Novelty Detection from Data Streams

João Gama
LIAAD-INESC Porto,
University of Porto, Portugal
jgama@fep.up.pt
Overview

- Data Streams
- Histograms
- Change Detection
- Clustering
- Decision Trees
- Rules Learning
- Evaluation
- Multiple Models
- Frequent Patterns
- Time Series
- **Novelty Detection**
Novelty Detection

One-Class Classification

Density Methods

Cluster-based novelty detection

Illustrative Example
Outline

1. Novelty Detection
2. One-Class Classification
3. Density Methods
4. Cluster-based novelty detection
5. Illustrative Example
Definition

- Novelty Detection refers to the automatic identification of unforeseen phenomena embedded in a large amount of normal data.
- *Novelty* is a relative concept with regard to our current knowledge:
  - It must be defined in the context of a representation of our current knowledge.
- Specially useful when novel concepts represent abnormal or unexpected conditions
  - Expensive to obtain abnormal examples
  - Probably impossible to simulate all possible abnormal conditions
Context

- In real problems, as time goes by
  - The distribution of known concepts may change
  - New concepts may appear

- By monitoring the data stream, emerging concepts may be discovered

- Emerging concepts may represent
  - An extension to a known concept (Extension)
  - A novel concept (Novelty)

- Several interesting applications: Early Detection of Fault in Jet Engines, Intrusion Detection in computer networks, Breaking News in a flow of text documents (news articles), Burst of Gamma-ray (astronomical data),
Perspective

- Stable state, nothing is learned
- Disturbance, something Unknown
- Reasoning for an explanation: Learning

Awareness vs. Time Graph
Desiderata for Novelty Detection

- **Principle of robustness and trade-off**: A novelty detection method must be capable of robust performance on test data that maximizes the exclusion of novel samples while minimizing the exclusion of known samples. This trade-off should be, to a limited extent, predictable and under experimental control.

- **Principle of generalization**: The system should be able to generalise without confusing generalised information as novel (Tax and Duin, 1998).

- **Principle of independence**: The novelty detection method should be independent of the number of features, and classes available. It should also show reasonable performance in the context of imbalanced dataset, low number of samples, and noise.

- **Principle of uniform data scaling**: In order to assist novelty detection, it should be possible that all test data and training data after normalisation lie within the same range (Singh and Markou, 2003).
Desiderata for Novelty Detection

• **Principle of adaptability**: A system that recognises novel samples during test should be able to use this information for retraining (Saunders and Gero, 2000).

• **Principle of computational complexity**: A number of novelty detection applications are online and, therefore, the computational complexity of a novelty detection mechanism should be as low as possible.

• **Principle of parameter minimization**: A novelty detection method should aim to minimise the number of parameters that are set by the user.
Illustrative Applications

- Fault Detection of Fault in Jet Engines
  Hayton et al., 2000, using SVMs (nu-SVC)
Illustrative Applications

- Fault Detection
- Image analysis for cancer identification in mammograms
  Tarassenko et al., 1995, using Parzen Windows
- Video-surveillance
Illustrative Applications

- Fault Detection
- Image analysis
- Video-surveillance
- Intrusion Detection in computer networks
  Yeung e Chow, 2002, using Parzen Windows
Illustrative Applications

- Fault Detection
- Video-surveillance
- Intrusion Detection
- Robot Navigation, to recognize changes in the environment; Marsland, 2002, using Neural Networks
Approaches for Novelty Detection in DS

- **One-class classification**
  - Model knowledge about a single profile
  - New examples may be identified as members of that profile or not

- **Methods based on Frequencies**
  A pattern is *surprising if the frequency of its occurrence differs substantially from that expected by chance, given the previously seen data.* (TARZAN; Keogh et al., 2002)

- **Methods based on decision structure**
  Considers decisions taken by each unit in a decision structure. In a stable state, the contribution of each unit is likely to remain constant. Changes in the participation of decision units may indicate a conceptual change
Outline

1. Novelty Detection
2. One-Class Classification
3. Density Methods
4. Cluster-based novelty detection
5. Illustrative Example
Autoassociator Networks

Concept-learning in the absence of counter-examples: an autoassociation-based approach Nathalie Japcowicz, 1999

- Three layer network
- The nr. of neurons in the output layer is equal to the input layer
- Train the network such that $\tilde{y}$ is equal to the $\tilde{x}$
- The network is trained to reproduce the input at the output layer
Autoassociator Networks

To classify a test example $\vec{x}$

- Propagate $\vec{x}$ through the network and let $\vec{y}$ be the corresponding output;
- If $\sum_{i=1}^{k} (x_i - y_i)^2 < \text{Threshold}$ Then the example is considered from class normal;
- Otherwise, $\vec{x}$ is a counter-example of the normal class.
Nearest Neighbour: Data Description

Nearest neighbour for novelty detection (Tax, 2001)

If $d_1 / d_2 > 1 \rightarrow$ reject
Outline

1 Novelty Detection
2 One-Class Classification
3 Density Methods
4 Cluster-based novelty detection
5 Illustrative Example
Density Methods

*Novelty Detection in Data Streams: A Small Step Towards Anticipating Strategic Surprise* C. Gazen, J. Carbonell

Density function is defined as:

\[ f(r) = \frac{dM(r)}{dV(r)} \]

where

- \( M(r) \) is the number of points within a sphere of radius \( r \) and
- \( V(r) \) is the volume of that sphere.

