# Introduction

In modern age, the intelligent systems are finding wide applications in various fields like Micro-Electro-Mechanical Systems, Robotics, Manufacturing, Medical, Aerospace, Drives and Controls, Business Intelligence, etc. A very important area for applications of the intelligent systems is the World as well as European Earth observation programs. This monograph is closely connected to two of them:

- **UN-SPIDER** – The United Nations Platform for Space-based Information for Disaster Management and Emergency Response;
- **GMES** – European programme for Global Monitoring for Environment and Security, which provides data useful in a range of issues including climate change and citizen's security.

## UN-SPIDER

In its resolution 61/110 of 14 December 2006, the United Nations General Assembly agreed to establish the "United Nations Platform for Space-based Information for Disaster Management and Emergency Response – UN-SPIDER" as a new United Nations programme, with the following mission statement:

*"Ensure that all countries and international and regional organizations have access to and develop the capacity to use all types of space-based information to support the full disaster management cycle".*

The United Nations Office for Outer Space Affairs (UNOOSA) implements the decisions of the General Assembly and of the Committee on the Peaceful Uses of Outer Space and its two Subcommittees, the Scientific and Technical Subcommittee and the Legal Subcommittee. The Office is responsible for promoting international cooperation in the peaceful uses of outer space, and assisting developing countries in using space science and technology.

UN-SPIDER is the first programme of its kind to focus on the need to ensure access to and use of space-based solutions during all phases of the disaster management cycle, including the risk reduction phase, which will significantly contribute to the reduction in the loss of lives and property.

Whereas, there have been a number of initiatives in recent years. They have contributed to making space technologies available for humanitarian and emergency response, UN-SPIDER is the first to focus on the need to ensure access to and use of such solutions during all phases of the disaster management cycle, including the risk reduction phase, which will significantly contribute to reducing the loss of lives and property.

The UN-SPIDER programme is achieving this by being a gateway to space information for disaster management support, by serving as a bridge to connect the disaster management and space communities and by being a facilitator of capacity-building and institutional strengthening, in particular for developing countries.

UN-SPIDER is being implemented as an open network of providers of space-based solutions to support disaster management activities. Besides Vienna (where UNOOSA is located), the programme also has an office in Bonn, Germany and will have an office in Beijing, China. Additionally, a network of Regional Support Offices multiplies the work of UN-SPIDER in the respective regions.

UN-SPIDER aims at providing universal access to all types of space-based information and services relevant to disaster management by being a gateway to space information for disaster management support; serving as a bridge to connect the disaster management and space communities; and being a facilitator of capacity-building and institutional strengthening.

UN-SPIDER ensures that all countries and international and regional organizations have access to and develop the capacity to use all types of space-based information to support the full disaster management cycle. UN-SPIDER is achieving this by being a gateway to space information for disaster management support; serving as a bridge to interlink the disaster management and space communities; and being a facilitator of capacity-building and institutional strengthening.

Capacity-building refers to the process of facilitating the strengthening of the competency of individuals, teams, and agencies to use space-based information to prevent, mitigate, and respond effectively to the challenges posed by natural hazards and related humanitarian crises.

The objective of the capacity-building efforts of UN-SPIDER is to ensure that countries recognize the value of all types of space-based information, and therefore access it to reduce the impacts of disasters and to respond more efficiently in case of such disasters through improved use of this type of information.

To increase capacities regarding the use of space-based information for disaster-risk management and emergency response at the national level, UN-SPIDER pursues three lines of action:

- Institutions: UN-SPIDER will promote the adoption of policies, which ensure that operating procedures within institutions incorporate the use of space-based information, and that they support activities targeting all phases of the disaster management cycle.
- Individuals: UN-SPIDER facilitates access to training programs conducted by Centers of Excellence and Specialized training centers to enhance the knowledge and skills of staff working in institutions, which conduct activities targeting all phases of the disaster management cycle.
- Infrastructure: UN-SPIDER facilitates access to infrastructure (hardware, software, related equipment) to ensure the capacity to access and make use of space based information.
- The respective activities are conducted in a systematic fashion to ensure that capacities within agencies are strengthened. This is coordinated with the network of Regional Support Offices and the National Focal Points.

Curricula are being elaborated targeting both disaster-risk management and emergency response in order to design training courses to be conducted through the Regional Centers for Space Science and Technology Education affiliated to the United Nations, Centers of Excellence, UN training centers linked to UN-SPIDER, and other national or regional training centers where remote sensing and earth observation applications are taught. Complementary efforts are being made to establish an e-learning environment as a means to build a repository of learning material, and information is being compiled to provide a database of training opportunities.

