Methods for Anomaly Detection: a Survey

© Leonid Kalinichenko © Ivan Shanin © Ilia Taraban
Institute of Informatics Problems of RAS
Moscow
leonidandk@gmail.com ivan_shanin@mail.ru tarabanil@gmail.com

Abstract

In this article we review different approaches to the anomaly detection problems, their applications and specific features. We classify different methods according to the data specificity and discuss their applicability in different cases.

1 Introduction

Anomalies (or outliers, deviant objects, exceptions, rare events, peculiar objects) is an important concept of the data analysis. Data object is considered to be an outlier if it has significant deviation from the regular pattern of the common data behaviour in a specific domain. Generally it means that this data object is “dissimilar” to the other observations in the dataset. It is very important to detect these objects during the data analysis to treat them differently from the other data. For instance, the anomaly detection methods are widely used for the following purposes:

- Credit card (and mobile phone) fraud detection [1, 2];
- Suspicious Web site detection [3];
- Whole-genome DNA matching [4, 5];
- ECG-signal filtering [6];
- Suspicious transaction detection [7];
- Analysis of digital sky surveys [8, 9].

The anomaly detection problem has become a recognized rapidly-developing topic of the data analysis. Many surveys and studies are devoted to this problem [1, 3, 4, 5, 10, 11]. The main purpose of this review is to reveal specific features of widely known statistical and machine learning methods that are used to detect anomalies. All considered methods will be categorized by the data form they are applied to.

The paper is organized as follows. In Section 2 we introduce three generic data representations that are most commonly used in anomaly detection problems: Metric Data, Evolving Data and Multistructured Data. In Sections 3, 4 and 5 these data forms are discussed in detail, each form is related to a certain class of problems and appropriate methods that are presented with the application examples. In Section 6 we discuss specific features of the anomaly detection problem that make strong impact on the methods used in this area. Section 7 contains conclusions and results of this review.

2 Data forms

The precise definition of the outlier depends on the specific problem and its data representation. In this survey we will establish a correspondence between concrete data representation forms and suitable anomaly detection methods. We assume that the data are usually presented in one of three forms: Metric Data, Evolving Data and Multistructured Data. Metric Data are the most common form of data representation, when every object in a dataset has a certain set of attributes that allows to operate with notions of "distance" and "proximity". Evolving Data are presented as well-studied objects: Discrete Sequences, Time Series and Multidimensional Data Streams. Third form is the Multistructured Data, under this term we understand the data that are presented in unstructured, semi-structured or structured form. This data form may not have a rigid structure, and yet it can contain various data dependencies. The most usual task with this type of data is to extract attributes that would allow using metric data oriented methods of the outlier analysis. In our survey the Multistructured Data are specialized as the Graph Data or Text Data.

3 Metric Data Oriented Methods

In this section the methods are considered that use the concept of “metric” data: such as the distance between objects, the correlation between them, and the distribution of data. We assume that the data in this case represents the objects in the space, so-called points. Then the task is to determine regular and irregular points, depending on the specific metric distance between objects in the space, or the correlation, or the spatial distribution of the points. In this case, we consider a structured data type, i.e., objects, which do not depend on time (time series are discussed in Section 4). Metric data form is the most widely-used, usually due to the fact that almost all entities can be represented as a structured object, a set of attributes, and thus as a point in a particular space [12]. Thus, these methods are used in various applications, e.g., in medicine and astronomy. We subdivide methods based

Proceedings of the 16th All-Russian Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections” — RCDL-2014, Dubna, Russia, October 13–16, 2014.

on the notion of distance, based on the correlations, data distributions and finally related to the data with high dimension and categorical attributes. We now turn to a more detailed review of certain types of these methods.

3.1 Distance-Based Data

Basic set of methods that use the notion of distance includes *clustering* methods, *K nearest neighbors* and their derivatives. Clustering methods use the distance defined in space to separate the data into homogenous and dense groups (clusters). If we see that the point is not included in large clusters, it is classified as anomaly. So we can assume that small clusters can be clusters of anomalous objects, because anomalies may also have a similar structure, i.e., be clustered. *K*-nearest neighbors method [13] is based on the concept of proximity. We consider *k* nearest points on the basis of certain rules, that decide whether the object is abnormal or not. A simple example of such rule is the distance between objects, i.e., the farthest object from its neighbors the more likely is abnormal. There are various kinds of rules starting from the distance-based rules to the neighbor distribution-based. For example, *LOF* (Local outlier factor) [14] is based on the density of objects in a neighborhood. Examples of clustering methods of anomaly detection in astronomy can be found in [15, 16, 17]. Besides classic clustering methods, many machine learning techniques can be used: e. g. modified methods of neural networks – *SOM* (Self-organizing map) [18, 19].

