Brief overview of novelty detection methods for robotic cumulative learning [⋆]

E. Kerr, Y. Gatsoulis, N. H. Siddique, J. V. Condell and T. M. McGinnity
`[ep.kerr, i.gatsoulis, nh.siddique, j.condell, tm.mcginnity]@ulster.ac.uk`

Intelligent Systems Research Centre, Faculty of Computing and Engineering,
University of Ulster, Magee College, Northland Road, Derry, BT48 7JL, NI

Abstract. Novelty detection is a key aspect of cumulative learning and is an important issue for the effective and long-term operation of intelligent robots designed to learn, evolve and make decisions based on their cumulative knowledge and experience. Our approach to novelty detection is from the perspective that the robot ignores perceptions that are already known, but is able to identify anything different.

As it is such an important aspect of any cumulative learning agent, the topic has drawn considerable attention over the last decade and there has been a large number of recent surveys published. This paper discusses some key characteristics for the effective operation of novelty detection filters in cumulative learning robot tasks. It also presents a brief overview of the most recent surveys, and describes in more detail some novelty detection methods that appear to meet some of the key characteristics, and have the potential to be efficient novelty detectors. The work presented here is part of an EU project. A brief description of the project is given.

1 Introduction

Novelty detection is an important factor for the effective and long-term operation of intelligent robots designed to explore, evolve and make decisions based on their cumulative knowledge and experience. Although there is no strict definition of novelty detection it is widely regarded [10, 12, 18] as the process of identifying novel stimuli that are 'different from anything known before; new, interesting and often seeming slightly strange' (definition of 'novel' [20]).

Based on this description, we approach novelty detection from the perspective that the robot ignores perceptions that are similar to those seen during training, but is able to highlight anything different. In this sense, novelty detection can be seen as a form of negative learning: examples are provided of those features that should not be detected, and the novelty filter aims to highlight anything that differs from the inputs used in training. As such, novelty detection is the ability to identify perceptions that have never been experienced before and constitutes an important component for the effective and long-term operation of intelligent

---

[⋆] This work was supported by EU FP7-ICT-2007-3: 2.2 Cognitive Systems, Interaction and Robotics PROJECT: IM-CLeVeR (2009-2013).

robot systems allowing computationally-efficient, unsupervised and incremental exploration and learning of new skills and environments.

This paper initially discusses some individual characteristics and key differences of novelty detection, particularly when used in cumulative learning tasks, from the often synonymous terms of outlier or anomaly detection. A short review of recent surveys in the field is then presented. Considering the page limits, we prefer to present the strengths and weaknesses of the general types of novelty detection methods as categorised by those surveys reviewed here. In this way, it is hoped that the reader can quickly gain a general overview of the current state of the art in the field. Two specific case study methods that meet some of the key characteristics are presented in more detail.

2 Key characteristics and open challenges

A number of novelty detection methods have been proposed in the literature, mainly focusing on detecting anomalies, i.e. identifying patterns that do not conform to expected behaviour, in application domains such as diagnosing medical problems, detecting machine faults, network security, surveillance, fraud detection, etc. [4, 9]. Typically for these problems there is substantial data about the normal classes (e.g. healthy subjects, normal operation of a machine, etc.), but very little data displaying the features that should be detected. Hence, it is common to learn a model of the normal dataset and then attempt to detect deviations from this model. This means that a training set of normal data is sufficient for the effective operation of a system. If there is a need for further expansion then this is easier done in batch processing, which also allows the complete retraining of the system.

