Rule Learning from Data Streams: an overview

Jesús S. Aguilar–Ruiz

School of Engineering
Pablo de Olavide University
Seville, Spain

Abstract. Classification is a very well studied task in data mining. In the last years, important works have been published to scale up classification algorithms in order to handle large datasets. However, due to the high rate of streams of data, a number of emerging applications are demanding new approaches. Rule learning is an efficient alternative to address non–stationary environments. The talk presents an overview of rule–based learning algorithms for data streams and emphasizes some important aspects of these techniques.

Keywords: Data Streams, Rule–based learning.

1 Introduction

The advances in hardware technology have paved the way for the development of algorithms that can process the real–time information at a rapid rate. Streams of data grow at an unlimited rate and traditional data mining algorithms cannot process them efficiently. In spite of the great increase of storage capacity, it is not even enough for hundreds or thousands of instances arriving per second. Nowadays, typical problems such as clustering, classification, frequent pattern mining, change detection or dimensionality reduction are being reconsidered in the realm of data streams. What was initially finite data is now infinite data, thus giving rise to many challenges in machine learning, data mining and statistics.

Classification and rule learning are important, well studied tasks in machine learning and data mining. Classification methods represent the set of supervised learning techniques where a target categorical variable is predicted based on a set of numerical or categorical input variables. A variety of methods such as decision trees, rule based methods, and neural networks are used for the classification problem. Most of these techniques have been designed to build classification models from static data sets, where several passes over the stored data are possible.

In order to classify and model large–scale databases, important works have been recently addressed to scale up inductive classifiers and learning algorithms. In environments where high–rate streams of detailed data are constantly generated, memory and time limitations make multi pass scalable algorithms unfeasible. Also, real–world data streams are not generated in stationary environments, requiring incremental learning approaches to track trends and adapt to changes in the target concept.

Furthermore, the classification process may require simultaneous model construction and testing in an environment which constantly evolves over time. However, within incremental learning, a whole training set is not available a priori as examples arrives over time, normally one at a time t and not time dependent necessarily (e.g., time series). Despite online learning systems continuously review, update, and improve the model, not every online system is based on an incremental approach.

2 Some aspects of learning from data streams

Formally, a data stream D can be defined as a sequence of examples (also called transactions or instances), $D=(e_1,e_2,\ldots,e_i,\ldots)$, where e_i is the i-th arrived example. To process and mine data streams, different window models are often used. A window is a subsequence between the i-th and j-th arrived examples, denoted as W[i,j]=$(e_i,e_{i+1},\ldots,e_j)$, $i < j$. There are three common models:

- Landmark window. The model computes examples from a starting e_i to the current e_t. If $i = 1$, then the model processes the entire data streams.
- Sliding window. This model is based on the size of the windows w, and then computes all the examples of the subsequence W[t-w+1,t], where e_t is the current example.
- Damped window. This is a variant that can be applied to the aforementioned models, as it assigns weights to examples in order to give more importance to more recent examples.

Ideally, a rule based system should not use *sampling*, and the window size should be one. Some important aspects of the design of incremental algorithms for data streams are:

- **Influence of the arriving:** both the order of examples, the arriving rate and the possible time–dependency are critical.
- **Adaptation to change:** knowledge obtained may not be useful in the future, so any change of tendency must be detected.
- **Learning curve:** at initial stages, the incremental model is not accurate enough.

3 Rule–based systems

In general, rule based classification algorithms for data streams are quite difficult to build, particularly when the window size is one, as the necessary statistics to maintain the performance are hard to keep and update. This factor is even more critical when concept drift is present in data.

3.1 AQ11 Algorithm

The first incremental algorithm based on set of rules, which are described in DNF. AQ11 does not store examples in memory, and relies on the learning process, i.e., on the decision rules obtained. For any new example covered by a rule with a different label from the one of the example, AQ11 refines such rules iteratively, until the wrongly classified example is not covered. Rules are newly generalized from new examples.

3.2 GEM Algorithm

The learning scheme is very similar to AQ11, although each process of generalization or specialization takes into account every example arrived until that moment. In practice, it is not an useful technique.

3.3 STAGGER Algorithm

The first incremental algorithm that provides decision rules, and designed to be robust in the presence of noise and able to deal with concept drift. STAGGER does not update the rules constantly, but only when there is a high level of inconsistency. STAGGER saves statistics instead of examples, and adopts a conservative policy in order to detect changes in the long time.

3.4 SCALLOP Algorithm

SCALLOP is a scalable classification algorithm for numerical data streams. The algorithm produces a set of decision rules that is constantly verified during the learning process, in order to maintain the tendency of data. The verification is a process that involves modification, updating and deleting of statistics to handle concept drift properly. One of the main features of this system is the use of several sliding windows.

4 Conclusions

The development of learning or classification systems for data streams is a very difficult task. There are many temporal and spatial constraints and the algorithm must provide accurate response at any time. Furthermore, if it is required the system to be descriptive, as decision rules for instance, then the level of complexity increases, since the updating of descriptive models takes longer.

References

1. Maloof, M., Michalski, R.: Incremental learning Artificial Intelligence 154 (2004).
2. Muthukrishnan, S.: Data streams: algorithms and applications. In: Proc. of The 14th annual ACMSIAM Symposium on Discrete Algorithms. (2003).
3. Jin, R., Agrawal, G.: Efficient decision tree construction on streaming data. In Proc. of The 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining KDD03, ACM Press (2003).
4. Wang, H., Fan, W., Yu, P., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In Proc. of The 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining KDD03, ACM Press (2003).