As an example, consider [20]. Authors propose their own clustering algorithm that also classifies anomalies. The main task in this case is to find erroneous values and interesting events in sensor data. Using Intel Berkeley Research lab dataset (2.3 million readings from 54 sensors) and synthetic dataset their algorithm reached Detection rate = 100%, False alarm rate = 0.10% and 0.09% respectively. These experimental results show that their approach can detect dangerous events (such as forest fire, air pollution, etc.) as well as erroneous or noisy data.

3.2 Correlated Dimension Data

The idea of these methods is based on the concept of correlation between data attributes. This situation is often found in real data because different attributes can be generated by the same processes. Thus, this effect allows to use linear models and methods based on them. A simple example of these methods is the linear regression. Using the method of linear regression of the data we are trying to bring some plane, which describes our data, then as the anomalous objects we pick those that are far away from this plane. Also often *PCA* (Principal component analysis) [21] can be used aiming at the reducing of the dimensionality of the data. Due to this the *PCA* is sometimes used in preprocessing data as in [15]. But it can also be directly used to separate anomalies. In this case, the basic idea is that at new dimensions it is easier to distinguish normal objects from abnormal objects [22].

3.3 Probabilistically Distributed Data

In probabilistic methods, the main approach is to assume that the data satisfy some distribution law. Thus, anomalous objects can be defined as objects that do not satisfy such basic rule. A classic example of these methods is the EM [23, 24], an iterative algorithm based on the maximum likelihood method. Each iteration is an expectation and maximization. Expectation supposes the calculation of the likelihood function, and maximization step is finding the parameter that maximizes the likelihood function. As well there are methods based on statistics, data distribution. These include the tail analysis of distributions (e.g., normal) and using the Markov, Chebyshev, Chernoff inequality.

An example of finding anomalies in sensors of rotating machinery is considered in [27]. In this task rolling element bearing failures are determined as anomalies. In practice, such frequent errors are one of the foremost causes of failures in the rotating mechanical systems. Comparing with other SVM-based approaches, the authors apply a Gaussian distribution. After choosing threshold and calculating parameters of distribution the anomalies are found. For testing they use vibration data from the NSF I/UCR Center for Intelligent Maintenance Systems (IMS – www.imscenter.net) and reach 97% accuracy.

Another examples of application of these methods can be found in [25, 26].

3.4 Categorical Data

The appropriate anomaly detection methods operate with continuous data - thus, one approach is to translate the categorical into continuous attributes. As an example, categorical data can be represented as a set of binary attributes. Certainly this kind of transformation may increase the dimension of the data, but this problem can be solved with methods of dimensionality reduction. Different probabilistic approaches also can be used for processing categorical data. It is clear that these approaches are not the only ones that can work with the categorical data. For example, some methods may be partially modified for using categorical data types: distance and proximity can be extended for categorical data.

3.5 High-Dimensional Data

In various applications the problem of the large number of attributes often arises. This problem implies the extra attributes, the incorrectness of the concepts of the distance between the objects and the sophistication of methods. For example, correlated dimension methods will work much worse on a large number of attributes. The main way of solving these problems is the search of subspaces of attributes. Earlier we mentioned the *PCA*, which is most commonly used for this task. But when selecting a small number of attributes other problems will be encountered. By changing the number of attributes, we lose information. Because of the small samples of anomalies, or the emergence of new types of anomalies, previously "abnormal" attributes can be lost.

More subtle approach for this problem is the Sparse Cube Method [28]. This technique is based on analysis of the density distributions of projections from the data, then the grid discretization is performed (data is forming a sparse hypercube at this point) and the evolutionary algorithm is employed to find an appropriate lower-dimensional subspace.

Many applications are confronted with the problem of high dimension. [29] will be taken as an example. Here authors searched for images, characterized by low quality, low illumination intensity or some collisions. They compare the PCA-based approach and the proposed one which is based on the random projections. After projection LOF works with neighborhood that was taken from source space. Both approaches show good results, but the second is much faster at large dimensions than PCA and LOF.