More information for UN-SPIDER is given at URL: [UN-SPIDER, 2010].

## Regional Support Office of UN-SPIDER in Ukraine

UN-SPIDER Regional Support Office (RSO) in Ukraine was established based on the Space Research Institute of the National Academy of Sciences of Ukraine and the National Space Agency of Ukraine.

The Institute signed the cooperation agreement with the United Nations Office of Outer space Affairs (UNOOSA) during the forty-seventh session of the Scientific and Technical Subcommittee of the Committee on the Peaceful Uses of Outer Space (COPUOS) on 10 February 2010 in Vienna, Austria.

UN-SPIDER aims at providing universal access to all types of space-based information and services relevant to disaster management

UN-SPIDER RSO is a regional or national centre of expertise that is set up within an existing entity by a Member State or group of Member States that has put forward an offer to set up and fund the proposed regional support office.

Activities of UN-SPIDER RSO in Ukraine:

- horizontal cooperation;
- outreach and capacity building;
- technical advisory support.

The Space Research Institute of the National Academy of Sciences of Ukraine and the National Space Agency of Ukraine (SRI NASU-NSAU) and the United Nations Office of Outer space Affairs (UNOOSA) signed the cooperation agreement on the establishment of a UN-SPIDER Regional Support Office (RSO) on the 47th Scientific and Technical Subcommittee sessions on 12 February 2010.

The Space Research Institute was established in 1996 at the National Space Agency of Ukraine and National Academy of Sciences of Ukraine for organization of scientific space researches in the country, conducting, and co-ordination of scientific and engineering activities in the area of peaceful exploration and use of outer space.

The Institute's main activities are:

- Pure and applied research in outer space, astrophysics research of objects in the universe, including in ranges unavailable from the earth surface;
- Development of strategy and principles of universe exploration means use in solving scientific and applied issues for the needs of the economy;
- Development and testing, in the space environment, of scientific space exploration equipment and relevant technological processes;
- Development of new spacecraft navigation and control systems and earth and space monitoring systems; improvement of existing ones; creation of information space systems;
- Working out suggestions on the conception and strategy for space programs

The Institute's main projects include:

- use of earth remote sensing and geographic information systems (GIS) for informational support of environmental control;
- monitoring, estimation, and forecasting of underwater petrochemical pollution in Ukraine;
- interball project on exploration of solar-earth relations;
- variant project on measurement of electromagnetic field and electric current flows in ionosphere;
- warning project: satellite complex for exploration of ionosphere phenomena related to seismic activity;

- planning and controlling system for science and engineering experimentation aboard the Ukrainian explorer unit of the International Space Station.

For additional information, please visit URL: [SRI NASU-NSAU, 2010].

## GMES – observing our planet for a safer world

Land, sea, and atmosphere – each Earth component is observed through GMES, helping to make our lives safer. The purpose of GMES is to deliver information, which corresponds to user needs. The processing and dissemination of this information is carried out within the "GMES service component" [GMES, 2010].

The thematic areas within the GMES service component comprise:

- land, marine and atmosphere information – ensuring systematic monitoring and forecasting the state of the Earth's subsystems at regional and global levels;
- climate change information – helping to monitor the effects of climate change, assessing mitigation measures and contributing to the knowledge base for adaptation policies and investments;
- emergency and security information – providing support in the event of emergency and humanitarian aid needs, in particular to civil protection authorities, also to produce accurate information on security related aspects (e.g. maritime surveillance, border control, global stability, etc.).

Managing natural resources and biodiversity, adapting to sea level rise, monitoring the chemical composition of our atmosphere: all depend on accurate information delivered in time to make a difference.

The GMES service component depends on Earth observation data, collected from space (satellites), air (airborne instruments, balloons to record stratosphere data, etc.), water (floats, shipboard instruments, etc.) or land (measuring stations, seismographs, etc.). These facilities are called the GMES infrastructure component; non-space based installations in the GMES infrastructure component are generally referred to as "in situ component".

By securing the sustainability of an information infrastructure necessary to produce output information in the form of maps, datasets, reports, targeted alerts, etc., GMES helps people and organizations to take action, make appropriate policy decisions and decide on necessary investments. GMES also represents a great potential for businesses in the services market, which will be able to make use of the data and information it provides according a full an open access principle.

