Modeling Precursors for Event Forecasting via Nested Multi-Instance Learning

Yue Ning, Sathappan Muthiah, Huzefa Rangwala, Naren Ramakrishnan

Introduction

Event forecasting, such as civil unrest movements, disease outbreaks, and elections is an important and challenging problem. From the perspective of human analysts and policy makers, forecasting algorithms must not only make accurate predictions but also **provide supporting evidence**. In our work, we present a framework that helps to identify the causes related to the event of interest. The identified key documents are referred to as **precursors**.

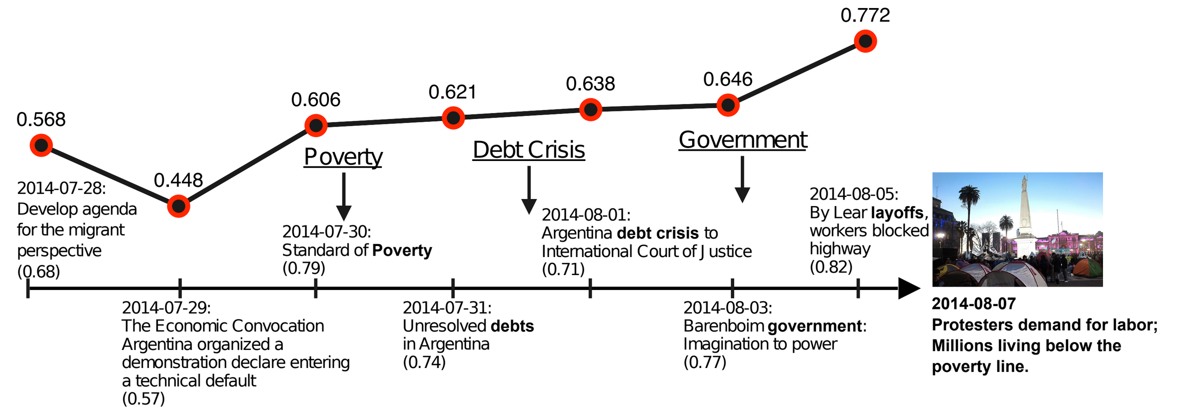

Figure 1: Precursor story line for a protest event in Argentina.

Objectives

Multiple instance learning algorithms are a class of supervised learning techniques that accept labels for groups of instances, but where labels for individual instances are not available. We formulate the precursor identification and forecasting problem in a novel multiple instance learning algorithm (MIL) setting.

1. A novel **nested framework of Multi-Instance Learning** for event forecasting and precursor mining.
2. Harness **temporal constraints** in Multi-Instance Learning.
3. Modeling for **various event categories in multiple geo-locations**.
4. Application and evaluation with comprehensive experiments.

Methods

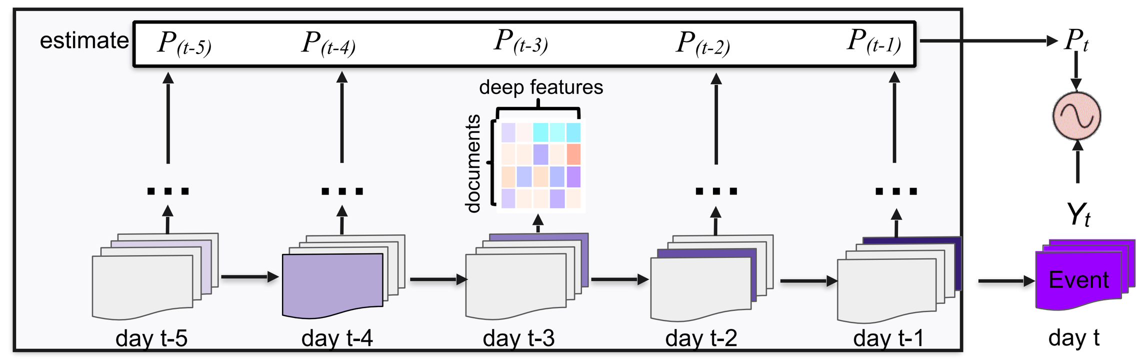

Figure 7: System Framework.

We model the instance level probability estimates p_{ij} for a news article j on day i to associate with a targeted event e with a logistic function.

$$p_{ij} = \sigma(\mathbf{w}^T\mathbf{x}_{ij}) = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}_{ij}}}. \quad (1)$$

The probability for a day (or bag) is then modeled as the average of probability estimates of all instances in a day. Hence, for each bag:

$$\mathbb{P}_i = \mathcal{A}(\mathcal{X}_i, \mathbf{w}) = \frac{1}{n_i}\sum_j^{n_i} p_{ij}, \quad (2)$$

The probability of a super-bag $\mathbb{S}$ (associated with an event e) being positive

$$P = \mathcal{A}(\mathbb{S}, \mathbf{w}) = \frac{1}{t}\sum_i^t \mathbb{P}_i \quad (3)$$