4 Evolving Data

It is very common that data is given in a temporal (or just consecutive) representation. Usually it is caused by the origin of the data. The temporal feature can be discrete or continuous, so the data can be presented in sequences or in time series. Methods that we review in this section can be applied to various common problems in medicine, economy, earth science, etc. Also we review methods suitable for "on-line" outlier analysis in data streams.

4.1 Discrete Sequences Data

There are many problems that need outlier detection in discrete sequences (web logs analysis, DNA analysis, etc. [3, 4]). There are several ways to determine an outlier in the data presented as a discrete sequence. We can analyze values on specific positions or test the whole sequence to be deviant. Three models are used to measure deviation in these problems: distance-based, frequency-based and *Hidden Markov Model* [10]. In the survey [30] the methods are divided in three groups: sequence-based, contiguous subsequence-based and pattern-based. The first group includes *Kernel Based Techniques, Window Based Techniques, Markovian Techniques*, contiguous subsequence methods include *Window Scoring Techniques* and *Segmentation Based Techniques*. Pattern-based methods include Substring Matching, Subsequence Matching and Permutation Matching Techniques [30].

In the work [34] the classic host-based anomaly intrusion detection problem is solved. The study is devoted to Windows Native API systems (a specific WindowsNT API that is used mostly during system boot), while most of other works consider UNIX-based systems. Authors analyse system calls in order to detect the abnormal behaviour that indicates an attack or intrusion. In order to solve this problem authors use a slide window method to establish a database of "normal patterns". Then the SVM method is used for anomaly detection, and in addition to that several window-based features are used to construct a detection rule. The method was tested on the real data from Win2K and WinXP systems (including logs of the important system processes such as svchost, Lsass, Inetinfo) and showed good results. One of the practical examples is given also in [31].

4.2 Time Series Data

If the data strongly depends on time, then we are facing the need to predict the forthcoming data and analyze the current trends. The most common way to determine an outlier is a surprising change of trends. The methods considered are based on well-developed apparatus of time series analysis including *Kalman Filtering*, Autoregressive Modeling, detection of unusual shapes with the Haar transform and various statistic techniques. Historically, the first approach to finding this sort of outliers used an idea from the immunology [33].

5 Multistructured Data

Sometimes the data is presented in a more complex form than numerical "attribute / value" table. In this case it is important to understand what an outlier is by using of the appropriate method of analysis. We will review two cases that need specific analysis: textual data (e.g., poll answers) and data presented as graph (e.g., social network data).

5.1 Text Data

In connection with the development of communications, world wide web, and especially with the advent of social networks, an interest in the analysis of texts on the Internet greatly increased. Considering the text analytics and anomaly detection, several major tasks can be distinguished: searching for abnormal texts – such as spam detection and searching for non-standard text – novelty detection. When solving these problems, the main problem is to represent texts in metric data. Thus we may use the previously defined methods. A simple way is to use the standard metrics for texts, such as the tf-idf. Extaction of entites from texts also is widespread. Using natural language processing techniques such as *LSA* (Latent semantic analysis) [34] it is possible to group text, integrating it with the standard anomaly detection methods. Due to the large number of texts, often the learning may have supervised character.

In [36] a study is focused on spam detection. Using the tf-idf measure their algorithm is based on computing distances between messages. Then it constructs "normal" area using training set. Afterwards area's threshold determines whether an email was a spam. LingSpam (2412 ham, 480 spam), SpamAssassin(4150 ham, 1896 spam) and TREC(7368 ham , 14937 spam) were selected as experimental data sets. The spam detector shows high accuracy and low false positive rate for each dataset.

5.2 Graph Data

In this section we review how methods of data analysis depend on the graph structure. The main difference is that the graph can be large and complex or, in the contrary, can consist of many smaller and simpler graphs. The main problem here is to extract appropriate attributes from nodes, edges and subgraphs that allow to use methods considered in Section 3. In the first case we will review methods that extract numerical attributes from smaller graphs and treat them like data objects using algorithms from Section 3. In case of a large and complex graph we may be interested in node outliers, linkage outliers and subgraph outlier. Methods that analyze node outliers usually extract attributes from the given node and its neighborhood, but in case of a linkage outlier detection the concept of an outlier itself becomes very complex [10, 3]. We will consider that edge is an outlier if it connects nodes from different dense clusters of nodes. The most popular methods are based on the random graph theory, matrix factorization and spectral analysis techniques [10]. Another problem in this section is to detect subgraphs with a deviant behavior and to determine its structure and attribute extraction [37].

