# REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188



| 1. REPORT DATE (DD-MM-YYYY) | 2. REPORT TYPE | 3. DATES COVERED (From - To) |
|---|---|---|
| | Technical Report | - |

4. TITLE AND SUBTITLE
IU Progress Report January 2013

5a. CONTRACT NUMBER
W911NF-12-1-0037

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS
Alessandro Flammini, Filippo Menczer, Sergey Malinchik, Qiaozhu Mei

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Indiana University at Bloomington
Trustees of Indiana University
509 E 3RD ST
Bloomington, IN 47401 -3654

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
61766-NS-DRP.16

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Our group has been working in the following directions: 1. Testing and tuning our time series classification algorithm, 2. Perfectioning the use of entropy in language models as a tool to reconstruct the context of a given meme/conversation, 3. formalizing in an operational sense the definition of campaign and its associated features, 4. collecting relevant datasets to test detection algorithms.

15. SUBJECT TERMS
entropy (in language models), time series classification, campaign detection

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 15. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Alessandro Flammini |
| UU | UU | UU | UU | | 19b. TELEPHONE NUMBER 812-856-1830 |

**Report Title**

IU Progress Report January 2013

**ABSTRACT**

Our group has been working in the following directions: 1. Testing and tuning our time series classification algorithm, 2. Perfectioning the use of entropy in language models as a tool to reconstruct the context of a given meme/conversation, 3. formalizing in an operational sense the definition of campaign and its associated features, 4. collecting relevant datasets to test detection algorithms.

**DARPA SMISC Project:**
DESPIC: Detecting Early Signatures of Persuasion in Information Cascades

**Teams:**
Indiana University: A. Flammini (PI) and F. Menczer
University of Michigan: Qiaozhu Mei
Lockheed Martin Advanced Technology Laboratories (ATL): S. Malinchik

## Progress Report – January 2013

**IU:** During January 2013 the IU Team has worked on the definition and implementation of a protocol to identify and generate time-series that represent ongoing campaigns on Twitter. In particular, we focused on the following points:

1. Define campaigns identification criteria
2. Formalize a classification of the various types of existing campaigns
3. Determine features that characterize data
4. Design methodologies of time-series data generation

According to our first goal, we determined four different classes of campaigns that are popular on Twitter: rumors, advertisement, astroturfing and anti-campaigns. Rumors are characterized by vast spread and controversy on credibility of information. Advertisement campaigns aim at attracting users' attention while clearly stating the brand/offer that is promoted. Astroturfing campaigns aim at smearing particular individuals/corporation by simulating grassroot discussions through orchestrated efforts. Anti-campaigns are all those campaign hijackings or defacing of orchestrated campaigns.

A first attempt has been done in the direction of rumors identification and classification. We manually selected, among all trending topics observed in Twitter during latest months, 10 specific rumors. This set includes false rumors such as Justin Bieber cancer, NASA announcing a global power outage for Dec. 22nd 2012, and others. After manually generating keyword-based search-queries for all these rumors, we collected data from our in-house Twitter gardenhose dataset, by fetching tweets across several months in the period previous of the given rumor appearance. We also started following specific memes (keywords and hashtags) via the Twitter "search and tracking API", to observe the evolution of each given rumor in the near future. The rumor dataset that we obtained will be available for future classification and validation purposes.

One of the main limits that we encountered dealing with rumors is the need of massive human efforts to identify relevant rumors and verify their nature across the enormous amount of memes produced every day on Twitter. The IU team has later decided to investigate a potential method to detect memes belonging to campaigns in a semi-automatic or fully automatic fashion. To this purpose we decided to focus our attention on the class of advertised campaigns ongoing on Twitter due to their

ease of identification and the possibility of automatic extraction and classification. In detail, Twitter offers, as an advertising solution, the chance of promoting particular hashtags or phrases, which will appear together with trending topics in users' pages.

We designed and implemented a system capable of extracting, continuously and at predetermined intervals, trending topics and promoted content from Twitter. Once new trending and promoted memes are identified, on an hourly base, the system automatically extracts from our data storage layer all tweets exhibiting each given meme (as of the date, this is done by accessing our in-house Twitter gardenhose dataset; later we expect to do the same, on a much larger scale, by querying the PeopleBrowsr API that is currently under development). These tweets are subsequently processed so that time-series related to each meme can be generated, for each feature that we would like to analyze.