As a key characteristic of cumulative learning is that previously gained knowledge can be used to recognise and solve future problems, that will arise in a dynamic system, such an approach to novelty detection therefore seems unsuitable, for this application domain, for a number of reasons. Firstly, offline learning might be inappropriate in cases where the learning task depends on a sequence of prior learnt actions; an online learning filter is then needed instead. Another reason is that retraining from scratch is inefficient, undesirable and is contrary to the whole purpose of cumulative learning. Furthermore, the notion of normal data as expressed in anomaly detection is not the same in novelty detection, particularly in the domain of cumulative learning; a novelty is not necessarily an anomaly, and as such there is no initial training dataset that can be complete. From a practical point of view, a novelty detector has to be able to reliably identify the novel perceptions that are distorted due to the noisy sensor readings of the robot, as well as operate online if an immediate response is required. In cases where a learnt action determines the next one, it is also necessary for the novelty detector to operate and learn online. Such a novelty detector requires learning architectures that support dynamic and incremental expansion of knowledge representation.

Despite these fundamental differences between novelty detection and outlier or anomaly detection, the methods used in each are often common. The next section provides a brief summary of the current state of the art of novelty detection methods, summarising a number of recent surveys in the field.

3 Overview of reviews of novelty detectors

A number of extensive surveys for novelty, anomaly and outlier detection have been published recently [4, 3, 9–12], and for this reason a comprehensive literature review of all novelty detection methods is currently unnecessary. Instead, this section summarizes these reviews as a general reference. The surveys presented here are recent general reviews of novelty detection methods over a broad range of application domains, rather than specific reviews focusing in one particular domain only (e.g. [1, 8, 21, 17, 2]). The emphasis is on the different categorisation schemes of novelty detection methods (see Table 1). Such a summary would allow an easier comparison of the strengths and weaknesses of each method. The surveys are grouped in reviews of methods for novelty detection, reviews of methods for outlier detection and reviews of methods for anomaly detection. Although there are fundamental differences between them, as already explained in Section 2, authors often use the terms novelty detection, outlier detection and anomaly detection interchangeably [4, 9, 12], as the methods used are usually common.

Table 1. Categories of novelty detection methods according to the surveys presented. 1: [10], 2: [11], 3: [12], 4: [9], 5: [4]. Part of this table is adopted from [4].

Categories	**1**	**2**	**3**	**4**	**5**
Classification based		✓	✓	✓	✓
Clustering based			✓	✓	✓
Nearest neighbour based			✓	✓	✓
Statistical	✓		✓	✓	✓
Information theoretic					✓
Spectral			✓		✓

3.1 Surveys in novelty detection

In the past, authors have distinguished between two main categories of novelty detection methods: statistical approaches and neural network based approaches [10, 11]. Statistical based methods typically test whether the test samples come from the same distribution as the one learnt from the training data. Their main strength is that they are transparent methods, meaning they can be analysed easily using standard numerical methods. Their main drawback is that they make a number of assumptions, e.g. parametric methods assume that the data distributions are Gaussian, which restrict their analytical power. In contrast, neural network based approaches make no prior assumptions on the form of the data distributions, requiring only the optimisation of a small number of training parameters [11]. However, they cannot be as easily retrained as statistical models. This is not a major issue in cumulative learning problems, where adaptation and expandability of the learning structures are more important factors. In one survey [11], three general algorithmic approaches are identified: regularization algorithms, pruning algorithms and constructive algorithms. Constructive algorithms are usually preferred, because they start with smaller networks that are faster to train, and expand as necessary. Furthermore, it is more difficult to decide on how big the initial network should be with pruning algorithms. Two important issues that a designer has to deal with when working with constructive algorithms are: 1) what is the most effective training method to enable integrating new units to the existing network, and, 2) when to stop adding new units.

Another extensive summary of novelty detection methods is provided in [12]. Although the methods are organised in specific categories based on their main algorithmic approach, rather than in generic categories that would allow an easy and structured comparison between them, the paper offers a concise description of each of the methods it presents. Most importantly, some important issues with novelty detection are emphasized, these being how different a stimulus should be before it is classified as novel, and how often a novel stimulus must be seen before it stops being novel, a question that the authors have investigated in their earlier work [13].

