Semi-Supervised Time Series Classification

Mojdeh Jalali Heravi

Introduction

- **Time series** are of interest to many communities:
  - **Medicine**
  - **Aerospace**
  - **Finance**
  - **Business**
  - **Meteology**
  - **Entertainment**
  - **....**

Introduction

- Current methods for time series classification:

  Large amount of labeled training data

  - Difficult or expensive to collect
    - Time
    - Expertise

Introduction

- On the other hand …

  Copious amounts of **Unlabeled data** are available

  - For example: **PhysioBank archive**
    - More than 40 GBs of ECG
    - Freely available
    - In hospitals there are even more!

  Semi-Supervised classification

  → takes advantage of large collections of **Unlabeled data**

The paper…

Li Wei, Eamonn Keogh, Semi-Supervised time series classification, *In Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006*

Outline

- Applications
- Value of unlabeled data
- Semi-supervise learning
- Time series classification
- Semi-supervised time series classification
- Empirical Evaluation

Applications

- Indexing of handwritten documents

  Google and YAHOO! are interested in making large archives of handwritten text searchable.

  – For indexing first the words should be classified.

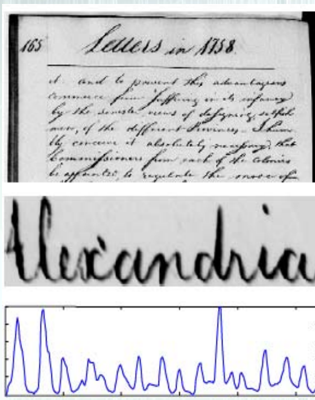
  – Treating the words as **time series** is an competitive approach.

Applications

- a classifier for George Washington will not generalize to Isaac Newton

•Obtaining labeled data for each word is expensive

•Having **few training examples** and using **semi-supervised approach** would be great!



A sample of text written by George Washington

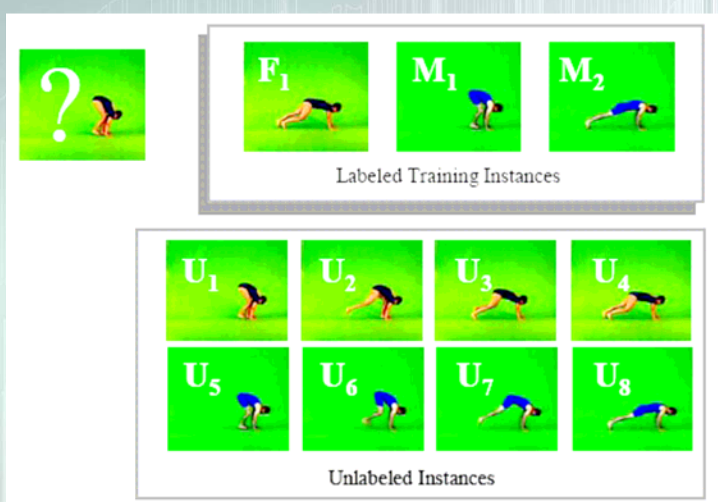
Applications

- Heartbeat Classification
  - PhysioBank
    - More than 40 GBs of freely available medical data
    - A potential goldmine for a researcher
    - Again, Having **few training examples** and using **semi-supervised approach** would be great!

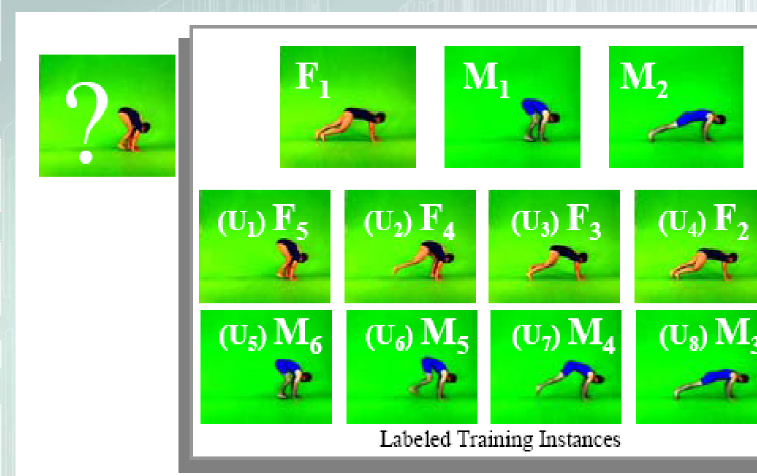
Outline

- Applications
- Value of unlabeled data
- Semi-supervise learning
- Time series classification
- Semi-supervised time series classification
- Empirical Evaluation