Earth observation-based services already exist in Europe, but they are dispersed at national or regional level. Because of this, they cannot rely on a sustainable observation capacity. With the exception of meteorological services, long-term availability and reliability of information is not guaranteed. This is why, in order to contribute to improve its response to ever-growing challenges of global safety and climate change, Europe develops a sustained and reliable Earth observation system of its own [GMES, 2010]. Let remark that one of the priority topics to be discussed on the second GMES Operational Capacity Workshop is planned to be in March 2011 again in Sofia is the Bulgarian initiative for regional networks for integrated risk and security management with main targets:

- strengthening and regional cohesion of GMES operational capacity in Europe, taking into account local and regional specific characteristics and capacities;

- space and in-situ infrastructure policy, taking into account a bottom-up approach – attention on "user" needs for spatial data, collected under the "space" and "in-situ" GMES components as well as better data access and efficiency in their practical use, without losing high standards of data security;
- regional "User" driven approach in defining and developing services in support for better prevention, better forecast and transborder/transnational preliminary measures, reducing nature and anthropogenic crisis and disasters impact.
- risk and security prevention based on regional network of servers and service providers, thus providing flexibility in the use of infrastructure and human power depending on the area of impact, complexity and importance of the problem.

## Intelligent data processing in GMES

Let consider a short example.

If one needs to know what is the time, he or she may ask someone and will know "the local" time. However, if one wants to know what time is pointed by all watches in the town or in the country, it is impossible to solve such problem because of the high number of devices – it is impossible in the same moment to scan and receive all information. During processing one part of watches, the other will show new time. So, one will never know "the total" time, i.e. what time is on all watches in the world. (Let make difference with "the global time" because there is a convention all watches to be synchronized with the standardized Greenwich Mean Time (GMT). It is the same all year around – there is no Summer Time, i.e. "Daylight Saving Time".)

In other words, along with modern technologies appears a lot of information. Often, when we want to collect and understand it, the traditional methods are not enough. They might be too general and they need additional processing and transformation. Therefore, we use advanced methods of data processing such as rough sets, genetic algorithms, neural nets, or fuzzy sets. Rough sets have application e.g. in problems of excessive data, problems with correct classification or problems with retrieval hidden relations between data. Placing these methods in environment of distributed applications, based on. Net platform and XML Web Services, introduces new possibilities in usage of artificial intelligence methods in data processing systems [Zielosko and Wakulicz-Deja, 2005].

Technological advancement using intelligent techniques has provided solutions to many applications in diverse engineering disciplines. In application areas such as web mining, image processing, medical, and robotics, just one intelligent data processing technique may be inadequate for handling a task, and a combination or hybrid of intelligent data processing techniques becomes necessary. The sharp increase in activities in the development of innovative intelligent data processing technologies also attracted the interest of many researchers in applying intelligent data processing techniques in other application domains [Wai-Wong et al, 2007].

A briefly classification of the existing methods of data processing by the complexity of results that we want to achieve is given in [Nguyen et al, 1997]:

- in some problems, all we want is one or several numerical values. It may be that we measure some characteristics, or it may be that we know the model, and we want, based on the experimental data, to estimate the parameters of this model. These problems are usually handled by statistical methods;

- in other problems, we want to know a function. We may want to reconstruct an image (brightness as function of coordinates), we may want to filter signal (intensity as a function of time), etc. These methods are usually handled by different regularization techniques;
- finally, there are many complicated problems where we want to reconstruct a model of an analyzed system. Methods that handle such problems are called **intelligent data processing methods**. Many of these methods are based on logic programming, a formalism that (successfully) describes complicated logical statements algorithmically, in a kind of programming language terms. Other methods are based on advanced information modeling technologies, which use rough sets, genetic algorithms, neural nets, fuzzy sets, etc. Very important case is the methods of data mining.

## Data mining in GMES

Data mining technologies have important place in Geosciences. Here we will remember only the main areas of data mining. For further readings see, for instance, [Ramachandran et al, 2006].

Over the past few centuries, the quantity of accumulated information is constantly growing. Because of rapid development of all areas of human activity in modern society, production, economic and social processes became greatly complicate. Most companies, which use information technology resources, collect and store large amounts of data. The challenge that most companies face today is not how to collect and store adequate amounts of data but how to derive meaningful conclusions from this mass information. The answer is in technology of data mining and, in particular, the association rules. Association rules can be classified as data based set of rules that are similar to expert system. They show worthy conditions – attributes that often occur together in a data set. Association rules provide information of the "if-then" structure. Those rules are calculated from the data, and unlike in the last rules of logic, association rules are probabilistic in nature. The first part ("if") is called support of the rule. The second part ("then") is like confidence of this rule. In data mining, association rules are useful for analyzing and predicting the behavior of customer, system, natural processes, etc. Developers use association rules to build programs capable of machine learning. Machine learning is a form of artificial intelligence that seeks to build programs with the opportunity to become more efficient without being explicitly programmed.