3.2 Surveys in outlier detection

An interchangeable synonym to novelty detection is outlier detection, and an extensive survey from this perspective is presented in [9]. The authors treat outlier detection based on the following definition of an outlier [7]: “an outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs”.

Three types of approaches are distinguished. The first type includes methods where the outliers are determined without any prior knowledge of the data, this being analogous to unsupervised clustering. In the second type, both normal and abnormal cases are modelled; the authors say that such an approach is analogous to supervised classification, and requires pre-labelled data tagged as normal or abnormal. Finally, methods of the third type mainly model normal cases together with a few abnormal cases also. It has been mentioned that this third type is generally acknowledged as novelty detection or novelty recognition, and it is analogous to a semi-supervised detection/recognition task [9]. Like in [10, 11], Hodge and Austin [9] also distinguish the methods used for novelty detection in statistical models and neural networks, and in addition to these two categories they suggest another two, these being machine learning methods and hybrid methods. Their critique to the statistical and neural network based approaches is similar to that in [10, 11], i.e. statistical models are easy to deploy and are transparent, but their computational demands increase if complex data transformations are needed before processing, while neural networks can generalise and learn complex class boundaries well, but they are opaque with respect to the underlying mechanisms and typically require longer training and tuning sessions. Because statistical based approaches and neural networks typically require vectors consisting of at least ordinal data, the machine learning category was suggested to cover methods capable of also dealing with categorical data, such as rule-based systems, further clustering methods, decision trees and other tree-structure based methods. The hybrid category includes the most recent developments in the field of outlier detection, and covers those methods that incorporate algorithms from the other three categories. This is done so in order to gain the benefits of each one of the methods used while eliminating their limitations. For these reasons these methods seem attractive, but Hodge and Austin warn that the use of more than one classifier, unless carefully designed, can lead to the exhaustion of the computational resources available.

3.3 Surveys in anomaly detection

The most recent comprehensive survey on novelty detection methods was written by Chandola et al [4]. The focus is on anomaly detection, i.e. "patterns in data that do not conform to a well defined notion of normal behaviour". [3, p. 15:2] Although such a notion makes a clear distinction between normal and abnormal data, the authors acknowledge that anomaly detection is related to the domain of novelty detection, as methods used in one area are often used in the other area also, and vice versa. The authors group the different methods into six categories, these being the same four as the ones presented in [9, 1], i.e. classification based methods, clustering based methods, nearest neighbour based methods and statistical methods, and two additional ones, these being information theoretic methods and spectral methods. An interesting suggestion is that anomalies can be classified in three categories. The first category includes point anomalies, which occur when an individual data instance is considered anomalous with respect to the rest of the data. The second category is contextual anomalies, which include data instances that are considered anomalous in a particular context but not otherwise; this type of anomaly is often investigated in time-series data. Finally, collective anomalies consist of sets of data instances that are anomalous with respect to the entire data set. A further agreement between the surveys of [9] and [4] is on the general approaches of the detection methods, these being supervised, semi-supervised and unsupervised.

The authors identify a number of challenges that exist in anomaly detection, but two of them are of particular interest to novelty detection as well. First, there is the issue of the boundaries between normal data and anomalies, or in terms of novelty detection, the boundaries between known data and novelties; an issue acknowledged also by other researchers [12].

The second issue is that of noise in the data. As the authors mention, noise tends to appear similar to actual anomalies, and hence, it creates serious difficulties in identifying the true novelties.

4 Case study methods of novelty detection

Two novelty detection methods are presented in more detail as they are of particular interest as they have fulfil of the foregoing key characteristics discussed in Section 2.

4.1 Kohonen self-organizing map with habituated synapses

A novelty filter that has the purpose of learning to recognize known features and evaluate their novelty based on the frequency with which these input stimuli have been seen recently was presented in [13]. It is based on a Kohonen map with habituated synapses linking the nodes of the network to an output node. A habituated synapse decreases in strength as its connected nodes fire, and increases in strength when already known stimuli are not seen for some time. The behavioural phenomenon of habituation has its roots in biology, and as mentioned by [13], cross-citing [19], it is thought to be one of the simplest forms of plasticity in the brain of a large number of organisms.