Loss function:

$$J(\mathbf{w}) = \frac{\beta}{n}\sum_{\mathbb{S}\in\mathcal{S}} f(\mathbb{S}, Y, \mathbf{w}) + \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\\ \mathcal{X}_i,\mathcal{X}_{i-1}\in\mathbb{S}}} \frac{1}{t}\sum_{i=1}^t g(\mathcal{X}_i, \mathcal{X}_{i-1}, \mathbf{w}) \quad (4)$$

$$+ \frac{1}{n}\sum_{\substack{\mathbb{S}\in\mathcal{S};\mathcal{X}_i\in\mathbb{S}\\ \mathbf{x}_{ij}\in\mathcal{X}_i}} \frac{1}{t}\sum_{i=1}^t \frac{1}{n_i}\sum_{j=1}^{n_i} h(\mathbf{x}_{ij}, \mathbf{w}) + \lambda R(\mathbf{w})$$

We update the cost function across days (bags) as follows:

$$g(\mathcal{X}_i, \mathcal{X}_{i-1}) = \Delta(\mathcal{X}_i, \mathcal{X}_{i-1})(P_i - P_{i-1})^2 \quad (5)$$

Predictive Performance

Overall Evaluation

	Argentina		Brazil		Mexico	
	Acc	F-1	Acc	F-1	Acc	F-1
SVM	0.611(±0.034)	0.406(±0.072)	0.693(±0.040)	0.598(±0.067)	0.844(±0.062)	0.814(±0.091)
MI-SVM	0.676(±0.026)	0.659(±0.036)	0.693(±0.040)	0.503(±0.087)	0.880(±0.025)	0.853(±0.040)
rMIL^{nor}	0.330(±0.040)	0.411(±0.092)	0.505(±0.012)	0.661(±0.018)	0.499(±0.009)	0.655(±0.025)
rMIL^{avg}	0.644(±0.032)	0.584 (±0.055)	0.509(±0.011)	0.513(±0.064)	0.785(±0.038)	0.768(±0.064)
GICF	0.589(±0.058)	0.624(±0.048)	0.650(±0.055)	0.649 (±0.031)	0.770(±0.041)	0.703(±0.056)
nMIL	**0.709**(±0.036)	0.702(±0.047)	**0.723**(±0.039)	0.686(±0.055)	**0.898**(±0.031)	**0.902**(±0.030)
nMIL$^{\Delta}$	0.708(±0.039)	**0.714**(±0.034)	0.705(±0.048)	**0.698**(±0.045)	0.861(±0.014)	0.868(±0.014)
nMIL$^{\Omega}$	0.687(±0.038)	0.680(±0.045)	0.713(±0.028)	0.687(±0.038)	0.871(±0.013)	0.879(±0.014)

Table 1: Event forecasting performance comparison based Accuracy (Acc) and F-1 score w.r.t to state-of-the-art methods. The proposed **nMIL** , **nMIL$^{\Delta}$**, **nMIL$^{\Omega}$**method outperform state-of-the-art methods across the three countries.

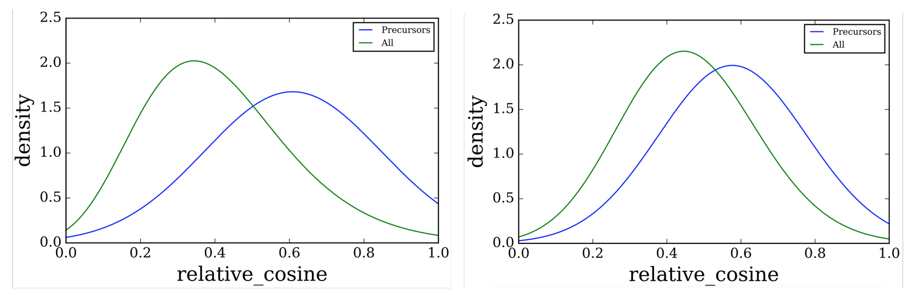

Figure 2: The distribution of relative cosine similarity for all documents (green line) and for precursor documents (blue line) with probability greater than 0.7

Precursors Evaluation

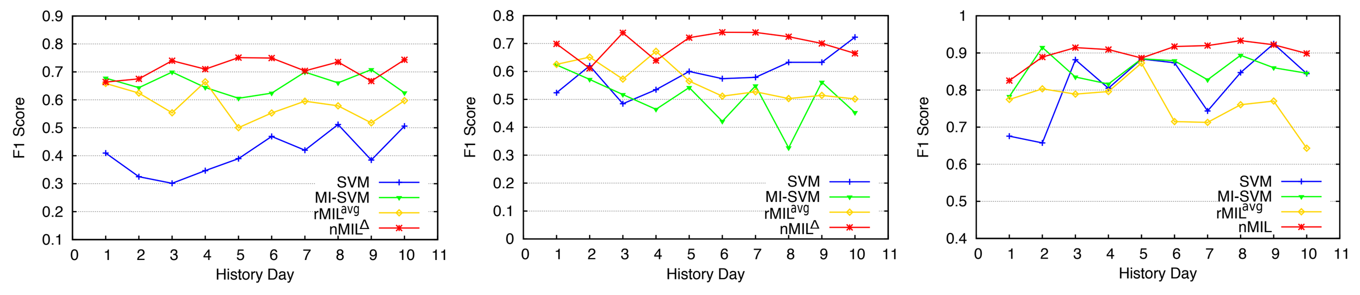

Figure 3: Forecasting evaluation on 3 countries with respect to F1 score. X-axis is the number of historical days used in the training process. Y-axis shows the average F1 score of 10 runs of experiments.

