Performance Analysis of Different Time Series Classification Techniques

Zachary Saegesser, Khayyam Saleem

CS559 – Machine Learning Fundamentals and Applications

Outline

The Dataset

Weak Classifiers

Advanced Classifiers

Conclusion

Topic

The Dataset

Weak Classifiers

Advanced Classifiers

Conclusion

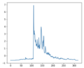
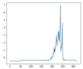
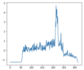
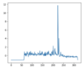
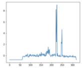
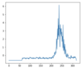
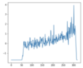
Lightning7

- The FORTE satellite detects transient electromagnetic events associated with lightning using a suite of optical and radio-frequency (RF) instruments.
- A Fourier transform is performed on the input data to produce a spectrogram.
- The spectrograms are then collapsed in frequency to produce a power density time series, which is then smoothed.

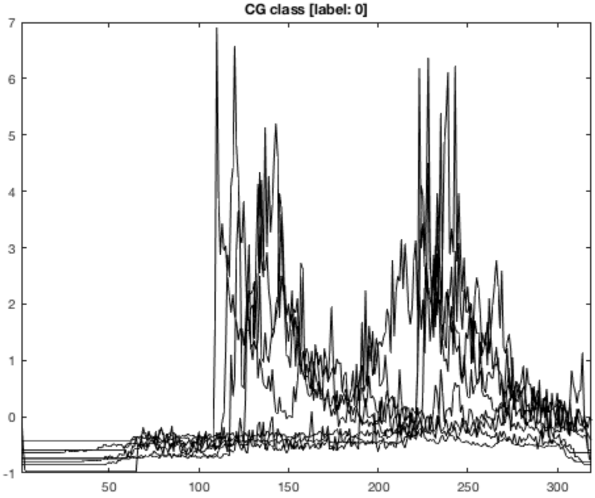
Classes

1. CG (Positive Initial Return Stroke)
2. IR (Negative Initial Return Stroke)
3. SR (Subsequent Negative Return Stroke)
4. I (Impulsive Event)
5. I2 (Impulsive Event Pair)
6. KM (Gradual Intra-Cloud Stroke)
7. O (Off-record) (special case)

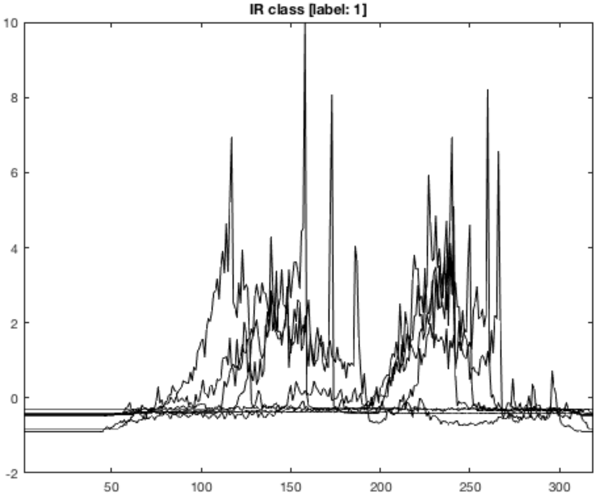
Visualization of Classes

Label	CG	IR	SR	I	I2	KM	O
Name	Positive Initial Return Stroke	Negative Initial Return Stroke	Subsequent Negative Return Stroke	Impulsive Event	Impulsive Event Pair	Gradual Intra-Cloud Stroke	Off-Record
# Samples	18	17	14	19	15	38	22
Ex. Time Series							

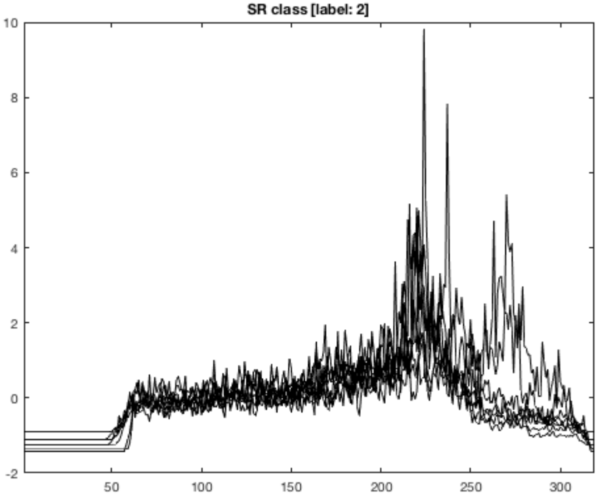
0. CG (Positive Initial Return Stroke)