Two sets of experiments were conducted to test the ability of the novelty filter to learn a representation model of an operational environment, and to investigate the effects of the habituated synapses on novelty detection. The experimental scenario consisted of a robot traversing a corridor. The robot was equipped with 16 sonar sensors that formulated the input vectors. Two further similar corridors were used as additional experimental environments.

In the first set of experiments the novelty filter was based only on the Kohonen map without any habituated synapses, and the aim was to demonstrate the ability of the novelty filter to learn a representation of an environment and recognize novel features in it. The experimental procedure was as follows: a number of learning and non-learning trials, succeeding each other in turn, were initially conducted in the first environment. The novelty filter was always active, even during the non-learning phase to better demonstrate what has been learnt in the previous learning trial. Once the first environment was learnt, as indicated by the lack of any novelties produced by the novelty filter, the robot was moved to another similar corridor where the current novelty filter was further trained using the same foregoing procedure. The experimental results showed that the system was able to firstly learn the first environment and then expand its knowledge to the second one. An interesting question would be to investigate the effect of further training in the second environment to what was already learnt in the first environment.

In the second set of experiments the effect of the habituated synapses was investigated, and the aim was to demonstrate the phenomenon of habituation and how stimuli that have not been observed for some time were being forgotten (dishabituation). Initially, only one environment was used, and features of it were deliberately altered between the trials. The experimental results confirmed the model of habituation used; the system was able to learn and habituate over the environment, but when a feature was changed and not observed for some time, then the novelty filter showed increasing levels of novelty over this feature.

To further investigate the effect of habituation, the robot was moved to another similar corridor after being trained in the first corridor. After each learning trial in the second environment, a non-learning trial, with the novelty filter still active, was conducted back in the first training environment. The experimental results showed that, as its training progressed in the second environment, the system exhibited increasing novelty levels in the first environment. Such a behaviour was observed largely because of the habituated synapses. However, dishabituation might not be the only reason for the increasing levels of novelty detection in this case. It might be possible that the further training received in the second environment may have had an impact on what has been already known, i.e. on the representation of the first environment.

Although this form of habituation is beneficial since it promotes temporal exploration, it has its limitations, mainly because it assumes that the frequency of seeing a stimulus and the amount of time needed to forget it are always the same for all stimuli and for every individual. Furthermore, the frequency and time parameters are set subjectively by the system designer. A more sophisticated mechanism of habituation, such as one that takes into account other aspects of the environment, the individual and the task goals at hand, may be more realistic.

One other issue is the static structure of the Kohonen maps. They are capable of being trained only up to a certain amount of data, depending on their fixed size, which is set initially. This is a drawback for the use of Kohonen maps in the domain of cumulative learning as at some point they will become replete and incapable of further learning.

4.2 Grow-When-Required (GWR) networks

In order to address the drawbacks caused by the static structure of the Kohonen maps, [15] proposed a dynamic network, called the 'Grow-When-Required' (GWR) network, capable of expanding as necessary when it seems to become replete and unrepresentative of new data.

The GWR network consists of two important components: nodes with their associated weights, and edges that separate clusters of nodes that represent similar perceptions. The edges are being updated using the competitive Hebbian learning method used by [6, 16]. In brief, each edge has a counter that is incremented when the edge is connected to the winning node. When an edge exceeds a threshold then it is deleted. A node can also be removed if it has no neighbours, i.e. when it has no edge connections.

The idea of a dynamically growing network is not novel (see [15, §2] for related work). The originality of the GWR network lies in the way that new nodes are added to it and how it deals with a common issue of expandable networks; this being the premature addition of a new node, before giving an opportunity for the last added node to settle. To overcome this problem many techniques allow a new node to be added after a certain number of steps (e.g. [5, 6]). In contrast to this method, the GWR network allows nodes to be added at any time. This is achieved as follows.