We determined a set of features that will be instrumental to build our classification infrastructure. At the current stage we designed and developed the system to extract those that we identify as network features. In the near future we will extend this system so that to be capable of extracting additional classes of features, such as sentiment-, content-, and geography-related ones. Network features currently available include, among others, general network statistics (e.g., no. nodes and edges), distributions (e.g., in/out node degree, weight and strength, etc.), largest connected component size, diameter, assortativity, and so on. We expect to expand this set of features including other potentially relevant network features (e.g., centrality, etc.) in the near future.

Starting from the beginning of January, we isolated more than 20 promoted content hashtags and phrases and more than three thousand trending topics. For each of them we built time-series of features including data from one week before and two weeks after the given meme has become trending or promoted. This is done on the purpose of isolating potential predictive patterns in the feature set that might help in our future work of classification of genuine or artificial campaigns. Moreover, for each feature, the system can concurrently produce data related to three different types of network: i) hashtag co-occurrence, ii) retweet, and iii) mention networks. The possibility of adopting different types of network will be instrumental to search for specific patterns of diffusions (e.g., considering the retweet network) or topic emergency (e.g., considering the hashtag co-occurrences). In the future we expect to extend the system so that to be able to exploit additional network types (e.g., follower networks).

This dataset is currently undergoing standard data cleansing and sanity check protocols so that to become available for future analysis in the next weeks.

**UM:** In the past month, we moved forward to study the time series of entropy of language models, which we found as a promising direction in the previous exploration. Specifically, given an individual meme, we are able to construct the language model that represents the context of the meme, both from the information

needs and from the general background. Entropy of the language models are computed and tracked over time. This microscopic analysis of the entropy time series helps us to understand how the discussion of particular topics concentrates and diverges. We believe that this analysis can help us extract signals of persuasion campaigns with the assumption that a campaign may intrigue either the concentration or the divergence of the discussion and people's information needs of a topic. Below are some preliminary results of the analysis.

In general, we can find from the examples that although the general trends of the volume of tweets correlate in information needs and background, the series of entropy in the two different contexts differ significantly. There are patterns that the trend of entropy differs completely in information needs and in background (see Figure 1 and 2), patterns that the trend of entropy in information needs predicts the trend in background (see Figure 3 and 4), and vice versa. These patterns will be explored in characterizing and identifying persuasion campaigns in the next step.

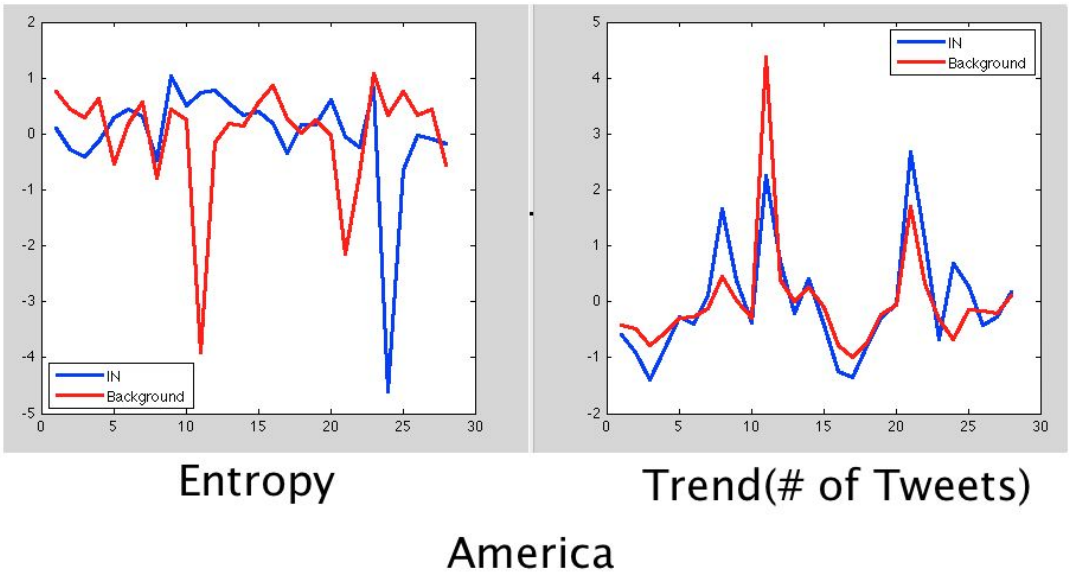



Figure 1: The trend of the volume of tweets and the entropy of context of selected keyword: America. Different series are provided for information needs and the general background tweets.