Data Mining binds the terms "knowledge discovery in database" and "intellectual analysis of data". The emergence of all these terms is associated with the emergence of new trends in development of tools and methods for data processing. Data mining is a process of analysis of stored databases in the direction of mining new useful information by revealing deep and hidden relationships between seemingly unrelated and unknown to each other values. An important feature of his is that it enables processing multi-dimensional arrays and retrieval of multidimensional relationships while automatically reveals exceptional situations – the information, which is not included in the general laws. Data mining analysis automatically makes assumptions to detect relationships between different components and parameters. The work of analysts who deal with these systems is limited to verification and clarification of the resulting hypotheses. The emergence of data mining is associated with the need to improve techniques for recording and storing data that summarize the work of thousands of people in huge flows of information in various fields. By the time, it became clear that without productive data processing, redundant information is generated. The need of the development of modern technologies of processing such data can be summarized as follows: unlimited data volume, variety, and heterogeneity of data (quantitative, qualitative, and textual); need for concrete and understandable results; processing tools providing data easy use.

Data mining derives its name from the similarities between searching for valuable information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new opportunities by providing these capabilities [Davenport and Harris, 2007]:

1. ***Automated prediction of trends and behaviors***. Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. Main goal is forecasting. It is a method to reveal the models in the relationship between data that are needed for forecasting and modeling. This type of algorithms prepares input and output data for the preparation of illustrative examples. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events. In GMES, the forecasting of disasters and collisions is one of the main tasks. The data mining methods may help to predict possible risk by analyzing large databases of satellite, in-situ, and other information.
2. ***Automated discovery of previously unknown patterns***. Data mining tools sweep through databases and identify previously hidden patterns in one-step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors. Very important is the possibility for segmentation – analysis of existing data form specific groups, based primarily on the parameters of the data, for instance, of the customer – demographics and purchasing power. Clustering algorithms make it possible to delineate homogeneous groups or types of customers for each group can determine the intrinsic and group clients. This enables us better to evaluate its customer base and planning of marketing activities. In GMES, clustering the prime causes of collapses and disasters is important to find regularities in the data and this way to recognize properly the risk situation.

Two technologies are in the base of data mining: machine learning and visualization.

The quality of visualization defines graphic representation of data in their colors, shapes and other elements representing the hidden data relationships.

The effectiveness of methods of machine learning is determined by the capabilities of data mining to explore the much larger amounts of data to reveal connections between them than a person.

The most commonly used techniques in data mining are [Davenport and Harris, 2007]:

- ***Artificial neural networks***: Non-linear predictive models that learn through training and resemble biological neural networks in structure. The knowledge is presented in the form of links, joining a set of conditions. The strength of the relationship is determined by the relationship between factors and data;
- ***Decision trees***: Tree-shaped structures that represent sets of decisions. It is designed to classify data using the gravity of the partition coefficients of elements of data in ever smaller and smaller groups. The decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID);

- ***Genetic algorithms***: Optimization techniques that use processes, such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution. They define natural "breakdown" of the data based on the target variables. Each branch of the tree is a separate part of the rules;
- ***Nearest neighbor method***: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where $k \geq 1$). Sometimes it is called the k-nearest neighbor technique;
- ***Rule induction***: The extraction of useful if-then rules from data, based on statistical significance. Such rules can be generated using the process for requesting and checking various combinations of rules or extraction of any of the decision tree.

How exactly could data mining foresee important things, which are still not known or which may occur later? The answer to this question relates to the technique used for data mining, namely modeling. ***Modeling*** is the act of creating a model of a situation based on existing experience and knowledge and understanding of the response to this situation. This model is then applied to another situation in which the answers are not known [Davenport and Harris, 2007]. This act of building models is something that people made long ago, certainly before the advent of computers and technology to retrieve data. What happens with computers is not very different from the way people build models. In computing, a set range of information about different situations in which knowing the answer, the software for data mining should be extended to these data and derive the characteristics of those data should appear in the model. Once the model is built, it can be used in similar situations where looking for unknown response. While data mining represents a significant progress in analytical tools that are available now, there are limitations on its capabilities. One of the limitations is that although data mining can help to identify patterns and relationships, it reveals what is the value and importance of these models. The user alone must make these distinctions. The second limitation is that although data mining can identify connections between behavior and/or variables, it does not necessarily identify causality. To be successful, data mining requires technical and analytical specialists who can structure the analysis and interpret search results [Davenport and Harris, 2007]. The associative rules represent causal links and determine the likelihood or the coefficient of reliability, allowing drawing appropriate conclusions. Rules presented in the form “IF <condition> THEN <conclusion>” is used for forecasting and evaluating the unknown parameters and meanings.