Concrete definition of the outlier node or edge can differ according to a specific problem. For example, in [38] several types of anomaly are considered: near-star, near-clique, heavy-vicinity and dominant edge. Anomalous subgraphs are often detected using the *Minimal Description Length principle* [39, 40, 41]. One of the most important application today is Social Network Data – many popular modern techniques are used in this area: *Bayesian Models* [42], *Markov Random Field*, *Ising Model* [43], *EM algorithm* [44] as well as *LOF* [45].

In [44] authors perform anomaly detection methods for social networks. Social network contains information about its members and their meetings. The problem statement is to find abnormal meeting and to measure its degree of abnormality. The problem specificity is that the number of meetings is very small compared to the number of members, that makes challenging to use common statistical methods. In order to solve the problem authors use the notion of hypergraph. The vertices of the hypergraph are considered as members of the social network and the edges are considered as meetings of the members (each edge of a hypergraph connects some set of vertices together). The anomalies are detected through density estimation of p-dimensional hypercube (the EM algorithm tunes a two-component mixture). The method is tested on a synthetic data and shows relatively low estimation error. It is also considered to be a scalable method, which makes it very valuable to use on large social networks.

6 Specific features of the anomaly detection methods comparing to the general machine learning and statistics methods

In this article we show the application for the anomaly detection of various data mining methods that can re-use of the general machine learning and statistical algorithms. The anomaly detection problem has its own specific features making possible to tune the appropriate general algorithms properly turning them into the more efficient ones.

Let us consider one of the basic concept of machine learning – the classification problem. The anomaly detection problem can be considered as a classification problem, in that case the data is assumed to have the class of anomalies. Most of the methods that solve classification problems assume that data classes have some sort of inner predictable structure. But the only prediction that can be made about anomalies is that these objects do not resemble non-outlier "normal" data. In this case, in order to solve the anomaly detection problem, the outlier class modeling can be senseless and unproductive. Instead of this, one should pay attention to the structure of the normal data, its laws of distribution.

The machine learning methods can be divided in three groups: supervised, semi-supervised and unsupervised methods. The first group is the most learned. It requires the labeled "training" dataset, and this is exactly the situation described above: the information about the outlier class is used to tune a model of it in order to predict it's structure, which has often very complex or random nature. The semi-supervised methods use information only about the "normal" class, so these methods have better specifications for anomaly detection problem as well as unsupervised methods, which do not use any information besides the structure and configuration of the unlabeled data.

Another important specific feature of the anomaly detection problem is that usually abnormal objects are significantly rare (compared to the non-outlier objects). This effect makes hard to construct a reliable training dataset for supervised methods. Also, if this effect is not presented in the data, most of known methods will suffer from high alarm rates [47, 48].

7 Conclusion

In this paper we introduced an approach to classify different anomaly detection problems according to the way the data are presented. We reviewed different applications of the outlier analysis in various cases. At the end we summarized specific features of the methods suitable for the outlier analysis problem. Our future plans include preparing of a university master level course focused on the anomaly detection as well as working on the anomaly detection in various fields (e.g. finding peculiar objects in massive digital sky astronomy surveys).

References

[1] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A Survey. ACM Computing Surveys, 41(3), 1–58. Doi:10.1145/1541880.1541882

[2] Kou, Y., Lu, C., & Sinvongwattana, S. (2004). Survey of Fraud Detection Techniques Yo-Ping Huang, 749–754.

[3] Pan Y., Ding X. Anomaly based web phishing page detection // Computer Security Applications Conference, 2006. ACSAC'06. 22nd Annual. – IEEE, 2006. – C. 381–392.

[4] Tzeng, J.-Y., Byerley, W., Devlin, B., Roeder, K., & Wasserman, L. (2003). Outlier Detection and False Discovery Rates for Whole-Genome DNA Matching. Journal of the American Statistical Association, 98(461), 236–246. doi:10.1198/016214503388619256

[5] Wu, B. (2007). Cancer outlier differential gene expression detection. Biostatistics (Oxford, England), 8(3), 566–75. doi:10.1093/biostatistics/kxl029

[6] Lourenço A. et al. Outlier detection in non-intrusive ECG biometric system // Image Analysis and Recognition. – Springer Berlin Heidelberg, 2013. – C. 43–52.

