Anomaly Detection in Emergency Call Data – the First Step to the Intelligent Emergency Call System Management

Petr Klement
MEDIUMSOFT a.s.,
Ostrava, Czech Republic
petr.klement@mediumsoft.cz

Václav Snášel
Department of Computer Science
FEECS, VŠB – Technical University of Ostrava
Ostrava, Czech Republic
vaclav.snasel@vsb.cz

***Abstract* — a collaborative Emergency call taking information system in the Czech Republic processes calls on the European 112 emergency number. Amounts of various incident records are stored in its databases. The data can be used for mining spatial and temporal anomalies. When such an anomalous situation is detected so that the system could suffer from local or temporal performance decrease, either a human, or an automatic management module could take measures to reconfigure the system traffic and balance its load. In this paper we describe a method of knowledge discovery and visualization with respect to the emergency call taking information system database characteristics. The method is based on Kohonen Self Organizing Map (SOM) algorithm. Transformations of categorical attributes into numeric values are proposed to prepare training set appropriate for successful SOM generation.**

Keywords-Emergency Call; Self Organizing Map; Knowledge Discovery in Databases; Data Clustering

I. INTRODUCTION

Statistical tools are exploited in public safety information systems to deal with historical incident-related data [2], [12]. Knowledge hidden in these data has to be derived by experts using common outputs like tables and graphs. Intelligent unsupervised knowledge discovery and visualization applied on the emergency situation data are not generally known. Even if research communities are encouraged to tackle the topic [6], the response is poor.

The emergency call taking in the Czech Republic is supported by a distributed collaborative information system operated at fourteen regional emergency call centres, or PSAP (Public Safety Answering Points). Each of these PSAP serves emergency calls from the home region primarily. In case that the home PSAP is occupied or out of order, the system automatically reroutes emergency calls to another PSAP where the call is processed in the same way as it would be processed by the PSAP of the home region. Every single operator knows actual operational status and language skills of all the other operators logged in the system. Thanks to the cooperative functionality based on an instant messaging subsystem, transparent to the user, operators can ask for help or offer their free capacity and skills in conference mode to the other operators. As all the descriptive and operational data are shared or replicated between system nodes in background, every operator can receive an emergency call from any region, regardless of his / her position with respect to the incident location.

Experiences from the emergency call taking system's operation suggest that in a normal situation calls are smoothly processed by operators of the region where the incident originated without special demands on system settings. In highly critical situations, when large amounts of incidents happen in a short period of time (e.g. during storms or floods), or many people are announcing the same incident (e.g. plane crash, gas explosion or big fire), the system could be locally overloaded. In this case an intelligent reconfiguring scheme would help to balance the system load with respect to the resources available. Routing schemes defining how calls are distributed, quality of service affecting network throughput for critical services as well as postponing certain less important data replication to lower network traffic in overloaded regions can be managed dynamically with a goal of improving system response in the critical situations.

To be able to apply proper and timely management actions, the system must firstly recognize the critical or anomaly situation. The system includes a central database, containing all records of emergencies, or incidents, from the whole territory of the Czech Republic. This database can be used for current situation monitoring as well as for learning from the historical incidents and mining spatial and temporal anomalies.

This paper is focused on a practical proof of the basic precondition of the process described above, the anomaly situation detection from the historical emergency data. Measures taken after the critical situation is detected are supposed to be applied by a human in this first stage and are not discussed here.

The goal of the presented work is:

(i) Design of a method searching for patterns in the database of incidents, which would point out certain interesting situation developments in time and space.

(ii) The method should reveal these patterns without human involvement and present results to the user in a way allowing fast comprehension and evaluation of the presented output.

In this paper we discuss knowledge discovery in database by unsupervised machine learning, specifically application of the Kohonen Self-Organizing Maps (SOM) algorithm.

In section II we compare our approach with related works, section III describes principles of the Kohonen SOM algorithm used for clustering, section IV deals with features of the SW tool used in the experiments. In section V we present a subset of experiments with the SOM algorithm tuning and data transformations applied to the training set successively to form suitable search space and reach satisfactory outputs. In section VI we discuss results and propose future work and in section VII we form conclusion.