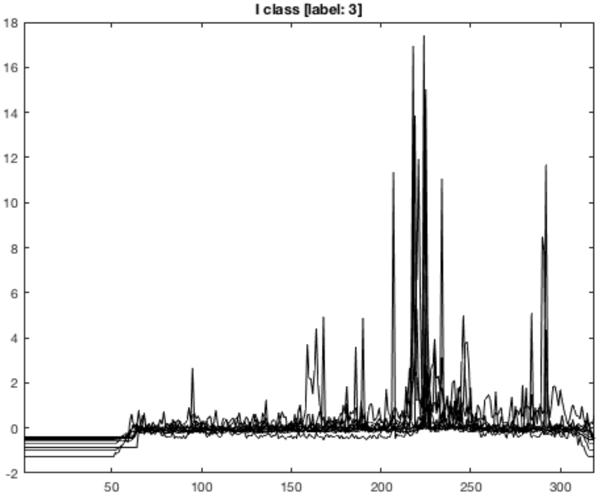
1. IR (Negative Initial Return Stroke)



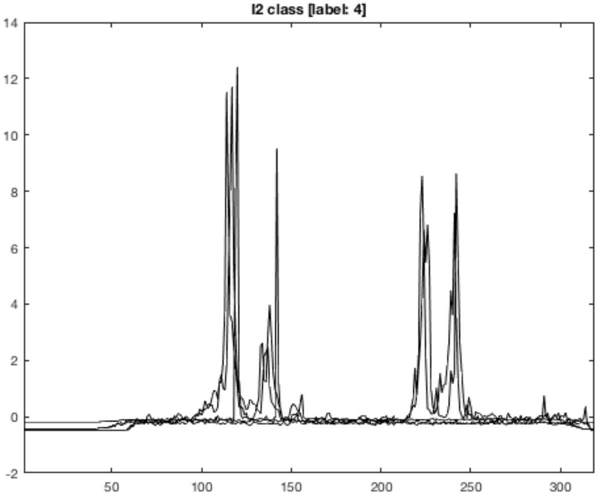
2. SR (Subsequent Negative Return Stroke)



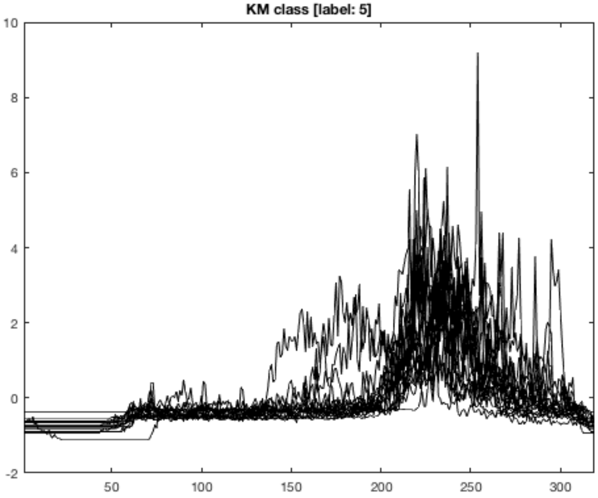
3. I (Impulsive Event)



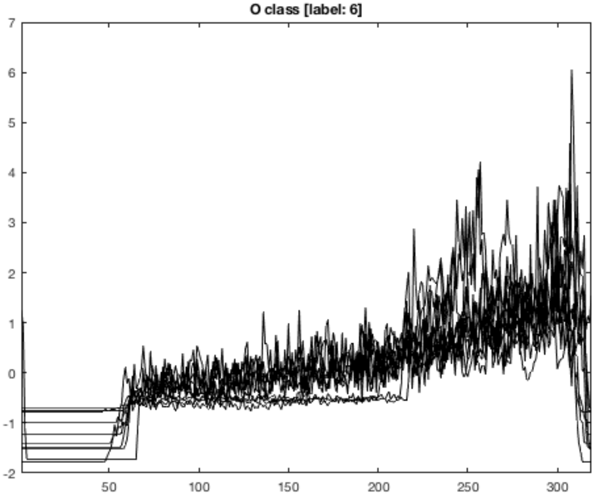
4. I2 (Impulsive Event Pair)



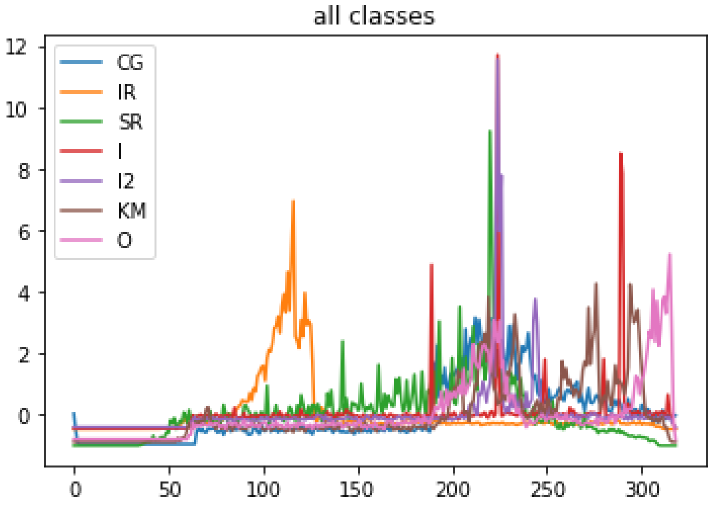
5. KM (Gradual Intra-Cloud Stroke)



