HYDROMETEOROLOGIC SOCIAL NETWORK WITH CBR PREDICTION

Tomáš Kocyan,  
Jan Martinovič,  
Andrea Valičková  
VŠB - Technical University of Ostrava  
FEECS, Department of Computer Science  
17. listopadu 15  
Ostrava - Poruba  
Email: tomas.kocyan@vsb.cz,  
       jan.martinovic@vsb.cz,  
       andrea.valickova.st@vsb.cz

Boris Šír,  
Michaela Hořínková  
VŠB - Technical University of Ostrava  
Institute of geoinformatics  
17. listopadu 15  
Ostrava - Poruba  
Email: sir.boris@vsb.cz,  
       michaela.horinkova@vsb.cz

Veronika Říhová  
VŠB - Technical University of Ostrava  
Institute of geological engineering  
17. listopadu 15  
Ostrava - Poruba  
veronika.rihova@vsb.cz

KEYWORDS
Case-Based Reasoning, Disaster Management, Case-Based Prediction, Information Retrieval, Episode-Based Reasoning, Social network

ABSTRACT
Human activities are contributing to more frequent natural extremes and climate change, which also come from the atmosphere, water or the Earth's crust. With the increasing development of infrastructure, the impacts of these changes and extremes leave more perceivable damage and increasing loss of lives and property. With the use of modern resources and technology we are able to minimize the impact of these extreme phenomena. There are in fact two main approaches - professional and non-professional - both meant from the aspect of data collection and information processing itself. The advantages of social networks have been increasingly utilized during the natural disasters as a way of communicating important information. This article describes the aim of our research - to create a hybrid system which would both enable the collection of data from the professional and non-professional public as well as communicate with other types of systems, to utilize and then process the data and use the data to predict new dangers. The principle is based on collecting data (knowledge, experience, etc.) from both main approaches to disaster management (professional and non-professional) and then applying this information to achieve new solutions. The practical application of a DIP system shows that it can be used for describing the risk of future natural disasters and enables us to deduce the threat imposed by them. Analyzing this data can help create new solutions in the fight to minimize the damage incurred by these disasters.

INTRODUCTION
Climate change as well as the occurrence of natural extremes, whose sources come from the atmosphere, water or even the Earth's crust, are the results of the natural variability of the atmosphere and the evolution of the Earth. These continual changes, caused not only by natural processes, are increasingly influenced by human activities. In natural ecosystems, these changes and extremes, including their occurrences and effects, are part of their natural development. But with the increasing development of infrastructure these changes and extremes leave more perceivable damage, i.e. loss of lives and property. Ecosystems and human society are both getting gradually equipped to adapt to recent and current climate. But there are fears that further adaptation to accelerating changes will be much more difficult. In a short time it can have big consequences for the fundamental values of life, the food or water system or public health, especially in many underdeveloped and economically disadvantaged countries. For almost all of the world there is an increased risk of extreme weather, with subsequent increased risk of disasters associated with them. As stated by the World Meteorological Organization, over the past 30 years, nearly 7500 natural disasters worldwide took the lives of more than two million people. Of that 72.5 percent of the disasters were caused due to weather, climate change or water hazards. In the context of the ongoing process of global warming and global climate change, the question arises to what extent these changes affect meteorological and hydrological extremes causing floods.

Floods are one of the most significant natural extremes in the Czech Republic, and are largely the result of anthropogenic factors, as well as meteorological and physical-geographical factors. Anthropogenically bur-
dened landscape is losing its ability to maintain stability and dynamic balance.

We may not be able to control the winds and rains yet, but in the modern era we fortunately have new tools and kinds of technology, which give us the opportunity to minimize the impact of these extreme phenomena. For example, by providing relevant and comprehensive information for decision support, i.e. creating forecasts for the situation, with the aim of limiting the adverse effects of natural phenomena and their consequences through modern computer and internet technologies. These activities can be supported by both professionals and nonprofessionals in the field, as well as by various levels of public safety administrations, and may reveal solutions for dealing with future disasters.