II. Related Works

Anomaly detection methods have been designed and well described for various professional domains. SOM has been used here e.g. for network intrusion detection [9], [13] and [16], fraud detection [1], mechanical fault detection [18] and anomaly detection in generic time series data [4]. A common approach in using SOM for anomaly detection is to build a classifier recognizing between anomalous and non-anomalous classes of data. The map is mostly trained over the non-anomalous data to create a model of the correct situation. After that a single input vector is presented to the trained SOM and the winner neuron closest to the input vector is found. If the distance from the winner's representative is smaller than certain limit, the input is classified as belonging to the winner's cluster of non-anomalous data. Otherwise the input is considered anomalous.

These classification methods assume that either non-anomalous subspace is known before the learning starts [13] or the resulting clusters are compared with an expert classification [1]. Artificial anomalous cases generation by Negative Selection Algorithm inspired by the immune system in combination with back-propagation neural network [4] falls in this category as well. Thus the anomaly detection in this classical view is based on supervised learning.

Our approach is different, based on two facts. Firstly, distinguishing between anomalous and non-anomalous cases within emergency calls is disputable. Secondly, if an emergency situation exceeded the "normal" scale, it would be probably reported by a set of single emergency cases having certain attributes in common. We are therefore interested in revealing certain patterns in the emergency call data rather than deciding whether a fresh new single case is somehow strange comparing to the experience. After the patterns are recognized automatically, the composition is presented to a human to analyze the situation.

Future work would enhance this concept. Providing that the algorithm is being run periodically, the new composition will be shown to the supervisor or to the network management module if the composition of patterns found in the current run is different from the composition formed in the previous run. SOM quality measures would be investigated like the index of the map's disorder [11] or the map goodness based on distances between the winner and the second best-match node [7].

While the SOM algorithm has been widely used for classification tasks, the way of using it for clustering data analysis has been relatively out of scope of the research community [17]. This paper describes another application of SOM in the cluster analysis.

III. Clustering and Self Organizing Maps

Cluster analysis groups objects (data records) into classes (clusters) in the way that objects in the same cluster are very similar while objects in different classes are quite distinct.

One of the possible clustering methods is competitive learning [3].

Commonly used application of the competitive learning is Kohonen Self Organizing Map [8], or SOM, described by Teuvo Kohonen in 1982.

SOM is inspired by the human brain cortex, where information is represented in structures of 2D or 3D grids. Formally, SOM is a type of artificial neural network [5] having two interconnected layers of neurons, the input layer and the output or Kohonen layer.

First step of the Kohonen learning is **competition**. Given the training vector on the network's input and weight vector for each neuron of the Kohonen layer, the neuron with minimal distance between weight and input vectors is excited or selected as the winner of the competition [3], [5].

Second step is **adaptation**. Neurons of the Kohonen layer are organized in one-, two- or three-dimensional lattice, reflecting its biological inspiration. A topological neighbour-affecting function is defined on the Kohonen layer, assigning degree of participation in the learning process to the neurons neighbouring with the winning neuron. In every learning step weight vectors of the winning neuron and its neighbours are adjusted to move closer to the input training vector.

In the batch version of the SOM algorithm [15], equivalent to the Lloyd's vector quantization [10], the winning neuron weights are not adapted immediately after the competition step. After the training set was consumed, weight vector of the output neuron N_i is replaced with the weighted mean value of training cases assigned to the clusters represented by neuron N_i and its neighbours, using neighbour-affecting function as the weight for the mean calculation.

The trained network finally sets its weights so that the topologically near neurons are representing similar training cases while distant ones reflect different cases. This is analogous with the human brain cortex, where similar knowledge is represented by adjacent parts of the cortex. The topology of such a trained SOM forms an inherently good base for clustering.

To get a satisfactory approximation of a data set with higher variance, the number of neurons in the static SOM should exceed the number of potential clusters. Agglomerative clustering [17] is therefore used over the trained SOM.

Initially, each of the SOM neurons represents a separate cluster. In each iteration a distance function is computed for every couple of clusters and those with the shortest distance are merged together to form a new cluster. Iteration process stops when the specified number of clusters is reached.