A new node is added when the activation value of the best-matching unit is below a threshold set by the experimenter. To deal with the issue of premature creation of a node, the nodes of the GWR network have a variable that decreases exponentially from 1, signifying that the node has not fired often and it is a new untrained node, to 0, signifying that the node has fired often and it had time to be trained. This variable is resembling the phenomenon of habituation presented in Section 4.1.

As such when the activation value of the best-matching unit is above an insertion threshold then a new node is added if the network is mature, otherwise the network is further trained. The weights of the new node are initialized to the mean average of the weights of the best-matching node and the input, i.e. the new node lies between the best-matching node of the network and the input.

A set of experiments was carried out to test the effectiveness of the GWR network as a novelty filter [14]. The experimental task was that of a robot learning the representation of a set of corridors by exploring them through a wall-following behaviour.

For each one of the environments a separate GWR network/filter was trained. The effectiveness of the system was assessed based on whether it was able to select the correct network representing the environment it was in. This was achieved by keeping normalized values, called familiarity indices, which indicated how well each one of the filters represented the current environment. The familiarity index f_i of a filter i was equal to $f_i + c \times n_i$, where n_i was the filter's output, and c was a scaling constant.

Each familiarity index was updated for the current perception, and then all of the familiarity indices were normalized so that their sum was equal to 1. The initial values of the familiarity indices were all the same and equal to $1/m$, where m was the number of filters.

The experimental results demonstrated that the robot is capable of learning representations of all environments, and in general of selecting the correct filter, although there were instances where the wrong one was chosen. This was because the robot was performing classification based only on the current perception, and as such it was incapable of distinguishing between similar environments.

One of the main advantages of the GWR networks is their ability to expand as necessary. By eliminating the drawbacks of the Kohonen maps caused by their fixed size, GWR networks provide an attractive method for the domain of cumulative learning. Furthermore, as self-organizing maps, they make it easy for the system designer to assess their performance and the similarities between objects. On the other hand, the decision on when to add a new node is determined by a fixed threshold set by the designer, a dynamic way of adjusting it is an open challenge.

5 IM-CLeVeR EU project

The work presented here is part of the EU funded project, “Intrinsically Motivated Cumulative Learning Versatile Robots” (IM-CLeVeR), which involves seven partners from Europe and one consultancy collaboration with a US academic institution. The aims of the IM-CLeVeR project are to develop new methodologies for designing robot controllers that can: 1) cumulatively learn new efficient skills through autonomous development based on intrinsic motivations, and, 2) reuse such skills for accomplishing multiple, complex, and externally-assigned tasks.

During skill-acquisition, the robots will behave like children at play which acquire skills autonomously on the basis of “intrinsic motivations”. During skill-exploitation, the robots will exhibit fast learning capabilities and a high versatility in solving tasks defined by external users due to their capacity to flexibly re-use, compose and re-adapt previously acquired skills. This overall goal will be pursued investigating three fundamental scientific and technological issues: 1) the mechanisms of abstraction of sensory information; 2) the mechanisms underlying intrinsic motivations, e.g. “curiosity drives” that learn to focus attention and learning capabilities on “zones of proximal development”; and, 3) hierarchical recursive architectures which permit cumulative learning.

The study of these issues will also be fueled by a reverse-engineering effort aiming at reproducing, with bio-mimetic models, the results of empirical experiments run with monkeys, children, and human adults.

The controllers proposed will be validated with challenging demonstrators based on a single humanoid robotic platform (iCub, www.robotcub.org). As a main outcome, the project will significantly advance the scientific and technological state of the art, both in terms of theory and implementation, in autonomous learning systems and robots. This overall goal will be achieved on the basis of the integrated work of a highly interdisciplinary Consortium involving leading international neuroscientists, psychologists, roboticists and machine-learning researchers.