[7] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2012). Isolation-Based Anomaly Detection. ACM Transactions on Knowledge Discovery from Data, 6(1), 1–39. doi:10.1145/2133360.2133363

[8] Djorgovski, S. G., Brunner, R. J., Mahabal, A. A., & Odewahn, S. C. (2001). Exploration of Large Digital Sky Surveys. Observatory, 1–18.

[9] Djorgovski, S. G., Mahabal, A. A., Brunner, R. J., Gal, R. R., Castro, S., Observatory, P., Carvalho, R. R. De, et al. (2001a). Searches for Rare and New Types of Objects, 225, 52–63.

[10] Aggarwal, C. C. (2013). Outlier Analysis (introduction). doi:10.1007/978-1-4614-6396-2

[11] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A Survey. ACM Computing Surveys, 41(3), 1–58. doi:10.1145/1541880.1541882

[12] Berti-équille, L. (2009). Data Quality Mining : New Research Directions. Current.

[13] Stevens, K. N., Cover, T. M., & Hart, P. E. (1967). Nearest Neighbor Pattern Classification. EEE Transactions on Information Theory 13, I, 21–27.

[14] Breunig, M. M., Kriegel, H., Ng, R. T., & Sander, J. (2000). LOF?: Identifying Density-Based Local Outliers, 1–12.

[15] Borne, K., &Vedachalam, A. (2010). EFFECTIVE OUTLIER DETECTION IN SCIENCE DATA STREAMS. ReCALL, 1–15.

[16] Borne, K. (n.d.). Surprise Detection in Multivariate Astronomical Data.

[17] Henrion, M., Hand, D. J., Gandy, A., &Mortlock, D. J. (2013). CASOS: a Subspace Method for Anomaly Detection in High Dimentional Astronomical Databases. Statistical Analysis and Data Mining, 6(1), 1–89.

[18] Networks, K. (n.d.). Data Mining Self – Organizing Maps, 1–20.

[19] Manikantan Ramadas, Shawn Ostermann, Brett TjadenDetecting Anomalous Network Traffic with Self-organizing Maps.(2003) Recent Advances in Intrusion Detection Lecture Notes in Computer Science. Vol. 2820, 36–54.

[20] Purarjomandlangrudi A., Ghapanchi A. H., Esmalifalak M. A Data Mining Approach for Fault Diagnosis: An Application of Anomaly Detection Algorithm // Measurement. – 2014.

[21] Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433–459. doi:10.1002/wics.101

[22] Dutta H. et al. Distributed Top-K Outlier Detection from Astronomy Catalogs using the DEMAC System // SDM. – 2007.

[23] Cansado, A., & Soto, A. (2008). Unsupervised Anomaly Detection in Large Databases Using Bayesian Networks. Network, 1–37.

[24] Zhu, X. (2007). CS838-1 Advanced NLP : The EM Algorithm K-means Clustering, (6), 1–6.

[25] Spence, C., Parra, L., & Sajda, P. (2001). Detection, Synthesis and Compression in Mammographic Image Analysis with a Hierarchical Image Probability Model, 3–10.

[26] Pelleg, D., & Moore, A. (n.d.). Active Learning for Anomaly and Rare-Category Detection.

[27] Fawzy A., Mokhtar H. M. O., Hegazy O. Outliers detection and classification in wireless sensor networks // Egyptian Informatics Journal. – 2013. – T. 14, № 2. – C. 157–164.

[28] Aggarwal C. C., Philip S. Y. An effective and efficient algorithm for high-dimensional outlier detection // The VLDB journal. – 2005. – T. 14, № 2. – C. 211–221.

[29] De Vries, T., Chawla, S., &Houle, M. E. (2010). Finding Local Anomalies in Very High Dimensional Space. 2010 IEEE InternationalConferenceonDataMining, 128–137. doi:10.1109/ICDM.2010.151

[30] Chandola V., Banerjee A., Kumar V. Anomaly detection for discrete sequences: A survey // Knowledge and Data Engineering, IEEE Transactions on. – 2012. – T. 24, № 5. – C. 823–839.

[31] Budalakoti S., Srivastava A. N., Otey M. E. Anomaly detection and diagnosis algorithms for discrete symbol sequences with applications to airline safety // Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. – 2009. – T. 39, №. 1. – C. 101–113.

[32] Wang M., Zhang C., Yu J. Native API based windows anomaly intrusion detection method using SVM // Sensor Networks, Ubiquitous, and Trustworthy Computing, 2006. IEEE International Conference on. – IEEE, 2006. – T. 1. – C. 6.