Examples of distance functions being used with SOM are the overall variance of the map [17], the Ward and the SOM-Ward distance [15].

In our work, the SOM algorithm with the clustering extension performs unsupervised cluster analysis over the training set without human assistance. If there is an anomaly situation described in the training set of records, we suppose that this anomaly would be testified by some subset of records which would be revealed in the form of an isolated cluster.

SOM realises transformation of relations of the objects from the m-dimensional input space in a two-dimensional map of nodes (neurons) of the resulting Kohonen network. Complexity of the input space is reduced significantly and in conjunction with colouring of nodes of the resulting network data clusters can be effectively visualised.

Thus the SOM-based method could satisfy both conditions of the goal stated in the beginning of this paper.

IV. Characteristics of the Exploited SW

One of the most advanced applications of the SOM method currently available, Viscovery® SOMine 5.0., was used in the practical part of this work [15]. SOMine uses the batch version of the SOM algorithm, the Gaussian neighbour-affecting function and the short cut winner search heuristic for building the map. Short cut winner search is based on the assumption that the winner for the given input vector in the current iteration is close to the winner for the same input vector in the previous iteration. Competition starts from the previous iteration winner and its neighbourhood. If there is a better match within the neighbourhood, the new winner is registered and the competition continues in its neighbourhood. If there is no better match, the competition stops without searching the rest of the network. This heuristic speeds up the competition phase and the algorithm itself significantly, with a certain risk of neglecting clusters formed by small sets of data.

SOMine also performs agglomerative hierarchical clustering with the Ward and SOM-Ward distance function [15]. The SOM-Ward function, used for experiments, computes the Ward distance for adjacent clusters only, omitting pairs of non-adjacent clusters. The clustering process is then faster.

Numeric attributes are normalized by SOMine into the ‹0, 1› interval to eliminate differences in magnitudes of the input data attributes.

Categorical attributes have to be transformed to numeric values to be usable in the SOM algorithm. SOMine applies the following method [15]:

Let $a_1, \ldots, a_k$ be values of the categorical attribute x. Attribute x is replaced by k new attributes $x_1, \ldots, x_k$, set to:

$x_i = 1$ in case that x has value a_i,

$x_i = 0$ otherwise.

Obviously, this kind of a transformation raises time complexity of the algorithm. It has also another substantial disadvantage. In Euclidean distance of two vectors with numeric and categorical attributes the difference in values of a categorical attribute adds 1 to the result while difference in the normalized values of a numeric attribute is adding only fractions far less then 1. Euclidean metric then can evaluate vectors having similar numeric attributes but different categorical attribute as distant and vectors matching in categorical attributes as close, even if they may differ substantialy in the numeric attributes.

Proposition 1: Described transformation of categorical attributes confuses competition phase and deteriorates results of the network learning process.

V. Experiments

Incident records from the period Feb. 1st – March 31st 2008 were used for experiments. On March the 1st 2008 the territory of the Czech Republic was swept by the Emma hurricane.

Input data set consists of about 25 000 records. The SOM-based procedure should find records related to the Emma hurricane (Emma-records), visualising the Emma-cluster formed by Emma-records in an automatic and effective way.

Emma is characterized by:

- Prevailing types in the incident classification (storm, danger status removal, obstacle removal);
- Raised frequency of incidents, namely of the above stated types, in time;
- Raised frequency of incidents in certain regions (districts).

Training vectors presented to the SOM algorithm consist of attributes “time of the incident beginning”, “incident classification”, “region of the incident origin” and artificial attributes derived from these primary ones. In the following text attributes are presented in the notation: ATTRIBUTE_NAME / Data Type / Description.

A. Experiment 1

We are starting with original attributes:

CALL_START / Datetime / Date and time of the call beginning;

CLASSIFICATION / Categorical / Classification of the incident type;

DISTRICT / Categorical / The district where the incident happened.

For the result, see Figure 1. Every picture shows a trained SOM map of 20x20 nodes, depicted as small hexagons, projected onto values of the selected attribute or results of a mathematical expression. The nodes are coloured according to the mean values of the presented attribute in all the records assigned to the specific node. Scales are shown at the bottom of every picture. The black line inside a map defines a cluster of nodes with related attribute values.

