**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.

**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 framework of Multi-Instance Learning for event forecasting and precursor mining.
3. Modeling for various event categories in multiple geo-locations.
4. Application and evaluation with comprehensive experiments.

**SELECTED REFERENCE**


**METHODS**

We model the instance level probability estimates $p_j$ for a news article $j$ on day $t$ to associate with a targeted event $e$ with a logistic function.

$$p_j = \frac{1}{1 + e^{-w^T x_j}}$$

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 $b$:

$$P_b = A(X_b, w) = \frac{1}{n_b} \sum_{j=1}^{n_b} p_j$$

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

$$P = A(S, w) = \frac{1}{l} \sum_{i=1}^{l} P_i$$

**Table 1:** Event forecasting performance comparison based Accuracy (Acc) and F-1 score w.r.t state-of-the-art methods. The proposed nMIL, $nMIL^2$, and $nMIL^3$ method outperform state-of-the-art methods across the three countries.

<table>
<thead>
<tr>
<th>Model</th>
<th>Argentina</th>
<th>Brazil</th>
<th>Mexico</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.611(±0.034)</td>
<td>0.628(±0.018)</td>
<td>0.698(±0.012)</td>
</tr>
<tr>
<td>MI-SVM</td>
<td>0.679(±0.026)</td>
<td>0.697(±0.016)</td>
<td>0.700(±0.017)</td>
</tr>
<tr>
<td>rMIL</td>
<td>0.705(±0.020)</td>
<td>0.704(±0.019)</td>
<td>0.705(±0.020)</td>
</tr>
<tr>
<td>nMIL</td>
<td>0.712(±0.022)</td>
<td>0.715(±0.019)</td>
<td>0.715(±0.021)</td>
</tr>
<tr>
<td>$nMIL^2$</td>
<td>0.786(±0.031)</td>
<td>0.789(±0.034)</td>
<td>0.788(±0.034)</td>
</tr>
<tr>
<td>$nMIL^3$</td>
<td>0.807(±0.033)</td>
<td>0.807(±0.036)</td>
<td>0.807(±0.036)</td>
</tr>
</tbody>
</table>

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

**Figure 6:** Sensitivity analysis on $\alpha$ and $\beta$.

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

**FUTURE RESEARCH AND ACKNOWLEDGMENTS**

**CONTACT INFORMATION**

Web: http://people.cs.vt.edu/~yning
Email: yning@vt.edu
Phone: +1 (703) 901 2596