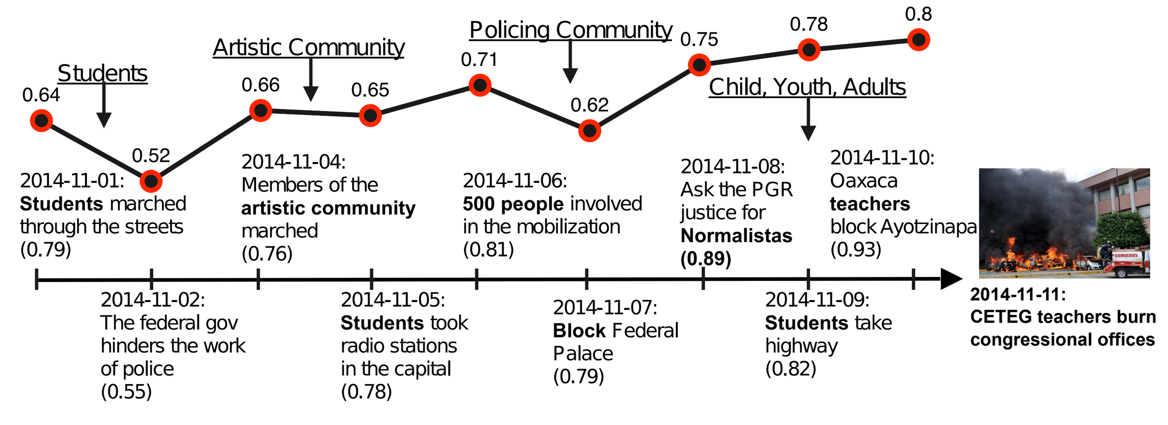

Figure 5: Precursor story line for a protest event in Mexico.

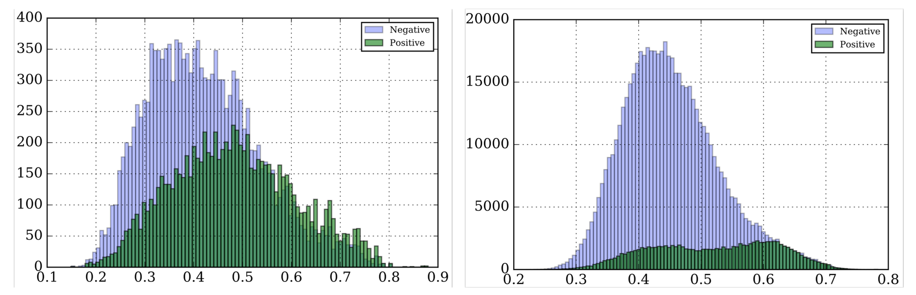

Figure 4: Estimated probabilities for negative examples (purple) and positive examples (green) for Argentina and Mexico.

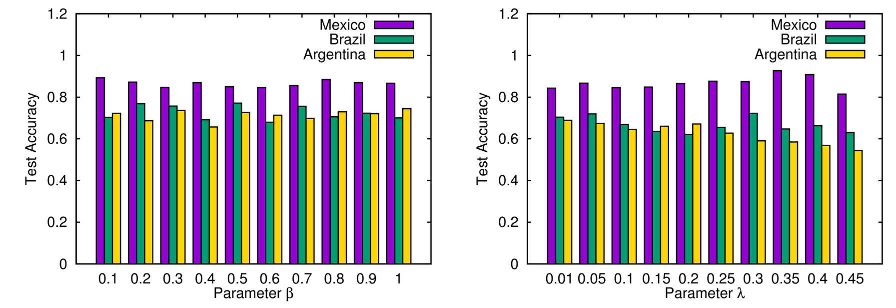

Figure 6: Sensitivity analysis on β and λ.

Selected Reference

[1] Zhi-Hua Zhou and Jun-Ming Xu. On the relation between multi-instance learning and semi-supervised learning. In *Proceedings of the ICML'07*, volume 227, pages 1167–1174, 2007.

Future Research and Acknowledgments

We plan to incorporate heterogeneous data sources in our method and also explore regularized Multi-Task Learning (MTL) approaches with spatial and temporal constraints.

Contact Information

Web http://people.cs.vt.edu/yning
Email yning@vt.edu
Phone +1 (703) 901 2956