The first and second maps show the CALL_START and CLASSIFICATION attributes, the rightmost map shows frequency of records assigned to the nodes of the trained SOM.

The darker the node colours in the frequency map, the more records were classified to the node. Quite a lot of empty white nodes suggest low quality of learning.

Shadowed nodes contain Emma-records together with other records not belonging to Emma.

.

In compliance with the Proposition 1, two categorical attributes used in this experiment make automatic Emma-cluster selection impossible.

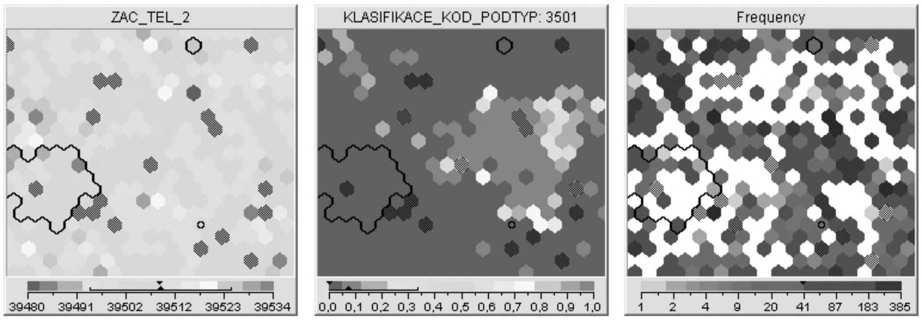

Figure 1. Experiment 1 results

B. Experiment 2

We are looking for a transformation of the categorical attributes into real numbers, which would preserve distribution characteristics of the original attribute values. We introduce a time quantum as the base for transformation. Let the quantum be the hour of the incident origin.

Categorical attributes CLASSIFICATION and DISTRICT are transformed into numeric attributes FREQ_CL_H and FREQ_DS_H. Used attributes are:

CALL_START_H / Numeric / Hour of incident origin;

FREQ_CL_H / Numeric / Frequency of the current incident CLASSIFICATION within subset of incidents related to the hour of the current incident origin;

FREQ_DS_H / Numeric / Frequency of the current incident DISTRICT within subset of incidents related to the hour of the current incident origin.

Results in Figure 2 look much better. Shadowed nodes with Emma-records compose continuous area, covering Emma-cluster identified by SOMine as the region defined by the black line in the left section of the map.

Nevertheless, nodes of the Emma-cluster contain records not belonging to Emma categories and Emma-records can be found in nodes outside Emma-cluster. Frequency map is suggesting good overall level of learning, but Emma-cluster contains substantial number of empty nodes.

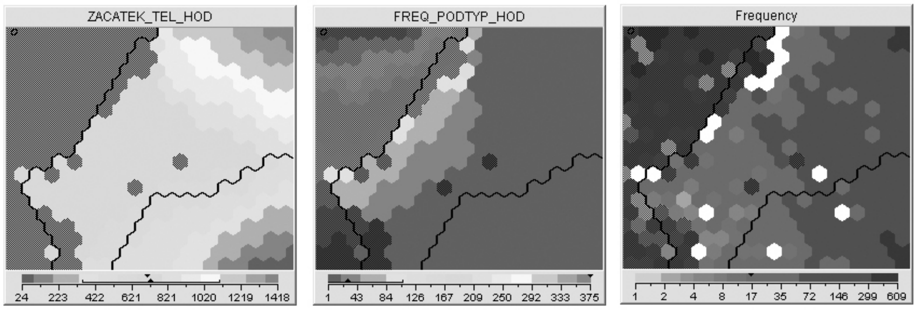

Figure 2. Experiment 2 results

C. Experiment 3

Hourly classification and district frequencies are further substituted by a function of relevance, combining classification or district frequency within the current hour and occurrence of the same classification or district in the remaining hours (adopted from the text classification TF-IDF weighing model [14]).