Value of unlabeled data



Value of unlabeled data



Outline

- Applications
- Value of unlabeled data
- Semi-supervise learning
- Time series classification
- Semi-supervised time series classification
- Empirical Evaluation

Semi-supervised Learning

- Classification → supervised learning
- Clustering → unsupervised learning

Learning from both labeled and unlabeled data is called
semi-supervised learning

Less human effort

Higher accuracy

Semi-supervised Learning

– Five classes of SSL:

1. Generative models

- the oldest methods
- Assumption: the data are drawn from a mixture distribution that can be identified by large amount of unlabeled data.

✓ Knowledge of the structure of the data can be naturally incorporate into the model

✗ There has been no discussion of the mixture distribution assumption for time series data so far

Semi-supervised Learning

– Five classes of SSL:

2. Low density separation approaches

- ***"The decision boundary should lie in a low density region"*** → pushes the decision boundary away from the unlabeled data
- To achieve this goal → maximization algorithms (e.g. TSVM)

✗ "*(abnormal time series) do not necessarily live in sparse areas of n-dimensional space*" and "*repeated patterns do not necessarily live in dense parts*". Keogh *et. al.* [1]

Semi-supervised Learning

– Five classes of SSL:

3. Graph-based semi-supervised learning

- *"the (high-dimensional) data lie (roughly) on a low-dimensional manifold"*
- Data → nodes
  distance between the nodes → edges
- Graph mincut [2], Tikhonov Regularization [3], Manifold Regularization [4]

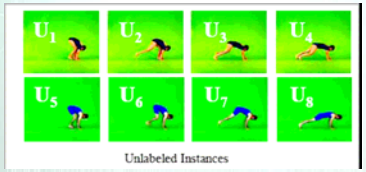
✗ The graph encodes prior knowledge → its construction needs to be hand crafted for each domain. But we are looking for a general semi-supervised classification framework

Semi-supervised Learning

– Five classes of SSL:

4. Co-training

- Features → 2 disjoint sets
  - assumption: features are independent
  - each set is sufficient to train a good classifier
- Two classifiers → on each feature subset
  - The predictions of one classifier are used to enlarge the training set of the other.



shape
color

✗ Time series have very high feature correlation

Semi-supervised Learning

– Five classes of SSL:

5. Self-training

- Train → small amount of labeled data
- Classify → unlabeled data
  - Adds the most confidently classified examples + their labels into the training set
  - This procedure repeats → classifier refines gradually

✓ The classifier is using its own predictions to teach itself → it's **general** with **few assumptions**

Outline

- Applications
- Value of unlabeled data
- Semi-supervise learning
- Time series classification
- Semi-supervised time series classification
- Empirical Evaluation

Time Series

- **Definition 1**. *Time Series:*

  A time series $T = t_1, \ldots, t_m$ is an ordered set of m real-valued variables.
  - Long time series
  - Short time series → subsequences of long time series

Definition 2. *Euclidean Distance:*

$$D(Q,C) \equiv \sqrt{\sum_{i=1}^{n} (q_i - c_i)^2}$$

Time Series Classification

- Positive class
  - Some structure
  - positive labeled examples are rare, but unlabeled data is abundant.
  - Small number of ways to be in class
- Negative class
  - Little or no common structure
  - essentially infinite number of ways to be in this class

$\boxed{\text{We focus on } \textbf{binary} \text{ time series classifiers}}$

Outline

- Applications
- Value of unlabeled data
- Semi-supervise learning
- Time series classification
- Semi-supervised time series classification
- Empirical Evaluation

