irregular Leadership Change using a gated recurrent unit sequence. They use mutual information and information gain to select most relevant features to each task among a feature set with size 1160. They also compare their methods with the static classification models, often used in previous work for these prediction tasks.

All mentioned methods are limited in terms of the type of the events or countries that they cover. They also demand additional effort for feature selection. These Knowledge Graphs are also too noisy to be fully analyzed using the traditional machine learning techniques. We then describe more general models that try to exploit all of the aspects of the data together. We believe they lie under two major categories:
• KG Embedding Based Category

1.1 Tensor Based Category

These methods represent data as a 4 dimensional matrix $Y$ of size $N \times N \times A \times T$, where $N$ is the number of country actors and $A$ is the number of action types and $T$ is the time period. Schein et al. in (Schein et al., 2015) study the problem of inferring the latent structure of the multilateral relationships from dynamic events. They propose a Bayesian latent factor model for predictive and exploratory analysis. Their approach provides nice interpretability as it captures interesting patterns from the data that is immediately clear or easy to discover. In a similar approach, schein et. al in (Schein et al., 2016) propose a Bayesian Poison Tucker decomposition model that is able to discover directed community interaction networks that are specific to "topics" of action types and temporal "regimes". An example of interaction can be: country $i$ took action $a$ toward country $j$ at time $t$.

Although these approaches have the advantage of useful visualization and understanding of the data, they can be used for forecasting. They offer an inference technique that outperforms the similar baselines. The general idea behind it is that, given a tensor with some of its elements missing and some observed, the decomposition of it followed by reconstruction using the model must preserve the observed elements.

1.2 KG Embedding Based Category

A category of papers study the problem as reasoning over large Knowledge Graphs. There are several approaches to embed entities and relations in a low dimensional space so that we can use it for link prediction and node clustering. One of the earliest work is TranE (Bordes et al., 2013) which assumes that if triple $(s, r, t)$ exists, the embedded vector of $s$ and $t$ should be close when translated by $r$ ($s \overset{r}{\mapsto} t$). The other works in this area are RESCAL (Nickel et al., 2011) and NTN (Socher et al., 2013), which then have been shown in (Yang et al., 2014) that the most existing frameworks can be generalized under a learning framework. In (Yang et al., 2014) also propose a method that search over the embedding states and mine logical rules which can then be used for reasoning such as $\text{BornInCity}(a, b) \land \text{CityInCountry}(b, c) \Rightarrow \text{Nationality}(a, c)$. However, these static embeddings are mostly suitable for general fact knowledge graph such as FreeBase and several others that are mostly used in Natural Language Processing tasks.

The problem of reasoning and embedding temporal Knowledge graph was first studied in (Trivedi et al., 2017). They propose a deep recurrent network, that captures the evolving non-linear representation of entities and relations over time. They also use a point process method to predict the occurrence of a fact, where at each time step, the learned representation of the entities and relations would be used as the feature for the intensity function. Authors show that their approach outperforms all the other static baselines. It is also worth mentioning that in (Johnson et al., 2018) a self exciting point process is used for political conflicts forecasting, however it is rather a timeseries forecasting approach that doesn't assume any latent structure for countries.

In (Sadeghian et al., 2016) Sadeghian et. al have taken the rule mining approach and extend it to a temporal version. They use customized Ontological Pathfinding (OP) and by adding additional time sequence constraint to it, create a set of sequence rules extracted from temporal knowledge graph. It captures interesting rules that the static version ignores. They also propose Chrono-Translation which aims to utilize the current state of the art static embedding and at the same time adding an evolving nature to the embedding vector such that $e_{t+1} = e_t \phi_t$.

References


Gizem Korkmaz, Jose Cadena, Chris J Kuhlman, Achla Marathe, Anil Vullikanti, and Naren Ramakrishnan.