[33] Dasgupta D., Forrest S. Novelty detection in time series data using ideas from immunology // Proceedings of the international conference on intelligent systems. – 1996. – C. 82–87.

[34] Susan T. Dumais (2005). "Latent Semantic Analysis". Annual Review of Information Science and Technology 38: 188. doi:10.1002/aris.1440380105

[35] Allan, J., Papka, R., & Lavrenko, V. (1998). On-line New Event Detection and Tracking.

[36] Laorden C. et al. Study on the effectiveness of anomaly detection for spam filtering // Information Sciences. – 2014. – Т. 277. – С. 421–444.

[37] Kil, H., Oh, S.-C., Elmacioglu, E., Nam, W., & Lee, D. (2009). Graph Theoretic Topological Analysis of Web Service Networks. WorldWideWeb, 12(3), 321–343. doi:10.1007/s11280-009-0064-6

[38] Akoglu L., McGlohon M., Faloutsos C. Oddball: Spotting anomalies in weighted graphs // Advances in Knowledge Discovery and Data Mining. – Springer Berlin Heidelberg, 2010. – С. 410–421.

[39] Noble C. C., Cook D. J. Graph-based anomaly detection // Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2003. – С. 631–636.

[40] Eberle W., Holder L. Discovering structural anomalies in graph-based data // Data Mining Workshops, 2007. ICDM Workshops 2007. Seventh IEEE International Conference on. – IEEE, 2007. – С. 393–398.

[41] Chakrabarti D. Autopart: Parameter-free graph partitioning and outlier detection // Knowledge Discovery in Databases: PKDD 2004. – Springer Berlin Heidelberg, 2004. – С. 112–124.

[42] Heard N. A. et al. Bayesian anomaly detection methods for social networks //The Annals of Applied Statistics. – 2010. – Т. 4, № 2. – С. 645–662.

[43] Horn C., Willett R. Online anomaly detection with expert system feedback in social networks // Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. – IEEE, 2011. – С. 1936–1939.

[44] Silva J., Willett R. Detection of anomalous meetings in a social network //Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on. – IEEE, 2008. – С. 636–641.

[45] Bhuyan M., Bhattacharyya D., Kalita J. Network anomaly detection: methods, systems and tools. – 2013.

[46] Portnoy L., Eskin E., Stolfo S. Intrusion Detection with Unlabeled Data Using Clustering (2001) // ACM Workshop on Data Mining Applied to Security (DMSA 01).

[47] Laorden C. et al. Study on the effectiveness of anomaly detection for spam filtering // Information Sciences. – 2014. – Т. 277. – С. 421–444.

[48] Fawzy A., Mokhtar H. M. O., Hegazy O. Outliers detection and classification in wireless sensor networks // Egyptian Informatics Journal. – 2013. – Т. 14, № 2. – С. 157–164.

[49] Yu M. A nonparametric adaptive CUSUM method and its application in network anomaly detection // International Journal of Advancements in Computing Technology. – 2012. – Т. 4, № 1. – С. 280–288.

[50] Muniyandi A.P., Rajeswari R., Rajaram R. Network anomaly detection by cascading k-Means clustering and C4. 5 decision tree algorithm // Procedia Engineering. – 2012. – Т. 30. – С. 174–182.

[51] Muda Z. et al. A K-Means and Naive Bayes learning approach for better intrusion detection // Information technology journal. – 2011. – Т. 10, №. 3. – С. 648–655.

[52] Kavuri V. C., Liu H. Hierarchical clustering method to improve transrectal ultrasound-guided diffuse optical tomography for prostate cancer imaging // Academic radiology. – 2014. – Т. 21, № 2. – С. 250–262.

[53] Li S., Tung W. L., Ng W. K. A novelty detection machine and its application to bank failure prediction // Neurocomputing. – 2014. – Т. 130. – С. 63–72.

[54] Cogranne R., Retraint F. Statistical detection of defects in radiographic images using an adaptive parametric model // Signal Processing. – 2014. – Т. 96. – С. 173–189.

[55] Daneshpazouh A., Sami A. Entropy-Based Outlier Detection Using Semi-Supervised Approach with Few Positive Examples // Pattern Recognition Letters. – 2014.

[56] Rahmani A. et al. Graph-based approach for outlier detection in sequential data and its application on stock market and weather data // Knowledge-Based Systems. – 2014. – Т. 61. – С. 89–97.