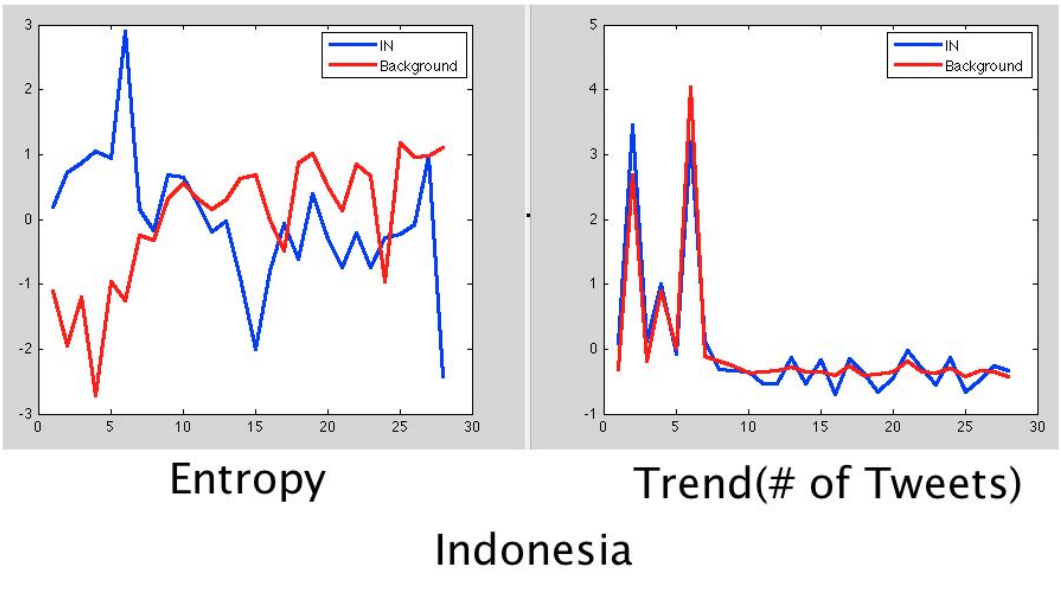



Figure 2: The trend of the volume of tweets and the entropy of context of selected keyword: “Indonesia”. Different series are provided for information needs and the general background tweets.

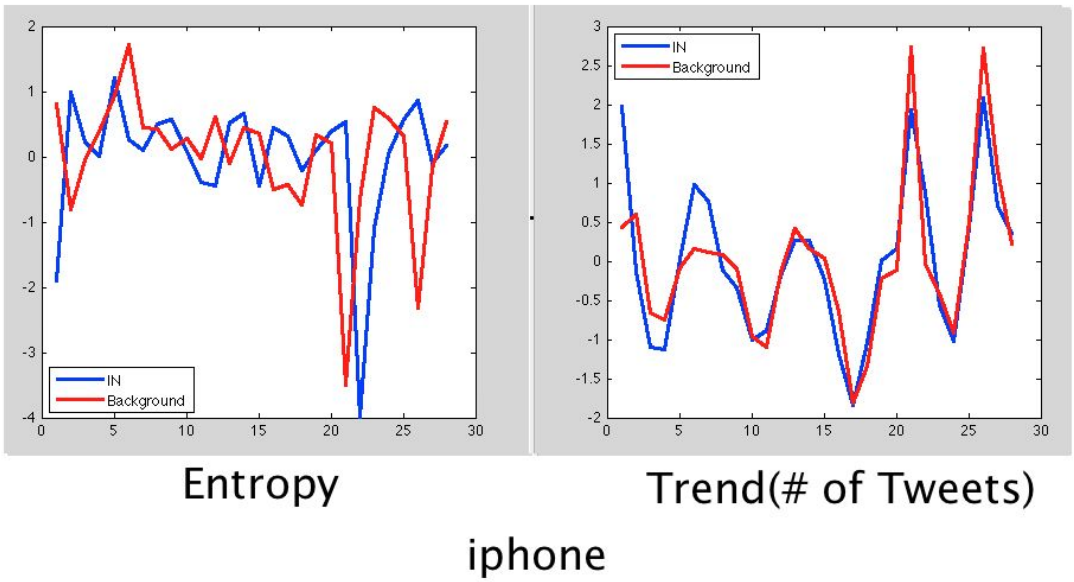



Figure 3: The trend of the volume of tweets and the entropy of context of selected keyword: “Iphone”. Different series are provided for information needs and the general background tweets.