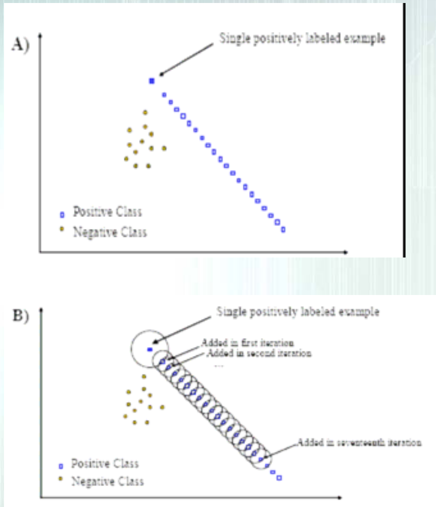
Semi-supervised Time Series Classification

- 1 nearest neighbor with Euclidian distance

Approach	Error Rate
One-nearest-neighbor with Euclidean distance [22]	**1.3%**
First order logic rules with boosting [38]	3.6%
Multi layer perceptron neural network [30]	1.9%
Multiple classifier system [11]	7.2%
Multi-scale histogram approach [12]	6.0%

On Control-Chart Dataset

- Training the classifier (example)



Semi-supervised Time Series Classification

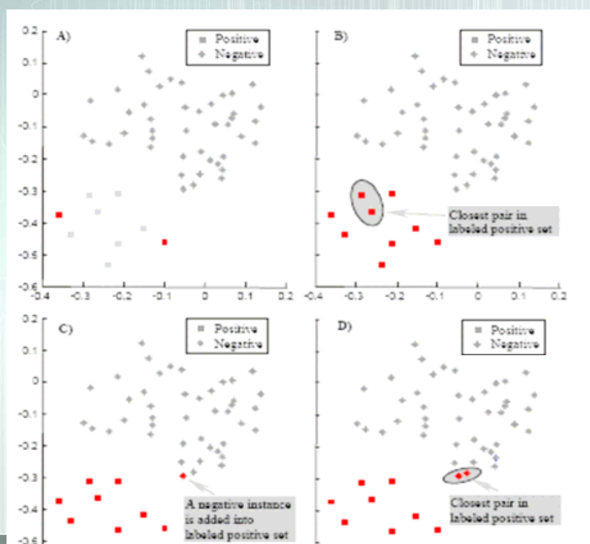
- Training the classifier (algorithm)

P→ positively labeled examples
U→ unlabeled examples

Function [*P*] = Semi_Supervised_Classification(*P*, *U*)
Until (some stopping criterion)
use *P* and *U* to train the one-nearest-neighbor classifier *C*
use classifier *C* to classify unlabeled set *U*
select the example that *C* most confidently labels as positive
add this example into *P*
delete this example from *U*
End

Semi-supervised Time Series Classification

- Stopping criterion (example)



Semi-supervised Time Series Classification

- Stopping criterion

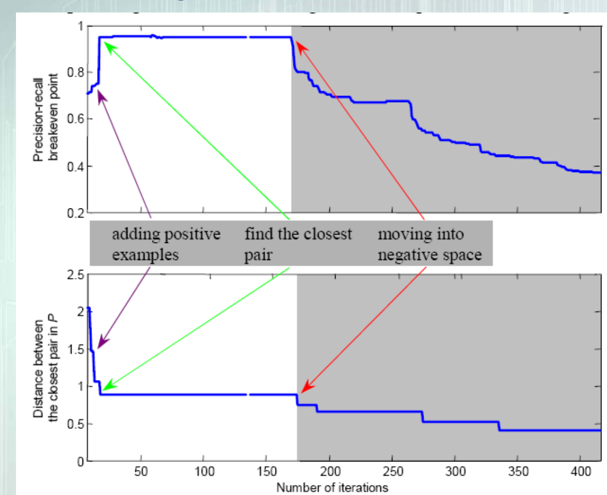


Figure 5: Statistics on ECG dataset

Semi-supervised Time Series

- Using the classifier
  - For each instance to be classified, check whether its nearest neighbor in the training set is labeled or not
  - the training set is huge

✗ Comparing each instance in the testing set to each example in the training set is untenable in practice.

