Tactical Analytics

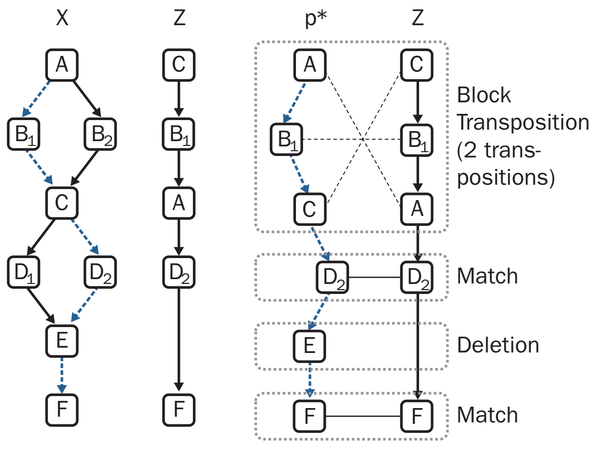
Recognizing Patterns of Life and Determining Credibility of Textual Data

Structural Multi-Task Transfer Learning

To support analysis of real-time streaming data for situational awareness, we created methods for recognition patterns in textual data and determining credibility of textual data.

Patterns of Life: To recognize patterns of life in textual data we use the concept of "scripts". A script is a series of ordered, related events that describe a stereotypical pattern that adversaries follow during military and other activities. Scripts allow analysts to recognize these patterns and make predictions about emerging events. This year's work was focused on automatically identifying scripts from streaming data, accounting for multiple pathways through the script.

Comparing Sequence Z against Script X:



Lessons Learned:

1. Scripts can be learned from streaming data
2. Constraints are necessary to avoid obviously invalid pathways
3. Even a simple test case is very complicated

Measures of Similarity

$$s(X,Z)=1-\arg\min_{p\in Path(x)}\left[\frac{\sum_{x\in p}\beta_x\delta(x,Z)}{\sum_{x\in p}\beta_x}\right]$$

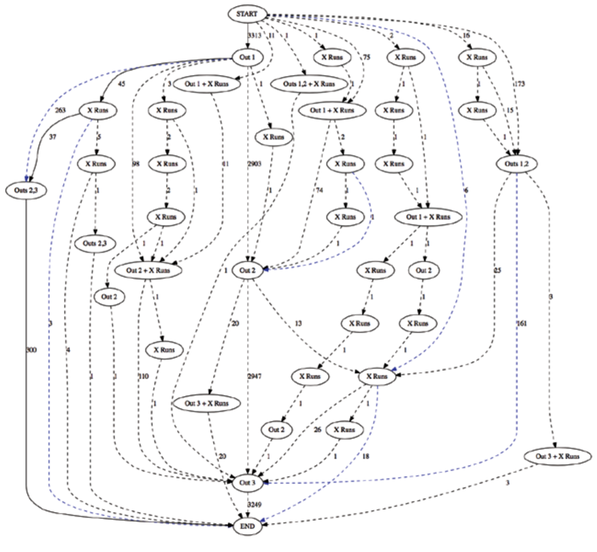
$$\delta(x,Z)=\begin{cases}0 & \text{a match B.T.}\\ \alpha_1 & \text{if insertion}\\ \alpha_1 & \text{deletion}\\ \alpha_1 & \end{cases}$$

$$\beta_x = 1 \text{ unless specified otherwise by the user}$$

Challenges:

1. State-of-the-art event recognition algorithms proved insufficient for our task. Solution: We used data from baseball box scores that allowed easy event extraction. FY17 work will extend DARPA algorithms for single & multiple sentence event recognition. Script recognition will ultimately require recognizing events across multiple dissimilar documents.
2. Establishing event relationships must be improved. Solution:FY16 work involved creating constraints for order and uniqueness. FY17 work will extend this work.