6. O (Off-record) (special case)



All Classes



The Problem

- Imbalanced data
- Unequal class representation
- Curse of dimensionality
- Distance Metric

Topic

Naïve Bayes

- Essence of Naïve Bayes $\rightarrow$ Given multiple pieces of evidence, treat each piece as independent.
- $P(outcome|evidence) = \frac{P(likelihood\_of\_evidence)*(Prior)}{P(evidence)}$
- Our job is to look at the evidence, to consider how likely it is to be this class or that class, and assign a label to each entity.
- The class that has the highest probability is declared the "winner" and that class label gets assigned to that combination of evidences.
- Gaussian Naïve Bayes $\rightarrow$ assumes intra-class time series means are normally distributed.
- With a (default) train-test split of 70-73, we achieved an accuracy of 0.573.

K-Nearest-Neighbors

- The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.
- In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point.

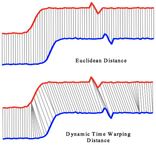


Figure 1: Note that while the two time series have an overall similar shape, they are not aligned in the time axis. Euclidean distance, which assumes the i^{th} point in one sequence is aligned with the i^{th} point in the other, will produce a pessimistic dissimilarity measure. The non-linear Dynamic Time Warped alignment allows a more intuitive distance measure to be calculated.

- Distance Metric: Euclidean vs DTW
- LBKeogh to battle computational complexity

1NN Performance (TTS: 70-73)

	precision	recall	f1-score	support
0.0	0.75	0.60	0.67	10
1.0	0.20	0.11	0.14	9
2.0	0.56	0.83	0.67	6
3.0	0.25	0.14	0.18	7
4.0	0.50	0.30	0.37	10
5.0	0.59	0.84	0.70	19
6.0	0.71	0.83	0.77	12
avg / total	0.54	0.58	0.54	73

	precision	recall	f1-score	support
0.0	0.73	0.80	0.76	10
1.0	0.80	0.44	0.57	9
2.0	0.50	0.67	0.57	6
3.0	0.46	0.86	0.60	7
4.0	1.00	0.20	0.33	10
5.0	0.77	0.89	0.83	19
6.0	0.92	0.92	0.92	12
avg / total	0.77	0.71	0.69	73

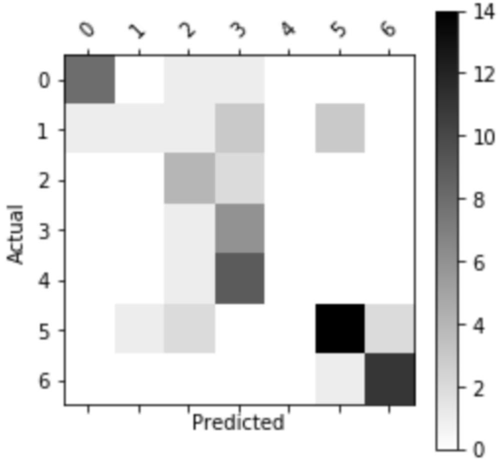
Topic

SVM

- Given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples
- Kernel function: linear
- Train-test split: 70-73
- Correct classification rate: 0.6027

SVM Performance

```
{0: 8, 1: 1, 2: 4, 3: 6, 4: 0, 5: 14, 6: 11}
```



CNN

- Based on model used to effectively classify human activity time series data
- Intuition for CNN's on images carry over to time-series. The main difference in the code is the stride argument we pass to the conv-layer.
- We want the kernel to stride along the time-series, but not along the second dimension that we would have used for images.
- Architecture: 3 Convolutional Layers, with Batch Normalization between and SoftMax activation functions in fully-connected layers
- Train-test split: 70-73
- Accuracy on validation data over 3000 iterations: 0.639

RNN

- Class of artificial neural network where connections between units form a directed graph along a sequence.
- Can use their internal state (memory) to process sequences of inputs.
- LSTM $\rightarrow$ cell is responsible for "remembering" values over arbitrary time intervals
- Train-test split: 70-73
- Accuracy: 0.8904

Topic

Ranking

1) LSTM-RNN	(0.8904)
2) 1NN with DTW Distance Metric	(0.7700)
3) CNN	(0.6390)
4) SVM	(0.6027)
5) Naïve Bayes	(0.5730)

Citations

- http://timewarping.org/TSCU/doc/html/tscu_tutorial.html
- http://robromijnders.github.io/CNN_tsc/
- https://github.com/titu1994/LSTM-FCN
- Damian R. Eads, Daniel Hill, Sean Davis, Simon J. Perkins, Junshui Ma, Reid B. Porter, James P. Theiler, "Genetic Algorithms and Support Vector Machines for Time Series Classification," Proc. SPIE 4787, Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation V, (6 December 2002);
- http://alexminnaar.com/time-series-classification-and-clustering-with-python.html
- http://timeseriesclassification.com/description.php?Dataset=Lightning7