Data mining stands at the crossroad of databases, artificial intelligence, and machine learning. Association rule mining is a popular and well-researched method for discovering interesting rules from large collections of data. The contemporary databases are very large, reaching giga- and terabytes, and the trend shows further increase. Therefore, for finding association rules one requires efficient scalable algorithms that solve the problem in a reasonable time. The efficiency of frequent item set mining algorithms is determined mainly by three factors: (1) the way candidates are generated; (2) the data structure that is used; and (3) the implementation details. Most investigations focus on the first factor, some describe the underlying data structures, and implementation details are almost always neglected [Bodon, 2003].

The main pillar of association rule mining algorithms is Apriori [Agrawal and Srikant, 1994]. It is the best-known algorithm to mine association rules, which uses a breadth-first search strategy to count the support of item sets and uses a candidate generation function, which exploits the downward closure property of support. Over the years, many improvements of Apriori, supported with different types of memory structures, are proposed.

Recent association rule mining algorithms, based on graph mining can be roughly classified into two categories. The first category of algorithms employs a breadth-first strategy. Representative algorithms in this category include AGM [Inokuchi et al, 2003] and FSG [Kuramochi and Karypis, 2001]. AGM finds all frequent induced sub-graphs with a vertex-growth strategy. FSG, on the other hand, finds all frequent connected sub-graphs based on an edge-growth strategy. Algorithms in the second category use a depth-first search for finding candidate frequent sub-graphs. A typical algorithm in this category is gSpan [Yan and Han, 2002], which was reported to outperform both AGM and FSG in terms of computation time.

A different approach for association rule searching is used in ECLAT (Equivalence Class Clustering and Bottom-up Lattice Traversal) [Zaki et al, 1997]. It is the first algorithm that uses a vertical data (inverted) layout. The frequent item sets are determined using sets of intersections in a depth-first graph.

In graph approaches, the bottleneck is the necessity of performing many graph isomorphism tests. To overcome this problem, alternative approaches use hash-based techniques for candidate generation. The representatives in this direction are DHP [Park et al, 1995] based on direct hashing and pruning, [Özel, and Güvenir, 2001] which proposed the use of perfect hashing, and IHP [Holt and Chung, 2002] that uses inverted hashing and pruning.

FP-Tree [Han and Pei, 2000], Frequent Pattern Mining is another milestone in the development of association rule mining, which breaks the main bottlenecks of the Apriori. FP-tree is an extended prefix-tree structure storing quantitative information about frequent patterns. The tree nodes are arranged in such a way that more frequently occurring nodes will have better chances of sharing nodes than less frequently occurring ones. The efficiency of FP-Tree algorithm has three reasons: (1) FP-Tree is a compressed representation of the original database; (2) it only scans the database twice; (3) it uses a divide and conquers method that considerably reduces the size of the subsequent conditional FP-Tree. The limitation of FP-Tree is its difficultness to be used in an interactive mining system, when a user wants to expand the dataset or change the threshold of support. Such changes lead to repetition of the completely mining process.

The Hmine algorithm [Pei et al, 2001] introduces the concept of hyperlinked data structure "Hyper-structure" and uses it to adjust dynamically links in the mining process. Hyper-structure is an array-based structure. Each node in a Hyper structure stores three pieces of information: an item, a pointer pointing to the next item in the same transaction and a pointer pointing to the same item in another transaction.

The innovation brought by TreeProjection [Agarwal et al, 2001] is the use of a lexicographical tree, which requires substantially less memory than a hash tree. The number of nodes in its lexicographic tree is exactly that of the frequent item sets. The support of the frequent item sets is counted by projecting the transactions onto the nodes of this tree. This improves the performance of counting the number of transactions that have frequent item sets. The lexicographical tree is traversed in a top-down fashion. The efficiency of TreeProjection can be explained by two main factors: (1) the transaction projection limits the support counting in a relatively small space; and (2) the lexicographical tree facilitates the management and counting of candidates and provides the flexibility of picking efficient strategy during the tree generation and transaction projection phrases.

Another data structure that is commonly used is a "trie" (or prefix-tree). Concerning speed, memory need, and sensitivity of parameters, tries were proven to outperform hash-trees [Bodon and Ronyai, 2003]. In a trie, every node stores the last item in the item set it represents, its support, and its branches. The branches of a node can be implemented using several data structures such as a hash table, a binary search tree, or a vector.