IMP_CL = FREQ_CL_H * LOG (N / CL_HRS)
IMP_DS = FREQ_DS_H * LOG (N / DS_HRS)

where N means number of hours in the analyzed period and CL_HRS respectively DS_HRS denote number of hours where classification or district occurred. New attributes are:

IMP_CL / Numeric / Relevance of CLASSI-FICATION derived from the frequency of the CLAS-SIFICATION within current hour and its occurrence in the other hours;

IMP_DS / Numeric / Relevance of DISTRICT derived from the frequency of the DISTRICT within current hour and its occurrence in the other hours.

The relevance function assigns higher rates to values with sparse and local occurrence comparing to the values commonly found in time.

Figure 3 shows that shadowed nodes with Emma-records are bound within Emma-cluster. Nodes of the Emma-cluster contain Emma-records only. We have almost met the target, but frequency map shows, that the Emma-cluster also contains most of empty nodes. The learning could be further refined.

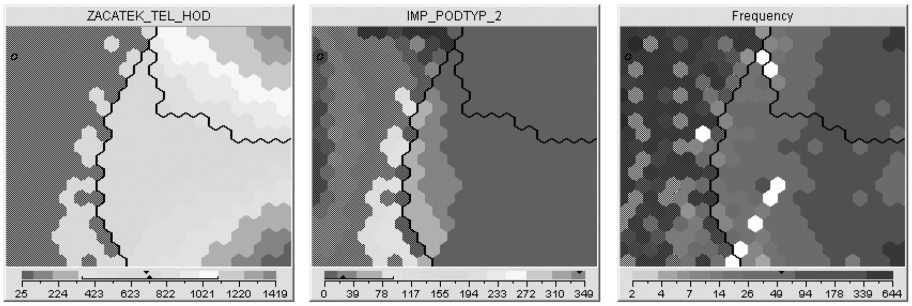

Figure 3. Experiment 3 results

D. *Experiment 4*

Another transformed attribute is the district relevance.

IMP_CL_DS = FREQ_CL_DS * LOG (N / CL_DS)

where N means number of districts in the Czech Republic, FREQ_CL_DS is frequency of the current incident classification in the current incident district during the whole analyzed period and CL_DS denote number of districts where the current incident classification occurred. The new attribute is:

IMP_CL_DS / Numeric / Relevance of DISTRICT derived from the frequency of the CLASSIFICATION in the current DISTRICT and its occurrence in the other districts.

To get an optimal result, the width of the SOM neighbourhood had to be raised, increasing influence of neighbouring neurons in the learning process.

Final output is in Figure 4. Learning quality within Emma-cluster is satisfying as 4 nodes only are empty in the frequency map and the quantization error, defined as the average of the squared distance of all data records associated with a node [15], depicted in the lowest right map has negligible 1.8 comparing to the standard deviation of 70 of the IMP_CL_DS attribute.

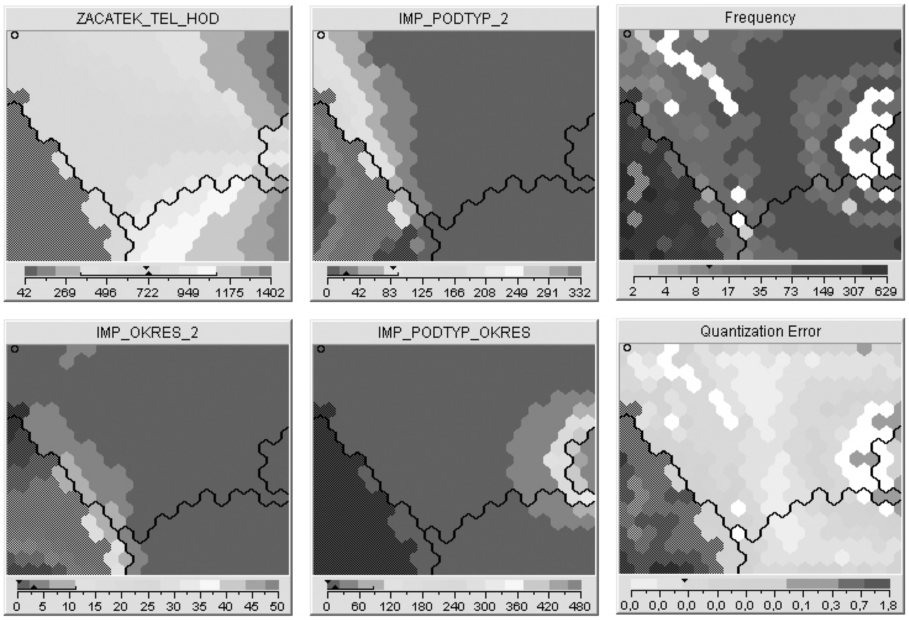

Figure 4. Experiment 4 results

Moreover, an unexpected anomaly was identified by the smallest cluster on the right side of the map, collecting incidents assisted by the Prague City Police and described by a classification which is used solely in Prague.