*To detect changes in the shape and density of clusters, we analyze the density of points within a cluster as a function of the distance to the cluster’s centroid.*
Density Methods

Figure 1 – Changes in density functions for some events
Outline

1. Novelty Detection
2. One-Class Classification
3. Density Methods
4. Cluster-based novelty detection
5. Illustrative Example
Cluster-based novelty detection

*Cluster-based novelty detection*, Spinoza, Carvalho, Gama, SAC 08.

**Initial Phase: Supervised, batch mode**

Start by modeling the normal condition. Learns a partial model about what is known. Based on a set of classified examples.
Cluster-based novelty detection

Cluster-based novelty detection, Spinoza, Carvalho, Gama, SAC 08.

Initial Phase: Supervised, batch mode

Start by modeling the normal condition.
Learns a partial model about what is known.
Based on a set of classified examples.

Second Phase: Process stream of unlabelled examples

For each incoming example:
- If it is explained by the current model: classify the example and discard
- If it is not explained: Store in a short-term memory
- Time to Time
  - Find clusters in the examples stored in the Short Term Memory
  - Clusters far away from existing ones: Novel concept.
  - Clusters closed to existing ones: Extend known concepts.
Perspective

- Unknown
  - Noise
  - Novelty
  - Concept drift

Sparse
Perspective

- **Unknown**
  - Noise
  - Novelty
  - Concept drift

- Sparse
- Dense
Perspective

- Unknown

- Noise
- Novelty
- Concept drift

Sparse

Dense, Dissimilar to what is known

Dense, Similar to what is known
Initial Phase - Generate Initial Model

![Diagram showing initial model generation with clusters labeled as Normal.](image-url)
Second-Phase: Process Stream Examples
Second-Phase: Look for new Concepts
Outline

1. Novelty Detection
2. One-Class Classification
3. Density Methods
4. Cluster-based novelty detection
5. Illustrative Example
Metrics

- Error rates
  - False-Normal
  - False-Unknown

- Distribution of profiles
  - Identified as Normal
  - Valid cluster of Extension of Normal Concept
  - Identified as Extension
  - Valid cluster of Novelty
  - Identified as Novelty
  - Unknown
Iris

- Iris dataset
- Iris-versicolor as Normal
- Discovery of Iris-setosa
- Discovery of Iris-virginica
- Extension of Iris-versicolor

Identified as Normal
Valid cluster of Extension
Identified as Extension
Valid cluster of Novelty
Identified as Novelty
Unknown
As new concepts are learned, the system is capable of explaining future examples of such concepts (dark orange / dark green).
Breast Cancer

- **Breast Cancer Wisconsin dataset**
- **Benign as Normal**
- **Benign examples correctly identified as Normal**
- **Benign concept extended**
- **Malignant concept identified as Novelty**

Identified as Normal
Valid cluster of Extension
Identified as Extension
Valid cluster of Novelty
Identified as Novelty
Unknown
OLLINDA: Main Algorithm

INIT: Generate Initial Model (Supervised, Batch)

Process Stream:

- Read Next Example
- If the current model explains the example
  - Classify the Example
- Otherwise: Store the example in a short-term-memory
- Time to Time
  - Cluster the examples in the short-term-memory.
  - For each New Cluster
    - New Concept: If the New Cluster is far away from the current Model
    - Merge the New Cluster with the closest existing cluster
Conclusions

- Having acquired knowledge solely about a Normal behavior (learning from positive-only examples), the proposed approach is able to identify emerging clusters that represent
  - Novel concepts
  - Extension of the Normal concept
Bibliography on Novelty Detection

- **Online Novelty Detection on Temporal Sequences**, by Junshui Ma, Simon Perkins, in the ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2003.
- **Efficient Elastic Burst Detection in Data Streams**, by Yunyue Zhu, Dennis Shasha, in the ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD) 2003.
- **Active Mining of Data Streams**, by Wei Fan, Yi-an Huang, Haixun Wang, Philip S Yu, in the SIAM International Conference on Data Mining (SIAM DM) 2004.
- **OLINDDA: a cluster-based approach for detecting novelty and concept drift in data streams**; E. Spinosa, A. Carvalho, J. Gama; SAC 2007: 448-452
Bibliography on Novelty Detection


- Karl-Michael Schneider, *Learning to Filter Junk EMail from Positive and Unlabeled Examples*

- Sameer Singh, Markos Markou, John Haddon; *Detection of New Image Objects in Video Sequences Using Neural Networks*; 1999