Another algorithm for efficiently generating large frequent candidate sets, which use different data structures, is proposed by [Yuan and Huang, 2005] and is called the Matrix Algorithm. The algorithm generates a matrix with entries 1 or 0 by passing over the cruel database only once, and then the frequent candidate sets are obtained from the resulting matrix. Finally, association rules are mined from the frequent candidate sets. Experiment results confirm that the proposed algorithm is more effective than the Apriori Algorithm is.

An interesting theoretical framework for association rule mining algorithm is brought in [Aslanyan and Sahakyan, 2009] that combines the known "unit cube chain decomposition structures" introduced in [Hansel, 1966] and [Tonoyan, 1976] into the frequent itemset generation algorithm. [Hansel, 1966] established the chain split theory. [Tonoyan, 1976] invented an excellent chain computation framework which brings chain split into the practical domain. Chain Split Data Mining integrates these technologies around the rule mining procedures. Effectiveness of this approach is related to the low complexity of rules mined, which is a usual case for many applications. Complexity of the procedure composed is complementary to the known Apriori family of algorithms.

In general, the "Affinity Analysis" (or association rules) technique discovers which events are likely to occur together. It is a data analysis and data mining technique that discovers co-occurrence relationships among natural events or, human activities performed by (or recorded about) specific individuals or groups. In general, this can be applied to any process where agents can be uniquely identified and information about their activities can be recorded [Shen et al, 2005].

The "class association rules" (CAR) algorithms have its important place in the family of classification algorithms. As it is mentioned in [Zaïane and Antonie, 2005], the advantages of associative classifiers can be highlighted in four major ones:

- the training is very efficient regardless of the size of the training set;
- training sets with high dimensionality can be handled with ease and no assumptions are made on dependence or independence of attributes;
- the classification is very fast;
- the classification model is a set of rules easily understandable by humans and can be edited.

The first associative classifier CBA was introduced by [Liu et al, 1998]. During the next decade, various other associative classifiers were introduced, such as CMAR [Li et al, 2001], ARC-AC and ARC-BC [Zaïane and Antonie, 2002], CPAR [Yin and Han, 2003], CorClass [Zimmermann and De Raedt, 2004], ACRI [Rak et al, 2005], Arubas [Depaire et al, 2008], etc.

The idea in CAR-algorithms is relatively simple. Given a training set with transactions where each transaction contains all features of an object in addition to the class label of the object, the association rules are constructed, which have as consequent a class label. Such association rules are named "class association rules" (CARs). Generally, the structure of CAR-algorithms consists of two major data mining steps:

(1) An association rule mining stage

(2) A classification stage, which uses the mined rules from the first stage directly.

During the first stage, several techniques for creating association rules are used, which mainly are based on:

- Apriori algorithm [Agrawal and Srikant, 1994] (CBA, ARC-AC, ARC-BC, ACRI, Arubas);
- FP-tree algorithm [Han and Pei, 2000] (CMAR);

– FOIL algorithm [Quinlan and Cameron-Jones, 1993] (CPAR);
– Morishita & Sese Framework [Morishita and Sese, 2000] (CorClass).

Generating association rules can be made from all training transactions together (such it is in ARC-AC, CMAR, CBA) or can be made for transactions grouped by class label (such it is in ARC-BC), which offers small classes a chance to have representative classification rules.

In order to reduce the produced association rules, pruning in parallel with (pre-pruning) or after (post-pruning) creating association rules is performed. Different heuristics for pruning during rule generation are used, mainly based on minimum support, minimum confidence, and different kinds of error pruning [Kuncheva, 2004]. In post-pruning phase, criteria such as data coverage (ACRI) or correlation between consequent and antecedent (CMAR) are also used.

In the classification stage, three different approaches can be discerned [Depaire et al, 2008]:

(1) Using a single rule

(2) Using a subset of rules

(3) Using all rules

An example, which uses a single rule, is CBA. It classifies an instance by using the single best rule covering the instance. CPAR uses a subset of rules. It first gathers all rules covering the new instance and selects the best $n$ rules per class. Next, it calculates the average Laplace accuracy per class and predicts the class with the highest average accuracy. CMAR, ARC-AC and ARC-BC use all rules covering a class to calculate an average score per class. CMAR selects the rule with the highest $\chi^2$ measure from the candidate set, whereas ARC-AC and ARC-BC use the sum of confidences as score statistics. Different approach is proposed in [Coenen and Leng, 2004], which suggests to consider the size of the antecedent and favor long rules before making an allowance for confidence and support.

During the pruning phase or in classification stage, different ranking criteria for ordering the rules are used. The most common ranking mechanisms are based on the support, confidence, and cardinality of the rules, but other techniques such as the cosine measure and coverage measure (ACRI) also exist.