Semi-supervised Time Series Classification

- Using the classifier
  - a modification on the classification scheme of the 1NN classifier

    using ***only*** the labeled positive examples in the training set

  - To classify:
    - within ***r*** distance to any of the labeled positive examples → positive
    - otherwise → negative.
  - ***r*** → the average distance from a positive example to its nearest neighbor

Outline

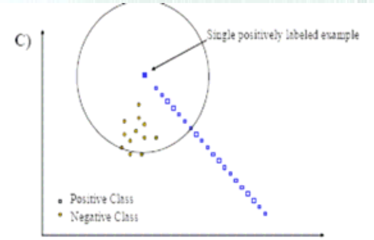
- Applications
- Value of unlabeled data
- Semi-supervise learning
- Time series classification
- Semi-supervised time series classification
- Empirical Evaluation

Empirical Evaluation

- Semi-supervised approach

Compared to:

- Naïve KNN approach
  - K nearest neighbor of positive example → positive
  - Others → negative
  - Find the best k



Empirical Evaluation

- Performance
  - class distribution is skewed → accuracy is not good

    96% negative

    4% positive

    if simply classify everything as negative

    accuracy = 96%
  - Precision-recall breakeven point
    - Precision = recall

$$\text{Precision} = \frac{\#\text{ of correct positive predictions}}{\#\text{ of positive predictions}}$$

$$\text{Recall} = \frac{\#\text{ of correct positive predictions}}{\#\text{ of positive examples}}$$

Empirical Evaluation

- Stopping heuristic
  - Different from what was described before

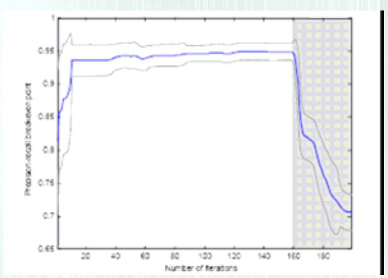
  Keep training until it achieves the highest precision-recall + few more iterations
- Test and training sets
  - For more experiments → distinct
  - For small datasets → same
    - still non-trivial → most data in training dataset are unlabeled

ECG Dataset

- ECG dataset form MIT-BIH arrhythmia Database

	Training Set	Testing Set
Positive (Abnormal)	208	312
Negative (Normal)	602	904
Total	810	1,216

  - # of initial positive examples = 10
  - Run 200 times
    - Blue line → average
    - Gray lines → 1 SD intervals



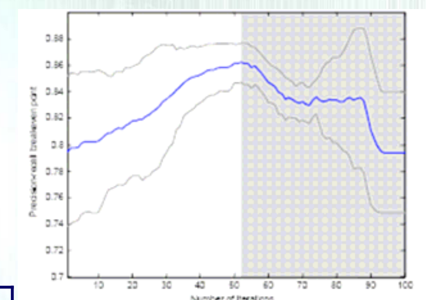
approach	P-R
Semi-supervised	94.97%
KNN (k = 312)	81.29%

Word Spotting Dataset

- Handwritten documents

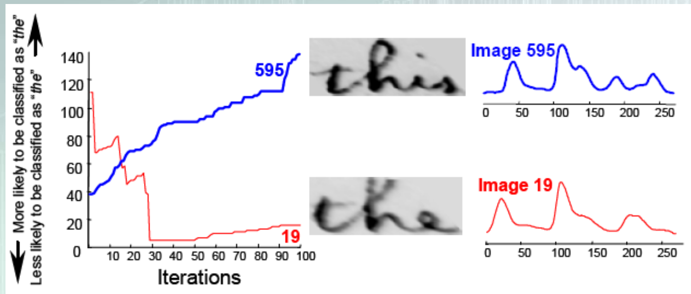
	Training Set	Testing Set
Positive (Word "the")	109	109
Negative (Other words)	796	796
Total	905	905