This result could serve as a proof of acceptability of the proposed method for anomalies discovery in the emergency call data in terms of the goal set in the introduction of this paper.

VI. Discussion

The emergency call system could be locally overloaded in large crisis situations, natural disasters or big accidents. As the system is distributed over the whole territory of the state, operators in regions not affected by crisis can receive calls from the affected regions. An intelligent system reconfiguration could help setting up routing policies, optimizing network traffic and balancing the system load with respect to the resources available.

The first step before the system management activities can be applied is an automatic anomaly or crisis situation detection. We are presenting a possible solution in a form of the following method:

(i) Selection of the relevant set of recent incident records from the emergency call taking system database;
(ii) Creation of the training set, enhancing incident records with artificial attributes and transforming categorical attribute values into the artificial numeric attribute values;
(iii) Processing the training set by SOM algorithm with learning parameters preset appropriately;
(iv) Making conclusions on inspection of records belonging to the nodes of clusters defined by SOM;
(v) In case that the anomaly situation was detected, issuing a signal for the network supervisor or the automatic management module.

We proposed transformations of categorical attributes, discrete values of which are not suitable for the SOM algorithm, into real numbers domain. The transformations are built on frequencies of incidents in time and place, expressing relevance of the incident with respect to the time and place of the incident origin. The learning proceeds on values of the transformed attributes which are nevertheless bound with original records. In this context, values of the original understandable attributes could be used for explanation of the result as well as for further processing.

Traditional anomaly detection methods [4], [13] and [18] use SOM for modelling the anomalous-free space from the set of data approved to be correct (bearing signs of supervised learning) and then classify a new case as anomalous if it falls beside the modelled non-anomalous space. As hardly any emergency situation can be considered anomalous, or any of them is anomalous on the opposite, we could not use the two-class classifier approach with supervised learning. Therefore we searched for certain patterns in data. After the patterns were recognized by unsupervised learning, the composition was always presented to human to analyze the situation.

We shall enhance this concept so that, providing that the algorithm is run periodically, the new composition would be further processed only if the composition in the current run is different from the composition in the previous run. SOM quality measures [11], [7] could help achieving good results here.

To obtain the best from the SOM algorithm, we designed transformations of categorical attributes to get a vector space over real numbers. Thus both precision and speed of the learning algorithm were improved significantly. Apparently, time is crucial for successful reaction in the beginning of the incoming crisis. At the same time the method proposed here has to spend some idle time interval in which records of incidents are collected in an amount allowing successful anomaly recognition. Further research could deal with variable time window for selecting records of the training set. It is expected that an optimal interval could exist in which transformations based on the frequency characteristics would work correctly, while in too long or short intervals extreme values could flatten or disappear.

Time complexity of the learning algorithm has to be taken into account as well. The Viscovery® SOMine SW tool used in our experiments applies batch version of the SOM algorithm with short cut winner search [15] heuristic speeding up the competition phase and the learning algorithm itself. The cost paid for the speed is a certain risk of omitting small clusters of data which could reveal anomalies in their beginning. We have an interesting paradox here: to react quickly on the evolving situation we need faster algorithm, which on the opposite loses its ability to detect small anomalies and therefore needs to wait until enough anomaly-positive reports arrive to reveal the anomaly.

Fortunately the large emergency phenomenon evolves in the order of minutes. In our experiments, creating a SOM over a set of 25 000 records on a common PC took around five minutes. Supposing that the phenomenon starts during the first run of the method, it could be detected in the third run of the method in the worst case. We could then loose from ten to fifteen minutes. Even if this seems to be an eternity in the world of computers, in a world of relatively slowly evolving and long lasting large emergencies this time is reasonable for successful application of the method described above.