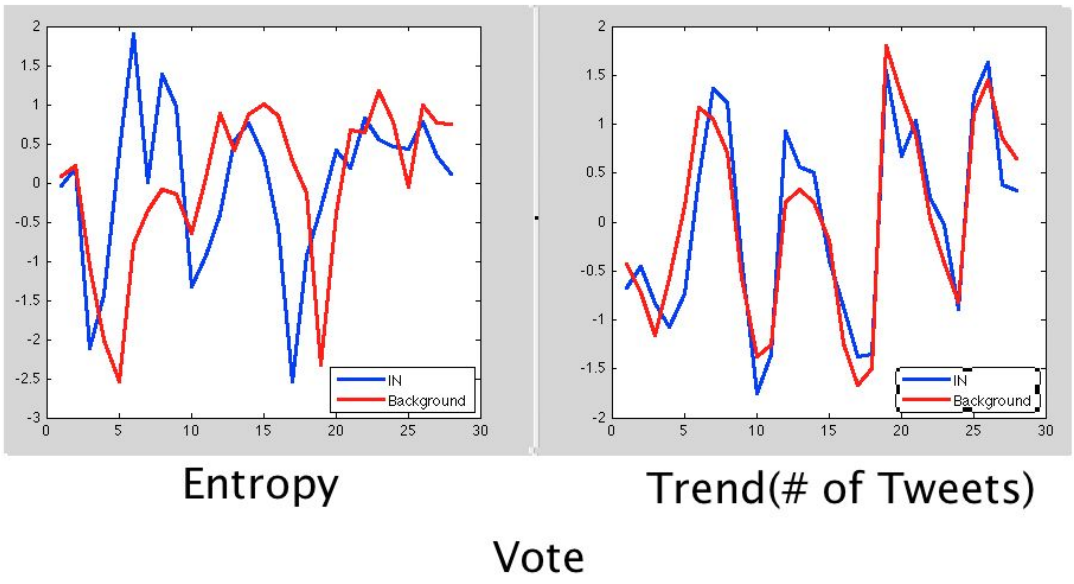



Figure 4: The trend of the volume of tweets and the entropy of context of selected keyword: "Vote". Different series are provided for information needs and the general background tweets.

**ATL:** In January 2013 ATL team was working on testing and tuning our SAX-TF*IDF time series classification algorithm. The basic idea of SAX [1] (Symbolic Aggregate ApproXimation) is to convert time series into a symbolic string with a small alphabet size (Fig.5).

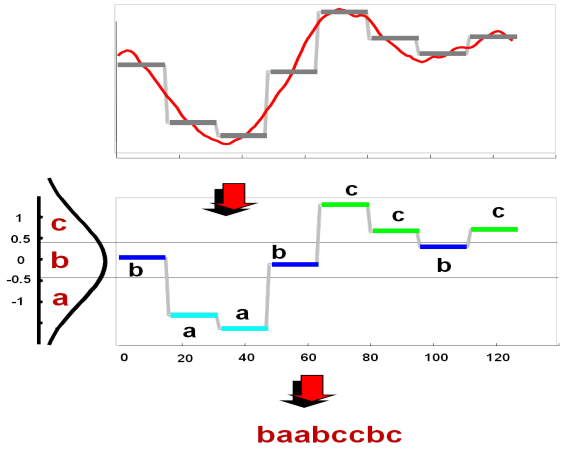
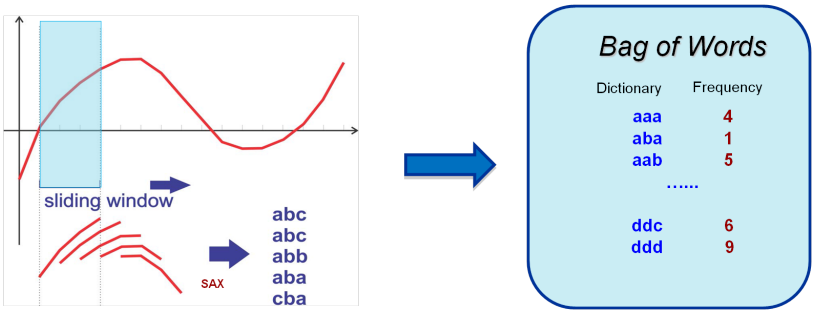



Figure 5: A Symbolic Representation of Time Series using SAX algorithm.