  - # of initial positive examples = 10
  - Run 25 times
    - Blue line → average
    - Gray lines → 1 SD intervals



approach	P-R
Semi-supervised	86.2%
KNN (k = 109)	79.52%

Word Spotting Dataset



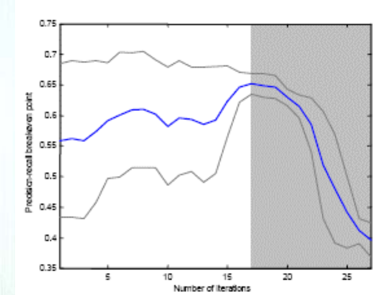
↑distance from positive class → ↑rank
→ ↓probability to be in positive class

Gun Dataset

- 2D time series extracted from video
- Class A: Actor 1 with gun
- Class B: Actor 1 without gun
- Class C: Actor 2 with gun
- Class D: Actor 2 without gun

	Training Set	Testing S
Positive (Class A)	27	
Negative (Class B,C,D)	95	
Total	122	

  - # of initial positive examples = 1
  - Run 27 times



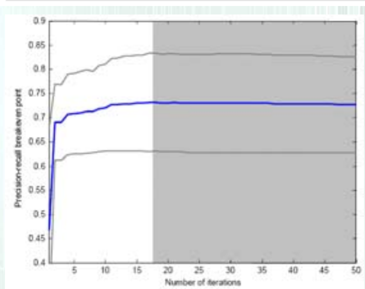
approach	P-R
Semi-supervised	65.19%
KNN (k = 27)	55.93%

Wafer Dataset

- a collection of time series containing a sequence of measurements recorded by one vacuum-chamber sensor during the etch process of silicon wafers for semiconductor fabrication

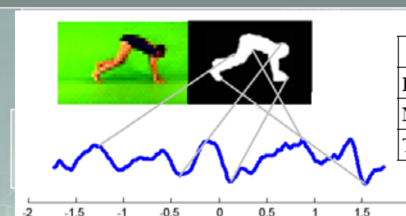
	Training Set	Testing Set
Positive (Abnormal)	381	381
Negative (Normal)	3,201	3,201
Total	3,582	3,582

  - # of initial positive examples = 1



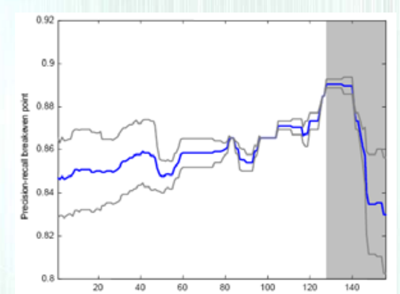
approach	P-R
Semi-supervised	73.17%
KNN (k = 381)	46.87%

Yoga Dataset



	Training Set	Testing Set
Positive (Female)	156	156
Negative (Male)	150	150
Total	306	306

# of initial positive examples = 1



approach	P-R
Semi-supervised	89.04%
KNN (k = 156)	82.95%

Conclusion

- An accurate semi-supervised learning framework for time series classification with small set of labeled examples
- Reduction in # of training labeled examples needed → dramatic

References

[1] Keogh, E., Lin, J., & Fu, A. (2005). HOT SAX: Efficient finding the most unusual time series subsequence. In *proceedings of the 5th IEEE International Conference on Data Mining (ICDM 2005)*, pp. 226-233, 2005.

[2] Blum, A. & Chawla, S. (2001). Learning from labeled and unlabeled data using graph mincuts. In *proceedings of 18th International Conference on Machine Learning*, 2001.

[3] Belkin, M., Matveeva, I., & Niyogi, P. (2004). Regularization and semi-supervised learning on large graphs. *COLT*, 2004.

[4] Belkin, M., Niyogi, P., & Sindhwani, V. (2004). Manifold regularization: a geometric framework for learning from examples. Technical Report TR-2004-06, University of Chicago.

Thanks …

"Thanks for your patience"

any question?