VII. Conclusion

In this work we devised and experimentally proved a method which could help optimizing the workload in a cooperative emergency call processing system. We experimented with unsupervised knowledge discovery in the emergency call data. Experiments were focused on revealing clusters of emergency records which would point to abnormal situation in time and place.

Method is based on the Kohonen Self Organizing Map (SOM) algorithm. We used comprehensive and user-friendly

SW tool Viscovery® SOMine 5.0 for SOM creation and visualization.

The SOM algorithm consumed the training set of records consisting of transformed attributes of the original incident records, identified subset of records belonging to the real emergency case, hurricane Emma, and built an isolated cluster over these records automatically. The method proposed here proved its ability of discovering anomalies hidden in data, visualising them in an effective user-friendly manner and indicated a way towards further development of intelligent management of the large distributed and cooperative emergency call information system.

References

[1] Brockett, P.L., Xia, X., Derrig, R.A..: Using Kohonen's Self-organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud, Journal of Risk and Insurance 65(2), 1998, pp. 245-274.

[2] Fire and Rescue Service statistical yearbook, http://www.hzscr.cz/clanek/statistical-yearbook-2007-312484.aspx (cited 22nd March 2009).

[3] Gan, G., Ma, Ch., and Wu, J.: Data Clustering: Theory, Algorithms and Applications, SIAM, Philadelphia, 2007.

[4] Gonzalez, F., and Dasgupta, D.: Neuro-Immune and Self-Organizing Map Approaches to Anomaly Detection: A Comparison, Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS), Canterbury, UK, Sept. 2002, pp. 203-211.

[5] Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edition. Upper Saddle River, NJ: Prentice-Hall, 1999.

[6] ISCRAM, Information Systems for Crisis Response and Management: http://www.efficient-response.com/iscram2008/ISCRAM2008_CFP130.pdf (cited 22nd March 2009).

[7] Kaski, S. and Lagus, K.: Comparing Self-Organizing Maps. In Proceedings of ICANN96, International Conference on Artificial Neural Networks, Springer, Berlin 1996, pp. 809-814.

[8] Kohonen, T.: Self-Organizing Maps. Springer-Verlag, Berlin, 1995.

[9] Lichodzijewski, P., Zincir-Heywood, A.N., and Heywood, M.I.: Dynamic Intrusion Detection Using Self-Organizing Maps. In The 14th Annual Canadian Information Technology Security Symposium (CITSS), 2002.

[10] Lloyd, S., P.: Least Squares Quantization in PCM. IEEE Transactions on Information Theory, vol. 28, no. 2, 1982, pp. 129-137.

[11] Mitra, S., Pal, S.K.: Self-organizing neural network as a fuzzy classifier. IEEE Trans. Systems,Man, Cybernetics 24 (3), 1994, pp. 385–399.

[12] NENA, National Emergency Number Association: http://www.nena.org/911-statistics (cited 20th December 2008).

[13] Ramadas, M., Ostermann, S., and Tjaden, B.: Detecting Anomalous Network Traffic with Self-Organizing Maps. Proc. of Recent Advances in Intrusion Detection, 2003, pp. 36-54.

[14] Salton, G. & Buckley, C.: Term-weighting approaches in automatic text retrieval. Information Processing and Management 24(5), 1988, pp. 513–523.

[15] Viscovery®. SOMine 5.0. Copyright © 1998-2007 by Viscovery Software GmbH, http://www.viscovery.net (cited 7th November 2008).

[16] Vokorokos, L., Balaz, A., and Chovanec, M.: Intrusion Detection System Using Self-Organizing Maps. Acta Electrotechnica et Informatica, No.1, Vol.6, TU Kosice, 2006.

[17] Yang, M. Y.: Extending the Kohonen self-organizing map networks for clustering analysis. In Computational Statistics & Data Analysis 38, 2001, pp. 161–180.

[18] Ypma, A., and Duin, R.P.W.: Novelty detection using self-organizing maps. Progress in Connectionist Based Information Systems, Vol. 2, Springer, 1998, pp. 1322-1325.