6 Conclusions

Novelty detection, i.e. the ability to identify stimuli that are different from what is already known, is an important component for the long term operation of any cumulative learning system and has drawn significant attention within the research community as shown by the increasing number of recent publications.

This paper discussed some key characteristics and open challenges that need to be tackled for the effective operation of novelty detection in cumulative learning tasks. A brief overview of the recent surveys on novelty detection methods was also presented, and two methods of interest were explained in more detail. The work presented here is part of an EU funded project, “Intrinsically Motivated Cumulative Learning Versatile Robots”.

References

1. Agyemang, M., Barker, K., Alhajj, R.: A comprehensive survey of numeric and symbolic outlier mining techniques. Intelligent Data Analysis 10(6), 521–538 (2006)
2. Bakar, Z., Mohemad, R., Ahmad, A., Deris, M.: A comparative study for outlier detection techniques in data mining. In: 2006 IEEE Conference on Cybernetics and Intelligent Systems. pp. 1–6. IEEE (2006)
3. Chandola, V., Banerjee, A., Kumar, V.: Outlier detection: a survey (2007)
4. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Computing Surveys 41(3), 1–58 (2009)
5. Fritzke, B.: Growing cell structuresA self-organizing network for unsupervised and supervised learning. Neural Networks 7(9), 1441–1460 (1994)
6. Fritzke, B.: A growing neural gas network learns topologies, pp. 625–632. MIT Press (1995)
7. Grubbs, F.E.: Procedures for detecting outlying observations in samples. Technometrics 11, 1–21 (1969)
8. Helali, R.G.M.: Data mining based network intrusion detection system: A survey. In: Sobh, T., Elleithy, K., Mahmood, A. (eds.) Novel Algorithms and Techniques in Telecommunications and Networking, chap. 86, pp. 501–505. Springer Netherlands, Dordrecht (2010)
9. Hodge, V., Austin, J.: A survey of outlier detection methodologies. Artificial Intelligence Review 22(2), 85–126 (2004)
10. Markou, M., Singh, S.: Novelty detection: a review part 1: statistical approaches. Signal Processing 83(12), 2481–2497 (2003)
11. Markou, M., Singh, S.: Novelty detection: a review part 2: neural network based approaches. Signal Processing 83(12), 2499–2521 (2003)
12. Marsland, S.: Novelty detection in learning systems. Neural Computing Surveys 3, 157–195 (2003)
13. Marsland, S., Nehmzow, U., Shapiro, J.: A real-time novelty detector for a mobile robot. In: Proc. of the EUREL Conference on Advanced Robotics Systems (2000)
14. Marsland, S., Nehmzow, U., Shapiro, J.: Environment-specific novelty detection. In: ICSAB: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior on From Animals to Animats. pp. 36–45. MIT Press (2002)
15. Marsland, S., Shapiro, J., Nehmzow, U.: A self-organising network that grows when required. Neural Networks 15(8-9), 1041–1058 (2002)
16. Martinetz, T.M., Schulten, K.J.: A neural-gas network learns topologies, pp. 397–402. Elsevier (1991)
17. Patcha, A., Park, J.: An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks 51(12), 3448–3470 (2007)
18. Saunders, R., Gero, J.: The importance of being emergent. In: Proc. of the Conference on Artificial Intelligence in Design (2000)
19. Thompson, R.F., Spencer, W.A.: Habituation: A model phenomenon for the study of neuronal substrates of behaviour. Psychological Review 73(1), 16–43 (1966)
20. Wehmeier, S. (ed.): Oxford Advanced Learner's Dictionary. Oxford University Press, 7th edn. (2005)
21. Zhang, Y., Meratnia, N., Havinga, P.J.M.: Outlier detection techniques for wireless sensor networks: A survey. IEEE Communications Survey & Tutorials (2010)