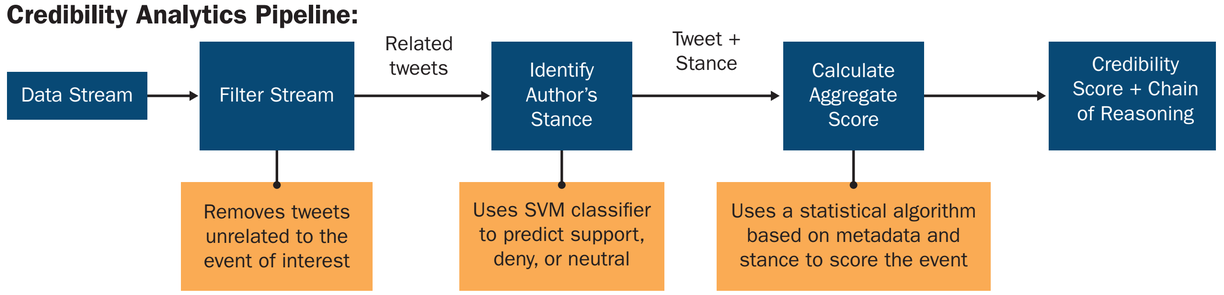
Generated Script for Baseball ½ Innings



Determining Credibility Scores of Streaming Social Media Data

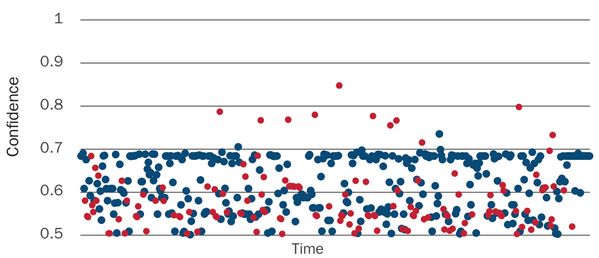
This work depends on accurate event detection. As a proxy, we used 3 celebrity death events and 80 diverse events from Twitter*

Credibility Analytics Pipeline:

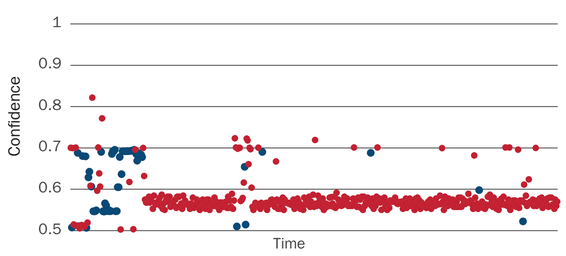


Twitter Events

True: Castaic Earthquake

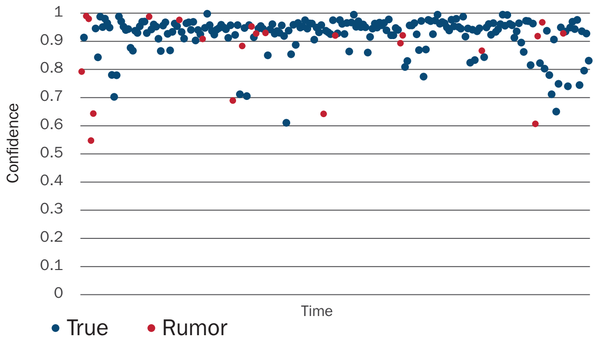


Rumor: Obama Bans Sprinkles

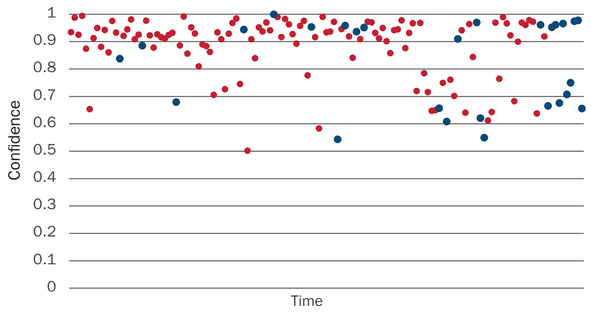


Celebrity Death Events

Whitney Houston Died



Paul McCartney Died



Dataset	All	True			False		
	Accuracy	P	R	F1	P	R	F1
Celebrity	.85	.82	.90	.85	.88	.80	.84
Twitter Events	.61	.58	.77	.67	.66	.45	.54

Lessons Learned:

1. Stance determination is essential
2. Noise is difficult to filter; we need accurate event recognition

Future Work: We need to remove more noise from the social media data in step #1 of the analytics pipeline. Step #2 must be improved to generalize to more event types. Step #3 requires external sources to improve the credibility assessment of the entities providing information.

*Zou, J., Fekri, F., & McLaughlin, S. W. (2015, August). Mining Streaming Tweets for Real-Time Event Credibility Prediction in Twitter. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015(pp. 1586-1589). ACM.

Distribution Statement A: Approved for Public Release; Distribution is Unlimited

Contact: Ed Morris | ejm@sei.cmu.edu