The "class association rules" (CAR) algorithms have their important place in the family of classification algorithms. The advantages of associative classifiers can be highlighted in several very important directions, such as very efficient training; possibility to deal with high dimensionality; no assumptions for the independence of attributes; very fast classification and the result that is easily understandable by humans. The latter two advantages make CAR algorithms an irreplaceable assistant in the processes of disaster risk management, where fast reaction and reliability of the systems are crucial.

## Discussion

The main idea of association rules mining is to discover regularities in the incoming data. Such approach is very important in Global monitoring for environment and security. The incoming data in GMES is as a rule in huge volume and finding regularities in it is critical to discover incoming disaster.

In data mining there are two key factors for success. The first is to determine precisely the problem to be solved. Bundled demand usually leads to best results. The second factor is the use of appropriate data. After choosing from the available data, or the likely purchase of external data, you may need them to be transformed or combined in some way. Retrieval does not give automatically solutions without initial direction [Two Crows Corp., 2005]. Moreover, although the best tools for data mining avoid intricate statistical techniques, it is necessary to understand how the selected tools and algorithms on which they rely.

Data mining is a tool, not a magic wand, which will not sit in the database and send e-mail, to draw attention to some interesting dependency model. These rules are structured in a certain way to retrieve data from the system, which eliminated the need to know the business to understand the data and analytical methods. Each method has its advantages and disadvantages. One of the main disadvantages of data mining association rules is that the support-confidence framework often generates too many rules. The advantage of associative rules lies in their understanding – they are similar to the natural language.

The association rule mining is very important technology for the GMES. It is clear, the floods and landslides are connected disasters, and finding regularities between them is very important. The same we may say for the drought and forest fires. Finding regularities make possible the recognition of the incoming disasters. This is true for anthropogenic collapses, where the regularities in the collected data from great number measurement devices may show the incoming collapse.

Arising from the field of market basket analysis for discovering interesting rules from large collections of data [Agrawal et al, 1993], the association rule mining easily finds its applicability to model relationships between class labels and features from a training set [Bayardo, 1997]. Since then, many associative classifiers were proposed, mainly differing in the strategies used to select rules for classification and in the heuristics used for pruning rules.

The short overview of available algorithms and used structures shows the variety of decisions in association rule mining. As we can see graph structures, hash tables, different kind of trees, bit matrices, arrays, etc., are used for storing and retrieving the information.

Each kind of data structure brings some benefits and bad features. Such questions are discussed for instance in [Liu et al, 2003] where the comparison between tree-structures and arrays is made. Tree-based structures are capable of reducing traversal cost because duplicated transactions can be merged and different transactions can share the storage of their prefixes. However, they incur high construction cost especially when the dataset is sparse and large. Array-based structures incur little construction cost but they need much more traversal cost because the traversal cost of different transactions cannot be shared.

We consider the data processing as intelligent if it is referred to any rational activity. Particularly, functions of comparison, estimation, generalization, systematization, aggregation and decision-making are realized in the intelligent data processing – in difficult conditions, when there is a lack of information, time etc. The application of intelligent processing in environmental problems/tasks is conditioned at least by three groups of factors:

- first, this information is complicated: it comes from various sources, in a considerable volume and irregularly, it has various nature, ranges, and units of measurements;
- second, data that is being collected relates to the significantly important areas – monitoring the quality of water and air, danger of fires and level of ground waters, weather and climatic anomalies, sharp dangerous situations and "creeping" accidents, etc.;
- third, processing of this information in general requires application of such operations, as estimation, comparison, analysis, classification etc.

All this area, i.e. the specified information, its collection and processing, its analysis and decision-making, management and taking actions, control and estimation of results – all this is in general poorly structured, has more likely qualitative rather than quantitative nature and therefore is insufficiently formalized, and is beyond usual methods of representation and computer processing.

Within the last decades, researches and application results of new methods of work with complicated information in difficult conditions were started in a variety of areas (management, economics, business, computer science, etc.). They have the name of intellectual methods, and it is not always a metaphor.

We may define the intelligence as "ability to think abstractly", as "ability to operate effectively in the present conditions", "ability to react correctly to certain problems", "ability to study", "ability to receive knowledge from experience", "skill to get the abilities that lead to desirable results", "ability to adaptation", etc.

In Webster dictionary, the dictionary of authority, there is the following interesting definition: intelligence is an ability to be taught (to be learned) or to achieve the comprehension due to experience. Another definition from the same source is to have a ready and quick apprehension.

In Oxford dictionary of current English (ed. A. Hornby) which is an excellent source, the following definitions (besides others) are presented: intellectual – having or showing good reasoning power; intelligence – the power of perceiving, learning, understanding, and knowing.