To preserve unique features of long time series we combine sliding window technique with SAX algorithm to transform time series into a set of strings as shown in Fig. 6a. Calculating frequencies of all strings (words) we represent the time series as a "bag of words" (Fig. 6b).




(a) (b)

Figure 6: Transformation of time series into a Bag-of-Words.

To compare Bags of Words we treat them as documents applying a Vector Space Model and calculating TF-IDF weight vectors. Two time series, **A** and **B,** can be compared by calculating cosine similarity of corresponding TF*IDF weight vectors:

$$sim(\vec{A},\vec{B}) = \cos(\theta) = \frac{\vec{A}\cdot\vec{B}}{\|\vec{A}\|\times\|\vec{B}\|} = \frac{\sum_{i=1}^{n} a_i \cdot b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\cdot\sqrt{\sum_{i=1}^{n} b_i^2}}$$

To test classifier based on our SAX-TF*IDF technique we used a CBF dataset - widely used synthetic time-series benchmark. The CBF curves are generated by choosing two random parameters, a and b, that characterize the beginning and the end of the signal (Fig. 7).

SAX-TF*IDF classification process can be described as follows:

- We generate three Mixed Bags-of-Words representing each of the classes.
- Choosing a training set size, N, we generate 3N labeled time series, processing them by SAX and accumulate results in three Bags-of-Words
- Calculate TF*IDF weight vectors for each class.
- Each unknown test sample is converted into a bag of SAX words and then into TF*IDF vector.
- The distance of the sample to all three TF*IDF vectors representing three known classes is computed using cosine similarity metrics.
- The unknown sample is assigned to the closest one of the known classes - Cylinder, Bell or Funnel.

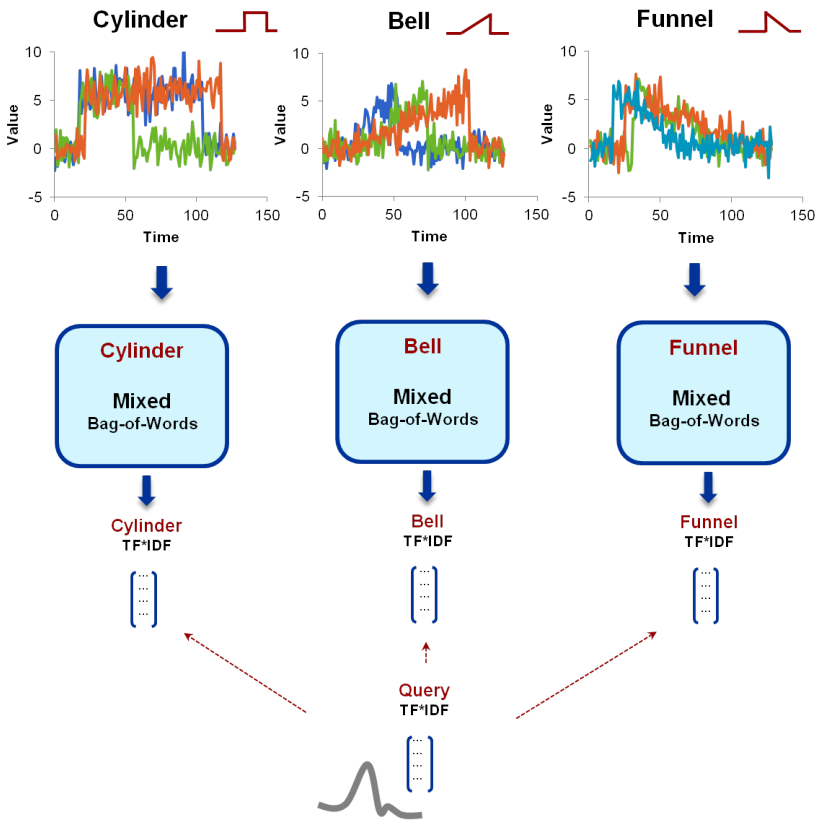



Figure 7: SAX-TF*IDF classification process applied to Cylinder-Bell-Funnel (CBF) families of synthetic time series.