In fact there are two main approaches to disaster management meant from the aspect of data collection and information processing itself. On the one hand there are some systems which enable the entry of data relative to disaster management and make it possible to trace it afterwards (NEDIES, 2010). These systems are very general and have almost no specialization. They are at the level of chronicles or encyclopedias. There are even systems, which, through the combination of GPS and mobile phone applications, allow for the entry of information about the events right from the location (Ohya et al., 2007). Often these are only data storages which are publicly accessible but with no other use.

On the other hand, there are models individually specialized for each natural phenomenon - floods (Vondrak el al., 2008), landslides etc., possibly for prescriptions and methodologies of how to integrate more miscellaneous models and to create joint interface such as OpenMI (Gregersen et al., 2007). Nevertheless, these systems are not publicly accessible.

Our aim was to create a hybrid system which would both enable the collection of data from the professional and non-professional public, and would be capable of communicating with the other types of systems, to utilize and process their data, and on the basis of this collected data would be capable of predicting new dangers.

Through research it was found out that advantages of social networks have been more and more utilized during natural disasters as a way of communicating important information (Palen et al., 2007). That information can be broadcasted quicker than by way of other news media. Faster communication of information could be the key point for protecting or even saving the lives of the people living in the area affected by a disaster.

It could seem that information provided by people in social networks may not be exact or reliable, but in the fact it is just the opposite. News coming from the administration authorities and big news media are often deliberately distorted (Palen et al., 2007). But it is not only because of that that social networks can help during the crises; they can help to bring people together through the flow of intensive information, and people can solve the problems together and also better resist the catastrophes and recover from their effects (Palen et al., 2007). Besides social networks such as Facebook, Twitter, MySpace and others that are used to that purpose, there is also the emergence of social networks specifically designed for crisis situations.

The first example of them is the social network IGLOO (IGLOO, 2007) which together with its members coming from more than 200 countries and the global connection of varied organizations aiming to solve complicated problems. Recently this network has connected more than 200,000 research workers, academics and specialists from various spheres including education and administration worldwide. Thus, it improves communication and co-operation which results in the better effectiveness of the crisis management and of Rotary’s reaction to the natural disasters.

Another social network is the Gustav Information Center, which was created in the days when Hurricane Gustav was approaching the Gulf of Mexico. This social network provided people with the necessary information to assist in organizing the help of volunteers and also in the evacuation of people before the storm. The network included links to other resources, as well as lists of volunteers, evacuation routes, blogs, photos and videos from the Gulf Coast, and many more resources to help during the hurricane (Edwards et al., 2009).

Microblogging and the Microsoft Vine warning system are other tools which will help to predict disasters such as Hurricane Katrina, earthquakes, pandemics or to manage critical and emergency situations of any kind (MS Vine, 2010). People will choose the field of certain problems and then they will be informed by way of short status messages and security alerts by using either a Vine desktop client or via email. The client will be able to link-up with Facebook and Twitter and will also have the opportunity to monitor the location of his or her relatives on a map background in a similar manner as Google Latitude function in mobile Google Maps for smart phones. Vine is currently available only in America. It should be a sort of system of last communication, which people would use to communicate with family when the telephone or mobile network fails (MS Vine, 2010).

SYSTEM OVERVIEW

In the following paragraphs we will describe a system prototype we have made which currently runs in the testing regime and which is already capable of receiving the inputs from the users. That system is based on the Case-Based Reasoning methodology (Aamodt et al., 1994) and it is described below.

System architecture

An essential element of the structure is characterized as a vertical cross-section of the entire architecture which consists of these six layers: The user enters a natural phenomenon (from a particular territory that has caused
some damage or impairments) into the system. The way the situation is resolved then becomes more reliable.

Based on this collected data, the system is capable of deriving solutions for new cases via Case-Based Reasoning methodology. The accuracy of estimation will be in direct proportion to the increase in the number of cases already correctly entered into the system.

**Phenomena**

One of the key elements of the system is natural phenomena. Based on a specific phenomena’s strength power in given territory, the system is able to search existing cases of similar situations from which potential damages and consequences may be predicted. We have decided that the phenomena in the system will be divided into a hierarchy and stored in the database in order to develop a system modularly and independent of the phenomena. This structure will clarify individual categories making them easier to work with.

Based on consultations with experts, we created a tree of categories. To distinguish between the strength of individual phenomena, we have divided each into several degrees of intensity. These degrees of intensity were determined based on consultations with experts or with the aid of tested methods for gauging the strength of a phenomenon.

A weighted vector is a part of every end category of phenomena and identifies the importance of individual aspects when comparing two cases. Each event includes three weighted vectors that form a hierarchical structure. At the micro level, there is a vector for the surface according to the specifications of Corine Land Cover (Bossard et al., 2000). Seven components with a strong emphasis on the specification of the surface are in the middle part, namely:

- influence of the location of areas of interest, influence of Corine Land Cover as a complex, influence of river networks, influence of the slope, influence of the orientation, influence of altitude, influence of neighboring territories.

The highest weighted vector (or vector for the whole case) then determines: weight of strength significance, weight of phenomenon duration, weight for similarity (on which the phenomenon operates).

**Territory**

Exposure of phenomena to a certain type and composition may naturally have the effect of natural and environmental disasters. As already mentioned, this system will serve to prevent and mitigate these disasters. The aim of this section is to find an efficient method for describing a territory.

The territory is separated into two parts. The first part (raster), determines the composition of a given area or, more specifically: the percentage of individual components of the surface of a given territory. The second part (vector), simply describes the river network, which plays a major role in the operation of any element.

The Corine Land Cover method, which offers various types of surfaces and their locations, is used to describe the territory.

The approximate structure of water is characterized by two indicators - nodes and flows.

**Nodes:** nodes are places where the river branches, runs, or breaks at an angle exceeding the limit angle.

**Flows:** flows are used as vectors starting in FNODE and orientation to TNODE. FNODE and TNODE are labels of individual nodes.

The territory described by a user will be expressed as a vector containing the specific parameters for a given area (the vector describing the territory), and will be presented in the following format.

The first part of the vector is location, to be determined using the S-JTSK coordinate system (CUZK, 2010). The second part of the vector is created by using Corine data, which is made up of individual components of Corine Land Cover, and expressed in percentages (or values in the range of 0-1). These values are identified for the particular field of interest as well as its surroundings. This may affect the territory in which there has been a phenomenon. The third part is the vector river network. As is the case with Corine data, the river network consists of two parts - the area of interest and its surroundings.

River inflow and outflow, river junctions, river branching, and river segments are all searched in a given territory. The slope, orientation, and digital terrain model is used for a more detailed description of the field. These properties constitute the last three attributes of our vector (see Fig. 1).

**Solutions, the consequences and damage**

The solution is a set of measures and other actions bound to a specific phenomenon, leading to the minimizing of consequences or damage. The term “solution” is therefore used to present sandbags, for example, which are used to prevent the flow of water during floods. Since our goal was to create an intelligent system, we did not settle for a mere determination of whether or not a given solution was used in a specific case.

This is why we define the structure as illustrated in Figure 2.

The solution is composed of indicators assigning its jurisdiction to the phenomenon and mainly outlines potential values that may be assigned during this step. If we define the above example with sandbags as a solution (field values), we also determine the recommended height for stacking such bags.

The solutions defined as above are only general rules for a situation (the mere abstraction). To be able to convert this information into a "tangible" form, we define a particular solution that represents the specific action in specific circumstances. The main media information is made up of attributes: the success of solution and index value. Since the minimum and maximum for
each solution is defined, the following formula is sufficient: 
\[ \text{RealValue} = \text{Solution}_{\text{min}} + \text{RatioValue} \times (\text{Solution}_{\text{max}} - \text{Solution}_{\text{min}}). \]

The value ratio was introduced in order to unify the scale of all phenomena, thus providing an improved overview. Maximum and minimum values may obviously provide inconsistent values, thus, the introduction of new, specific solutions to exceed the scale interval is automatically extended and its values are converted.

Consequences and damage. A formal definition of consequences and damages is no longer necessary. Now unwanted phenomena occurring after exposure to a natural element, despite all efforts and measures taken to prevent it, becomes the priority.

Derivation and solution. The system DIP which we develop is not only a warehouse of recorded cases of the past, but is able to intelligently on the basis of experience to derive a solution for the situation which currently threatens. Selecting the most appropriate action is carried out using the methodology of CBR.

Activity (score) of the searched phenomenon can be summarized as follows:

1. The user chooses a list of phenomenon that directly threatens the field of action, intensity, estimated duration and other parameters, described in the scenarios above.
2. The system then monitors all cases, which relate to similar situations, with corresponding information describing the territory.
3. The system monitors all solutions used for the given phenomenon.
4. The deductible matrix is created and evaluates a list of appropriate consequences and damage.
5. The deductible matrix is created and evaluates a list of appropriate measures.

A detailed approach of individual items is described in the following paragraphs.

Observations of similarities. The first step for deriving an appropriate solution is the calculation of similarity (Watson, 1997) found in sample cases which the user entered. This number is expressed in the interval \( \text{CaseSimIndex} \in (0, 1) \), where 0 indicates the maximum diversity and the number 1 symbolizes the identity. The similarity is now calculated on two levels as defined by the phenomenon and its weight vector.

The landscape and its surface structure are derived at the lowest level of similarity, creating the attribute \( \text{LandSimIndex} \in (0, 1) \) unit scope. This value, combined with the duration and intensity of the phenomenon, determines the final congruity of the entire case \( \text{CaseSimIndex}(CSI) \)

\[ CSI = \text{SimilarityVector} \times \text{WeightedVector}, \]  

where \( \text{SimilarityVector} \) consists of the following members:

\( \text{LandSimIndex} \): similarity of the territory,
\( \text{WeightedVector} \): similarity of the length of exposure to the phenomenon.

An important part of the formula is \( \text{WeightedVector} \) where its components determine how important different parameters of vector similarity \( \text{SimilarityVector} \) are
for the calculation. This enables us to easily determine which components can be ignored and which we need to highlight, thus speeding up the calculation.

Creation of deductive matrix. Once we have traced cases and evaluated their similarity, we can create a deductive matrix that will help us to derive specific solutions. The matrix has a form which is shown in Figure 3.

Rows form our searched cases, while columns define all the solutions used for the phenomenon. The last line of the matrix presents the ideal case.

The calculation of the ideal values. Each particular solution is a system characterized by the proportionate value "success" which is defined as the number $SolutionSuccess \in \langle 0, \infty \rangle$. A value of less than 1 indicates that the solution was insufficient, whereas a higher index indicates an unnecessary waste of resources. Ideally calculated values then become the basis upon which the entire algorithm works - seeking solutions for crisis situations.

Derivation of the ideal solution is carried out as follows:

1. For each specific solution in the current column of a deductive matrix:
   (a) Calculate the ideal value and solution for the given event.
   (b) Add to the calculated value the potential impact of a phenomenon’s duration.
   (c) Add to the calculated value the potential impact of exposure to a phenomenon.
   (d) The system derives a weight for the similarity of the earth’s surface to surface similarities of both cases and the weight vector.
   (e) The system derives a calculation for the total weight of the final solution.

2. The system calculates the recommended weight average using collected data and adds the item to a list of recommended measures for the situation.

The calculation of probable consequences. Even after measures are evaluated, it is still necessary to alert the user of the consequences of this type of phenomenon and to what extent it is likely that this situation will affect the user. Derivation is equivalent to the calculation of the recommended solutions of a deductive matrix; the calculation is simplified by the fact that consequences are monitored only as a binary value (whether it happened or not). The result is the number of operations in the interval $AfterEffectProbability \in \langle 0.1 \rangle$, where we obtain the probability of % after we multiply the number by 100.

CASE STUDY
In the system there are inserted data which you can see in the table in a very simplified sample. The table shows a few columns of the projection of many attributes inserted into our case database.

The sample consists of geographical location (latitude, longitude), time specification of the phenomenon (start time and duration), phenomenon details (type and intensity) and a group of affected objects.

An actual version of the DIP system user interface is a simplification of the original DIP design which was presented in (Martinovic et al., 2009). Deployment into the real conditions shows that users are not able to describe situations by specifying several constants and generally by a set of numbers.

The number of details users had to specify was too high and we think that they were intimidated by it. So we decided to adapt input form the capabilities of an amateur community.

The results of our efforts can be seen in the following screenshots. The Figure 4 shows the main application window where you can see inserted events symbolized by pushpins on the map. Events may be filtered and sorted in several ways, such as location or time.

The user can click on one of these pushpins to display a detailed information window of a selected event. The event can be modified, or additional information can be added in this window (see Fig. 5).

However, the user can include his or her own event description (see Fig. 6) with various details of the events, including photos or other additional content.

CURRENT WORK
Tests show (Martinovic et al., 2009) the system provides satisfactory results for the cases where individual events have no time succession, and therefore this succession may not be essential. In that case we are satisfied with the basic techniques of information retrieval and statistics (Bjarne et al., 2003). During the practical application of a DIP system, the need to process not only a single moment but whole time intervals representing individual situations arose. Nonetheless, some information cannot be described by a single time moment; possibly this single moment would give almost no evidence.

Our aim is to provide users with a freely accessible tool which would be capable of helping the public with a progressive description of an event which would last longer and would be inserted into the system. It should be satisfied in the simplest way. These situations will be progressively put together from the individual snapshots and each user will be allowed to enter some parameters, and thus gradually help to describe the whole situation.

We do believe that the social network which will arise within the system will achieve to sufficiently motivate the user to enter details of the events. The exact information in the numerical form, verbal description as well as illustrative pictures will be welcome, both coming directly
from the users or, for example, automatically from other systems. For longer lasting phenomena (meant longer than one moment) the term episode was introduced by (Sanchez-Marre et al., 2005). Each of these episodes consists practically of 2...n cases (so called snapshots) which has the form of film tape consisting of single photos. At least one parameter needs to be fulfilled on each snapshot (see the Fig 7).

It is evident that the denser the series of shots is, the more exact picture about the situation we will have. Via the suitable interpolation or approximation of the points (see the Fig. 8) we are capable of getting close to a real picture of the situation (respectively about the progress of an individual phenomenon) which is inserted into the system.

This extension of the situations’ descriptions leads to a change in the requirements of saving the data and especially their extraction from the database. Thus CBR methodology for these kinds of situations is replaced by its extension called Episode Base Reasoning (Sanchez-Marre et al., 2005), which operates with only these episodes. In the case of successfully obtaining the episodes, the knowledge from the field of time series could be used to add precision to the prediction. At the same time we could thus discover a suitably derived parameter for better indexing and tracing of the episodes.

**DISCUSSION**

Besides the comments directly inserted by the users, the system could be suitably extended by the automatic download of additional information. Data could come
from measurement and observation gauges as well as from other systems. Suitable parameters could be, for example, precipitation depth, temperatures etc. coming not only from professional gauges but also from the amateur meteorological stations which often publish their measurement in accessible sources.

As was stated in the Introduction, the goal of the work was the development of a hybrid system combining the possibilities of the two disaster management approaches discussed above. It will be accomplished by the connection of the DIP system with the FLOREON+ system (Unucka et al., 2009; Vondrak et al., 2008; Martinovic et al., 2008).

Information about the events manually inserted into DIP by its users will be continually and without the necessity of the operation updated by FLOREON+ system, which could be thus understood like another - automatic - user of the DIP system. FLOREON+ is a complex and modular system for hydrologic and environmental modelling (Unucka et al., 2009). System FLOREON+ partly disposes of automated hydro-meteorological data collecting, partly of the automated computational cascade of rainfall-runoff and hydrodynamic models. The hydro-meteorological data is collected from the network of gauges professionally run by the Czech hydro-meteorological institute and the Povodi Odry state enterprise (the River Odra basin management state enterprise) and are used as the inputs of the hydrologic models. On the side of meteorological inputs, data about precipitation depth is most important, followed by data of air temperature and snow pack (thickness and water equivalent). The time step of the obtained precipitation depth and temperature data is 1 hour. On the side of hydrological inputs the data about hourly discharges at the hydrologic gauges are obtained as well. This discussed data is used within the FLOREON+ system as inputs of the calculation of the runoff response to causal rainfall (possibly runoff caused by snowmelt) or as the reference data to optimize the model parameters (to model calibration). The outputs of these models (hydrographs, also in the profiles over and above the professionally monitored gauges and gauges of professional hydrologic prediction) together with the observed discharges are then used as the inputs to the hydrodynamic models which solve the water routing in the riverbeds, and possibly outside the riverbeds during the floods. One of their outputs is the spatial localisation of potential flood lakes. The whole cascade including the postprocessing of the models’ outputs and the web visualisation of the results is the constituent of a fully automated server solution and
nowadays its outputs are used as a basis for decision-making within the disaster management. In the future the FLOREON\textsuperscript{+} system outputs should be among others used as an automatic user of DIP system, which is intended for the level of the near real-time information (update of hydro-meteorological data measured at the gauges, signalisation of achievement or exceedance of the flood emergency degrees etc.) as well as on the level of hydrologic situation prediction (hydrographs and flooding predictions, prediction of achievement or exceedance of the flood emergency degrees etc.).

The classic imperfection of the professional gauges’ networks is their spatial density, which can be insufficient for some purposes (e.g. estimation of convective rainfall effects in the landscape etc.). A possible way to solve these problems can be the use of the precipitation radar measurement outputs. At the level of the gauge network the solvers team works on the automatic collection of the data measured by amateur meteorologists. Organized efforts in linking the hobbyist and the nation’s demand for weather data is not new (e.g. (Bristow et al., 2005; COOP, 2000)). These meteorology enthusiasts could supplement the professional gauge network and they could serve as another user with automatic data update to the DIP system and its social network. On top of that they could provide input data for the hydrologic models of FLOREON\textsuperscript{+} system.

It is evident that data obtained by the amateur meteorological stations measurements are tainted by some amount of uncertainty which could often be quite significant. Measurement of the meteorological factors at the professional level follows the strictly given rules and
standardised procedures and the gauges are calibrated. Localization of these gauges is not random. Amateur meteorological stations can hardly fulfil these conditions. However this is not a problem that could impede their implementation into the DIP system. The information coming from these kinds of users can very suitably and effectively supplement the knowledge of actual conditions, particularly during spatially limited events and possibly localised between the points of the professional observation network. In the question of their potential use within FLOREON+ system the solutions team is aware of the necessity of the homogenisation of this data coming from two different measurement sources (professional and amateur). Thus the use of the amateur data will first need to be carefully and theoretically elaborated and practically tested besides the official use and outputs of FLOREON+ system.

ACKNOWLEDGEMENT

We acknowledge the support of projects SP/2010196 Machine Intelligence and SP/2010101 Integration of GIS and numerical models for analyzing the vulnerability and operational crisis management in relation to selected natural and anthropogenic hazards.

REFERENCES


Microsoft Vine, URL: http://www.vine.net/default.aspx (22nd February 2010)

NEDIES - Natural and Environmental Disaster Information Exchange System. URL: http://nedies.jrc.it (23rd Feb. 2010)


Watson, I. (1999). Case-based reasoning is a methodology not a technology. AI-CBR, University of Salford, Salford M5 4WT, UK