About the intelligence, Schlesinger and Hlavach wrote, "You and we altogether should not see such a great nonsense in that one can learn about something, which has never been observed. The entire intellectual activity of individuals, as well as that of large human communities, has for long been turned to those parameters which are inaccessible to safe observation. We will not be speaking about such grandiose parameters as good and evil. We will choose something much simpler at first glance, for example the temperature of a body, which is regarded as an average rate of motion of the body's molecules.

The path leading to knowledge about directly unobservable phenomena is nothing else than an analysis of parameters which can be observed, and a search for a mechanism (model) explaining the relations between the parameters. This means an effort of exploring the relations between the observed parameters and the impossibility to explain them in another way (or more simply) than as an existence of a certain unobservable factor that affects all the visible parameters and thus is the cause of their mutual dependence. Recall astronomers who have been predicting a still unobservable planet by encountering discrepancies in observations from assumed elliptical orbits of observable planets since Kepler laws have been known. Such an approach is a normal procedure for analyzing unknown phenomena. The capability of doing such exploring has since long ago been considered to be a measure of intelligence" [Schlesinger and Hlavach, 2002].

The reality cannot be given in one definition. There exist many definitions of the concept "***intelligence***". For instance, a definition given from practical point of view is given in [Fritz, 1997]: the "intelligence is the ability to reach ones objectives. A system is more intelligent if it reaches its objectives faster and easier. This includes the ability to **learn** to do this. The intelligence of a system is a property of its mind. The mind is the functioning of its brain. An ***intelligent system*** is a system that has its own main objective, as well as senses and actuators. To reach its objective it chooses an action based on its experiences. It can learn by generalizing the experiences it has stored in its memories. Examples of intelligent systems are persons, higher animals, robots, extraterrestrials, a business, a nation. An ***artificial intelligent system*** is a computer program. We can say that it is like the proverbial black box; it has inputs and **learns** which outputs get the most approval by human beings. It stores experiences in its memory, generalizes them, and thus can deal with new circumstances (new inputs)".

The philosophical and practical points of view are two sides of the same idea. Nevertheless, from these definitions it is not clear what the main characteristics of the intelligence are. We need a definition that is more detailed and in the same time to be universal to cover natural and artificial intelligence.

In general, the definitions of the intelligence are covered by next definition [Mitov et al, 2010], which follows from the General Information Theory [Markov et al, 2006] and especially from the Theory of Infos [Markov et al, 2009].

***The intelligence is synergetic combination of:***

- ***(primary) activity for external interaction***. This characteristic is basic for all open systems. Activity for external interaction means possibility to reflect the influences from environment and to realize impact on the environment. For instance, in Walter Fritz' definition [Fritz, 1997] these are "senses" and "actuators";
- ***information reflection and information memory***, i.e. possibility for collecting the information. It is clear; memory is basic characteristic of intelligence for "the ability to **learn**";
- ***information self-reflection***, i.e. possibility for generating "secondary information". The generalization (creating abstractions) is well known characteristic of intelligence. Sometimes, we concentrate our investigations only to this very important possibility, which is a base for learning and recognition. The same is pointed for the intelligent system: "To reach its objective it chooses an action based on its experiences. It can learn by generalizing the experiences it has stored in its memories";
- ***information expectation*** i.e. the (secondary) information activity for internal or external contact. This characteristic means that the prognostic knowledge needs to be generated in advance and during the interaction with the environment the received information is collected and compared with one generated in advance. This not exists in usual definitions but it is the foundation-stone for definition of the concept "intelligence";
- ***resolving the information expectation***. This correspond to that the "intelligence is the ability to reach ones objectives". The target is a model of a future state (of the system) which needs to be achieved and corresponding to it prognostic knowledge needs to be "resolved" by incoming information.

In summary, ***the intelligence is creating and resolving the information expectation*** [Mitov et al, 2010].

This definition of the concept "intelligence" is a common approach for investigating the natural and artificial intelligent agents. It is clear; the reality is more complex than one definition. Fortunately, there exits good theoretical ground, especially in the areas of pattern recognition and data mining as well as the decision-making. Presented understanding of intelligence is important for realizations of the intelligent computer systems. The core element of such systems needs to be possibility for creating the information expectation as well as the one for resolving it. The variety of real implementations causes corresponded diversity in the software but the common principles will exist in all systems. Summarizing, the artificial system is intelligent if it has: (1) Activity for external interaction; (2) Information reflection and information memory; (3) Possibility for generalization (creating abstractions); (4) Information expectation; (5) Resolving the information expectation.