We use as a reference a common simple 1-NN classifier with Euclidean distance as a similarity measure. Both classifiers are sensitive to the size of training set but for small training sets the SAX-TF*IDF classifier outperforms 1-NN classifier (Fig.8). Computationally SAX-TF*IDF classifier becomes less expensive for large training sets. As can be seen in Fig.8 for training size of 100 the SAX-TF*IDF classifier gives ~99.9% accuracy.

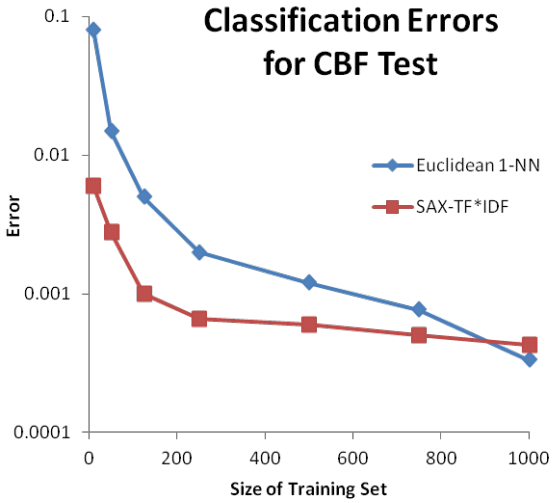
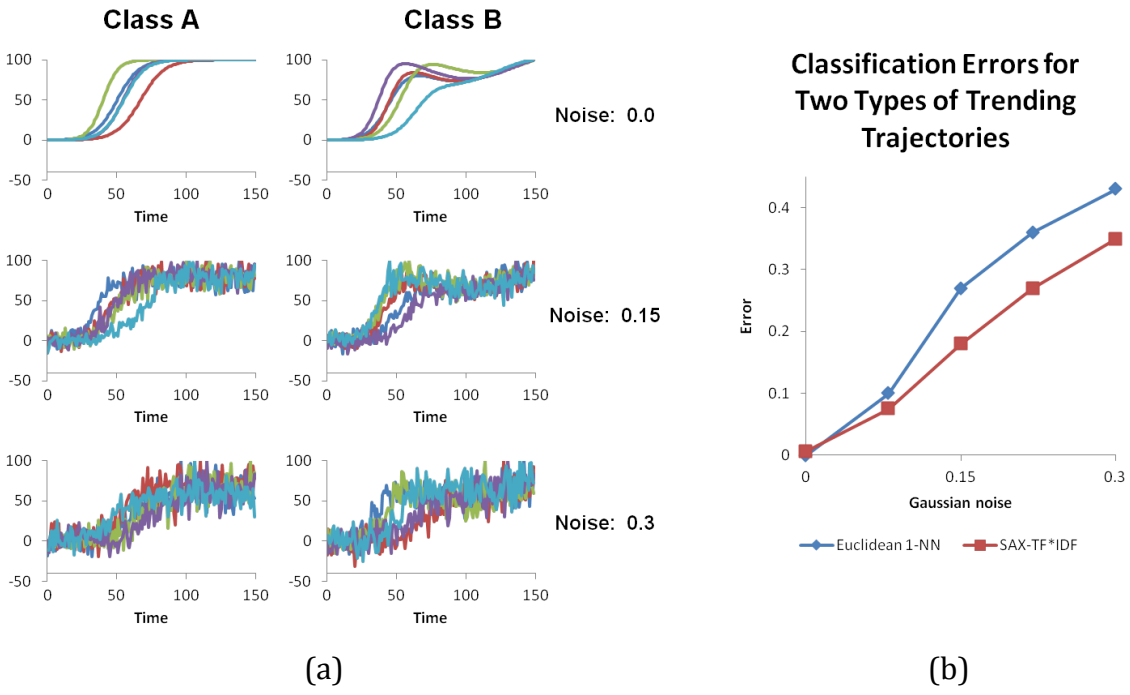



Figure 8: Cylinder-Bell-Funnel (CBF) classification benchmark for two types of classifiers.

To explore sensitivity of SAX-TF*IDF classifier to noisy data we created two classes of synthetic data with ability to control level of noise (Fig. 9a).




(a) (b)

Figure 9: Classification results of two families of noisy data.

As can be seen from Fig. 9b, SAX-TF*IDF classifier gives better results for the data with significant level of noise.

**References**

[1] Lin, J., Keogh, E., Lonardi, S. & Chiu, B. (2003) A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA