Novelty Detection

Meelis Kull

Apr 22, 2008

Presentation prepared and presented in the Machine Learning course lead by Phaedra Agius in University of Tartu
Outline

• Novelty detection
• Applications
• Statistical approaches
• Neural network approaches

• A Linear Programming Approach to Novelty Detection

Colin Campbell
Dept. of Engineering Mathematics,
Bristol University, Bristol BS8 1TR,
United Kingdom
C.Campbell@bris.ac.uk

Kristin P. Bennett
Dept. of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, New York 12180-3590
United States
bennek@rpi.edu
Used materials for novelty detection overview


Novelty detection: a review—part 1: statistical approaches

Novelty detection: a review—part 2: neural network based approaches

Markos Markou*, Sameer Singh
PANN Research, University of Exeter, UK
Novelty detection

Novelty detection is the identification of new or unknown data or signal that a machine learning system is not aware of during training.
Example: Fault detection

Training data:

Test data:

known  known  known  known  novel

This is one-class classifier, as all training data belong to the same class
Applications of one-class classifier

- Intrusion detection
- Jet engine fault detection
- Heart monitoring
- *i.e.* any system monitoring
Example: Face recognition

Training data:

Test data:
Example: Face recognition (1)

Training data:

Test data:

No similarity discovered, all novel
Example: Face recognition (2)

Training data:

Test data:

known  known  known  known  known  known  known

All images have skin-colored region in the middle
Example: Face recognition (3)

Training data:

Test data:

Only one test image has similar colors as test data
Example: Face recognition (4)

Training data:

Test data:

known  known  known  known  known  known  novel

Only one test image does not represent a human face
Example: Face recognition (5)

**Training data:**

![Images of four people with labels: known, known, known, known]

**Test data:**

![Images of four people with labels: known, known, novel, novel]

Four test images represent people known from training data.
Applications of multi-class novelty detection

• Face recognition
• Biometrics
• Hand-written digit recognition
• Robotics
Main problems to solve

• System must build a model of main data in order to recognize novel data
• System should be able to generalize without confusing generalized information as novel
Approaches (1)

• Statistical – statistical model of data
  – Construct a density function of the known data, report as novel if too improbable in this model – e.g. GMM (Gaussian mixture modeling; parametric)
  – Sequential patterns - HMM – Hidden Markov modeling (parametric) – intrusion detection!
  – Cluster known data, report as novel if does not belong to any of known clusters (non-parametric)
Approaches (2)

• Neural network based approaches
  – Multi-layer perceptron – e.g. each neuron in output layer corresponds to one class, if zero or more than one neurons fire, then reported as novel
  – ART, CNN, ONN, RBF, SOM, Neural tree
  – SVM-type approaches
    • Find minimal sphere that encompasses almost all points in the data set
    • Find hyper-plane maximally distant from origin with all data points lying on the opposite side from the origin
A Linear Programming Approach to Novelty Detection

Colin Campbell
Dept. of Engineering Mathematics,
Bristol University, Bristol
Bristol, BS8 1TR,
United Kingdom
C.Campbell@bris.ac.uk

Kristin P. Bennett
Dept. of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, New York 12180-3590
United States
bennek@rpi.edu
Algorithm – hard margin

Input space:

Feature space:

\[
\begin{align*}
\min_{\alpha_j, b} & \quad \sum_{i=1}^{m} \left( \sum_{j=1}^{m} \alpha_j K(x_i, x_j) + b \right) \\
\text{s.t.} & \quad \sum_{j=1}^{m} \alpha_j K(x_i, x_j) + b \geq 0 \\
& \quad \sum_{i=1}^{m} \alpha_i = 1, \quad \alpha_i \geq 0
\end{align*}
\]
Algorithm – soft margin

Input space:

Feature space:

\[
\min_{\alpha_j, b, \xi_i} \sum_{i=1}^{m} \left( \sum_{j=1}^{m} \alpha_j K(x_i, x_j) + b \right) + \lambda \sum_{i=1}^{m} \xi_i \\
\text{s.t.} \quad \sum_{j=1}^{m} \alpha_j K(x_i, x_j) + b \geq -\xi_i, \quad \xi_i \geq 0 \\
\sum_{i=1}^{m} \alpha_i = 1, \quad \alpha_i \geq 0
\]
Experiments

• Artificial dataset 1: hard margin, RBF kernels,

\[ e^{-\frac{||x_i - x_j||^2}{2\sigma^2}} \]

\( \sigma = 0.2 \)
Experiments

• Artificial dataset 2: soft margin, RBF kernels,

\[ e^{-\frac{||x_i - x_j||^2}{2\sigma^2}} \]

\[ \sigma = 0.2 \]
Experiments

- Artificial dataset 3: modified RBF kernels,

\[ e^{-|\mathbf{x}_i - \mathbf{x}_j|/2\sigma^2} \]

\[ \sigma = 0.5 \]
Experiments

- Biological dataset: 194 blood samples, 4 attributes measured in each
- 127 normal, 67 rare genetic disease
- 100 normal – training data
- 27 normal, 67 disease – test data
- Hard margin RBF
Experiments

error rate

abnormal

normal

$\sigma$
Experiments

• Engineering dataset: Ball-bearings
• Normal and 4 types of abnormalities
• 913 normal for training
• 913 normal, 747 type 1, 996 type 2 for validation
• Result: $\sigma = 320.0$
• 913 normal, 747 type 1, 996 type 2, 996 type 3, 996 type 4 – test data
• 98.7% normal, 100% type 1, 53.3% type 2, 28.3% type 3, 25.5% type 4
Summary

• Novelty detection is applied for one-class or multi-class classification where there exists an option “None of the above” / reject / novel / etc
• There are many methods for performing novelty detection, things to look at while choosing a suitable one:
  – Speed
  – One-class vs multi-class
  – Efficiency / transparency of the decision function