Collaborative Recommender System Development with Ubiquitous Computing Capability for Risk Awareness *

Nick VERCRUYSSEN 1, Cosmin TOMOZEI 2, Iulian FURDU 2, Simona VARLAN 2, Cristian AMANCEI 3

1 ISSCO - International Software Solutions, Bacău, Romania
   nick.vercruyssen@issco.ro

2 “Vasile Alecsandri” University of Bacau, Bacău, Romania
   cosmin.tomozei@ub.ro

3 Bucharest University of Economic Studies, Bucharest, Romania
   cristian.amancei@ie.ase.ro

Abstract: This paper outlines the architecture prototyping and development of ubiquitous computing for recommender systems in case of risk, with mobile and geographical capabilities based on collaborative and knowledge-based filtering approaches. Heterogeneous types of strategies developed towards minimizing risks occurrence effects are applied so as to efficiently allocate the efforts for reliable software development. A novel technique to improve risk assessment and validation at the time of event announcement involving risk type, risk severity level, users trust level, and users location is described and implemented within a dedicated mobile GIS recommender system for risk-awareness. The main purpose of research is to enhance the process of consolidation of the development cycle of GIS applications for early community risk-awareness.

Keywords: Collaborative filtering, Recommender Systems, GIS, Risk management

1. Introduction

Mobile and ubiquitous computing devices are endowed with multimedia capabilities for the geospatial information processing. Information packages are sent by the members of the community to each other or to the authorities and include the geo-coordinate, with the appropriate accuracy, photos from the risk areas and data captured by means of the device sensors. To reduce the effects of the identified risk, geospatial data is processed and added on specific layers by means of classes of objects in recommendation packages for the other members of the community. Furthermore, the cross platform facility of ArcGIS Runtime SDK is implemented for Windows Phone, iOS and Android being available to all community members that actually have a smartphone or a tablet. Mapping functions and geocoding of locations is implemented as well, in order to sustain the spatial risk analysis. Taking into consideration different types of risk notification recipients in the community and the potential increased volume of information packages during alert notifications, the most relevant chronological order of priority is utilized for actual delivery of messages.

Simultaneously, the potential delay is proactively analysed, measured and optimized using initiation of extra cloud computing infrastructure in order to support the acute increase of system performance, within predefined limits.

For the optimization of user experience, a special type of interface design is required so as to overcome the restrictions generated by the size of screen. The user interface has a dual role, to assist the collaborative work for the announcement of a potential or a real observed risk and to automatically advise the community members, mainly from the risk area, in case of need. Trust and reputation is measured by weighted indicators. On the first basis, a small community is considered as target group and further development will expand on larger communities.

An imperative objective of recommender systems development is to look forward to the future generation of mobile oriented software, with a private cloud back-end approach, which has more precise and complex algorithms, large amounts of data, and large-scale data mining functionalities, so as to provide high-quality recommendations and advice to the users.

* This paper is based on two previous presentations at the “Second International Conference on Natural and Anthropic Risks ICNAR2014” with the titles Requirements Analysis and User Scenario for Mobile GIS Recommender Systems Development and Colmars - Collaborative Risk Awareness Recommender System Development with Ubiquitous Computing Capability.
At least two main types of risks are to be managed by the recommender system, natural risks which arise from systems whose existence is beyond the human agent such as landslides, forest fires, floods, extreme weather, and risks derived from human activities, especially technological risks: pollution, severe accidental toxic emissions, explosions, fire, exposed electric wires or damaged gas pipelines. Traffic restrictions, temporarily closed roads or inadequate transport infrastructure can be easily managed and also different types of social risks such as violent and armed individuals are to be minimized. All these types of risks are first identified by common location and common time window as shared contextual information for the mobile recommender system.

Section 2 presents recent scientific achievements in the domain of mobile recommender systems MRS. A non-exhaustive list of popular mobile applications for natural and social risk management is also provided along with a brief description of each application. Section 3 describes the system requirements and architecture and section 4 presents aspects about the risk evaluation and validation of measures. Section 5 describes the ways data is processed and section 6 analyses the required infrastructure for the created mobile recommender system. Next section tackles the integration with other data providers and last section summarizes the work.

2. Background and Related Work

Mobile recommender systems are greatly used in many social and economic activities, among which we may mention tourism, financial activities and shopping. Efficiency is provided by recommender system due to the data filtering and collaboration. They both extract the most significant information in order to provide in a short amount of time, valuable information to the users. Data mining and information filtering are very significant, for the reduction of information overload in recommender systems and for their direct relation with the accuracy and significance of the results. Personalisation is also very important in recommender systems because using filtering or recommending interesting items, users receive relevant information matching their interests [1]. This type of systems are beneficial for countries like Romania, where are regions with high risk of natural disaster [2] or people can be exposed to different risk events.

MRSs are classified in three classes according to the techniques adopted [3]: i) Collaborative Filtering - CF which exploits the user’s feedback on resources; ii) Content-based Filtering – for a given user analyzes the resources to determine which of them are likely to be of interest for him. iii) Hybrid approach: combines the previous two methods to cover their disadvantages. Many of the algorithms used in RS come from the field of machine learning, i.e. algorithms for prediction learning, and decision-making. Similarity between users is usually computed with the Pearson correlation, clustering algorithms are used to organize different entities according to a desirable set of features, while Bayesian Belief Nets, Markov chains, and Rocchio classification are used in recommendation engines.

The MRSs application area which received the largest attention is tourism. Research on tourist guides is focusing on two directions: to find relevant services including transportation services, accommodation, restaurants, information offices [4, 5] and to find relevant attractions like city attractions (historical zones, museums, art galleries) or recommended routes [6, 7].

RSs for media content come into second place: in movies selection, criticism or approval of recommendations are used to provide adequate suggestions [8], when in [9] a semantically based approach is implemented. A hybrid content-based and item-based collaborative filtering approach to recommend TV programs is presented in [10]. Different filtering techniques or data structures are used in [11, 12] to tackle the problem of music recommendation. RS were also developed for document management [13], news or e-learning [14, 15].

The study from [16] proved that mobile technologies are reliable, effective, and affordable solutions for alerting communities in case of disaster. In early 2000, the interest for different mobile alerting systems started to grow and seems to be boosted by the creation in 2004 of Global Disaster Alert and Coordination System (GDACS: now involving UN, EU and disaster managers worldwide) [17]. GDACS aims to improve alerts, information exchange and coordination as soon as possible after sudden-onset disasters by sending SMS or email to prior registered users.
In [18] the use of Synthetic Aperture Radar (SAR) data for natural disaster mitigation in mobile environment is described. The design process of an emergency management system based on mobile communication infrastructure is described in [19] and in [20] is analyzed how to improve emergency announcements to mobile user devices.

An on-line system that informs users about environmental health impact evaluation of air pollution that helps users to make appropriate decisions is presented in [21].

Android market, followed by iOS and Windows Phone are the richest in mobile risk alerting applications both shareware and on-demand ones. Despite the abundance of applications real collaborative systems are very few; in fact the most complete is only ubAlert whose aim is to create a global, highly reliable social network, for all-hazard disaster alerting, by combining data from global institutions and data providers with crowd-sourced user accounts [22]. The majority of alert type mobile applications are dedicated to specific areas like Emergency AUS, Alberta Emergency (Australia), Hazard Alert, AlertFM (USA, Canada), Geohazard (US, Canada, countries along Pacific Ocean, Caribbean Sea, Indian Ocean) or to specific risk: Pacific/Atlantic tsunami, Earthquake Alert, Earthquake Alert, AMBER Alert (child missing). Lot of mobile apps e.g. Tsunami alert, NOAA eRadarHD and Alerts are just interfaces of warning centers, in this case NOAA, or only offer guidelines in case of specific disaster: Disaster Readiness, SAS Survival Guide, wikiHow, Hurricane, Tornado, Wildfire.

Some of the latest winning awards [23] mobile applications are noticeable, but few of them are collaborative RS like ubAlert for example: Quick Disaster – best 2014 apps which runs on Google Glasses and instruct the users about what to do during the natural disaster events, RiskPoint Alert keep the users informed on the latest severe weather events, Jakarta FloodAlert, Disaster Resilience, Anytime, Ehon etc. Regarding to the anthropic risks, such as the risks which come from traffic Inrix and Waze can be mentioned as ones of the largest community based traffic and navigation apps. Through Inrix or Waze different traffic problems are to be avoided, including traffic jams.

Further, COLMARS - a new collaborative risk awareness recommender system is described and its main characteristics outlined.

3. System Requirements and Architecture

In drawing up the development cycle of recommender systems, the identification of requirements plays a decisive role in obtaining quality specifications, accurate and complete, so as to be taken into account for the analysis phase. This purpose presumes that the categories of risks described in previous section, to which are exposed the Romanian communities, can be identified and categorized, and for each particular risk, a user scenario is to be defined. User scenarios, reunited into specialized diagrams, integrate use-cases, actors and interactions between them, as in Figure 1. Furthermore, the end-user risk scenarios are utilized in software requirements.
elicitation, refinement and validation so as to proceed into the development cycle.

The accuracy and reliability of software based recommendations should identify mainly the following categories of beneficiaries/actors and their corresponding roles in the user scenario:

- the members of the local communities, as the most important category of users, directly exposed to the risks and simultaneously the ones who have the possibility of offering help to the authorities, in case of need; trust and reputation of the community members play a decisive role in the decision making.

- the local authorities, which collect information from the citizens by means of mobile devices and specialized software; data stored on servers may be also offered to the citizens in a form of data services or service oriented architecture data flows, which the community members may access by their mobile devices;

- the national and regional institutions which, in case of necessity, are directly informed by the other categories of actors about the specificity of a risk occurrence, a natural disaster or an anthropic risk in a precise geographical location and with prediction facilities for any further resource allocation, aimed at the minimization of risks.

- the trustworthy information sources, as actors from the field, which are to verify the correctness and the accuracy of data provided by the citizens, in a collaborative way; each recommendation given to the authorities and to the citizens is based on collaborative filtering and social media channels with mobile user interfaces;

- knowledge base administrators, whose main duty is the management of historical data about the occurrence of any particular risk for a well-defined community; queries and filtering operations are implemented in correspondence with actual data from the social media channels;

- software developers and GIS specialists from various domains of activity; their main challenge is to harmonize the evolution of software to the new social and technical requirements; consequently, software will become more useful and accurate in the process of automatic decision making.

In Figure 1, the actors are placed in the context of a recommendation use-case scenario, in which their roles are identified and translated into use-cases. The overall set-up may be divided into reduced form scenarios, so as to precisely define the requirements of the software application in the analysis stage of the development cycle.

Taking into account the standard recommender systems classification the further analysis describing our hybrid recommender system is based on the knowledge base management and the collaboration cases in order to finally obtain accurate risk classification and proper recommendations for the community users. These relations and entities practically describe the architecture of the system.

**Knowledge base management scenario.**

Presented in Figure 1, includes the processes of data acquisition and storage, as a direct outcome of the historical background and user experience. Previously stored similar recommendation sessions are used by a prediction technique named case-based reasoning [24], to retrieve and reuse information stored in a case-base in order to speed up time to make a recommendation. Similarity in this case implies the same type of risk in the same area.

Further predictions and consequent recommendations are greatly being supported by reliable data sources, which include besides the date and time of the past natural and anthropic risks in certain areas, duration, impact, georeference, and procedures which were coordinated by the authorities with support of the community members. The knowledge base management involves large software and hardware resources, which should mainly be accessible to mobile users by means of private cloud services or high performance servers. Concurrency plays a significant role, regarding the regular updates on the knowledge base, as well as for the granting of access to the users, on specific areas, in a very short amount of time, in order to minimize the effects of the risks by sending the recommendations in time to definite geographical communities. Filtering and risk management assumes that because of the multitude of risks and areas exposed to risk, information overload may occur. Consequently, the automatic filtering of information is necessary for the elaboration of accurate recommendations. The authorized community
members are granted to the right of sending messages by mobile application and further describe the event, consisting of the choosing risk type, from a predefined list of risks, assessing risk level on a five degree risk impact scale from minor to severe, sending the geo-coordinate, and eventually texting the description of the risk, or posting multimedia files such as images, video and sound. The announcements involve collaborative filtering since at least two different persons/actors should announce and confirm the same risk within a given area described by spatial-temporal features like CenterDistance or AverageDistance, time and duration. Knowledge base must be invoked to obtain historical background of the area concerning risk occurrence.

The AverageDistance feature is used to estimate the average distance between two announcements viewed as GPS positions related to the nearest linear shape segment such as road or river. The CenterDistance feature computes a possible center of risk from the locations of members who announce the same risk and also estimate the radius of the circle area within all members will be notified. This radius is related to the estimated type and impact of risk and has to cover at least the distance between the farthest announcer and the risk center. Acceptance of the risk for analysis implies that all the reports should occur within a specific time window and the users who made the notifications need to have a certain degree of trust. Each user has a trust score which quantifies his reputation within the community members. Trust score are managed by knowledge vectors [24] that describe user’s knowledge about the risks they announced. This leads to the increase or decrease of the impact of their evaluation on the risk’s total score. The entire sequence diagram for risk classification and recommendation is given in Figure 2.

**Personalization** is based on data filtering and assures that appropriate information is given to a community of users which have common features, for instance a common geographic location, the same intentions or the same objectives about a specific activity. In this case, the application offers personalized advice to the users which belong to certain communities and share the same characteristics in order to minimize the effects of risk occurrences. From other perspective, the entire risk RS is personalized by integrating the actors which are specific to Romania or to any other country, such as the trustworthy information sources like inspectorate for emergency situations, the National Institute of Meteorology and Hydrology. These kinds of sources can directly contribute to information dissemination by using the mobile RS since recommendations they provide does not require any processing.

### 4. Risk Evaluation and Validation of Measures

A set of indicators, $SI^R$, for each category of risks, $IR_i$, defined by (1) is considered:

$$SI^R = \{IR_1, IR_2, \ldots, IR_n\}$$

(1)

Every indicator has an appropriate range of values, which determine the risk level from moderate to severe, based on trustworthy information sources and authenticated users from different locations sharing the common area where the risk has impact. In (2) the

![Figure 2. Sequence diagram for risk classification and recommendation](http://www.sic.ici.ro)
Users which identify the risk classification of the data is an important term without manual validation or corrections, circumstantial data needs to be processed fast, template, where data is provided both actors is based on a predefined informational classification. A user defines one or multiple recommendations, which are transformed into corresponding weights \( w_j \in (0,1] \). Therefore, the indicator takes the form of a weighted average (3), which has the possibility to be ordered in the following stages of analysis [25], where \( NRUSR \) represents the number of users \( US_j \), which identified the risk \( i \) in a certain area \( k \), based on their geo-coordinate values, and \( w_j \) is the correspondent user level of trust.

\[
IR_{i,k} = \frac{\sum_{j=1}^{NRUSR} w_j US_j}{NRUSR} + H_{i,k} \tag{3}
\]

For each user, the trust level is calculated considering user’s historical background: his involvement- contribution to RS collaborative filtering, expressed by the number of events confirmed/infirmed and the accuracy of the descriptions of the events he announced. Thus, user trust becomes a reliable measure since is based on a reward-penalty mechanism. The term \( H_{i,k} \) is related to case-base management: a positive score of previous occurrences of risk having the same type \( i \) within the \( k \) area, 0 otherwise.

5. Data Processing

5.1 Subscriber data, preferences and new risk data entry

Subscribers of the recommender system are able to define limited notification conditions, based on geo-location and risk level classifications. A user defines one or multiple geo-locations of interest for wherein he wants to receive risk notifications.

The creation of new risk events by different actors is based on a predefined informational template, where data is provided both automatically and manually. Since the circumstantial data needs to be processed fast, without manual validation or corrections, classification of the data is an important characteristic. For mobile recommendation, the data template completed by a human actor contains: a descriptive title, limited in size and describing the risk in a few words; a risk category, defining the risk classification, impact and potential measures; a risk level, in relation and based on the risk category, however, manually adjustable by the actor. Risk notifications are only relevant for a certain geo-location and duration; the initial values of the parameters are defined by the risk category and are gathered during the alerting process from the users which are subscribed. When risk events are created using smartphones, the metadata of a new risk event will contain the location of the user, defined by A-GPS or GPS systems present in the mobile device.

5.2 Data validation by trust and reputation management

A data validation algorithm, protecting the relevance of the risk notifications and thus maintaining a correct usage of the system by its subscribers, controls the credibility and accuracy of the recommender system. This data validation algorithm is based on the metadata of a risk event and the trust and reputation level of the users when creating and confirming the events, resulting in a continuous changing probability score, defining the probability of a risk being genuine. Every new subscriber in the system receives an initial trust level, which is adapted by each creation, confirmation or denial of risk events. The algorithm is furthermore dependent on the location of the author and the knowledge base of risks for that geo-location in relation to the risk classification. The initial probability score is used to generate an initial limitation of notification of subscribers, for whom the user preferences apply. This limited group of subscribers will then be notified of the risk and asked to confirm or deny the risk. Each human validation will contribute to the data validation algorithm and potentially increase or decrease the risk probability score. When the risk probability reaches the inner or outer limits, the risk notification process will stop or alert all subscribed users. During the validation and continuous altering of the probability, defined thresholds will trigger a further narrowing or expanding of the initial subscriber limitation.
5.3 Delivery priority of notification

The recommender system collects geo-location information of the subscribers, based on A-GPS and GPS data provided by the smartphone usage, only during active usage of the mobile application, limiting the geo-awareness of the system on the subscribers’ location. Priority of notification to subscribers is therefore based on the historical geo-location information stored in the system and the potential presence probability of the subscribers, in order to prioritize alert notifications during high loads.

5.4 Study case

In example from Figure 1, a street fight is announced by member $i$. If the estimated potential risk is low, a predefined area of 200m radius is assigned to the event, centered on the member $i$. If another person $j$, confirms the event within the area, this coverage remains the same. Instead, if the user $j$ is close to the scene but outside the previous zone, the radius is updated considering the new centroid of points given by users’ location, extended with 50%.

![Figure 3. Example of interaction in case of risk and the way the coverage zone is updated](image)

Thus, the coverage area is permanently updated even when the source of the event is moving, and recommendations will be further provided to the users.

6. Infrastructure

The infrastructural characteristics of the system are based on a variety of interfaced subsystems, such as mobile smartphones through a mobile application, push notification dispatching systems, browser based applications, interfaces with third party data providers and third party notification collecting systems. All sub-systems are interfaced through a secured and highly scalable API interface, providing data transfer capabilities between the different data consumers and providers. In order to optimize the data transfer duration and load balancing capabilities, localized gateways are providing an improved infrastructure.

The system contains a scalable central data storage component and asynchronous data processing modules, provisioned on a hybrid cloud infrastructure. Since data storage contains historical data, consolidated in a functional knowledge base component and multimedia data, scalable data storage is required. The cloud infrastructure provides endless storage scaling while maintaining data accessibility performance, using caching techniques.

The data storage is enhanced with data consolidation and metadata abstraction, facilitating an improved data retrieval performance. Different data storage subsystems are in place to store the different types of data, such as multimedia uploads, geo-locations, risk event history and geographical metadata. Each subsystem is optimized for the storage of its specific data formats, and retrieval requirements.

![Figure 4. COLMARS – infrastructural characteristics](image)
provisioning techniques ensure the activation of extra computing resources, facilitating the potential increase in infrastructural needs due to new risk creation or risk notification processes.

7. Integration with External Data Providers

The system foresees integration with third party knowledge base providers; through direct data interchange standards and data collecting processes. The incoming data is transposed contextual and rated on probability and validity using rating mechanisms. Third party knowledge bases can contain already validated risk information – through risk assessment agencies, news agencies, national public relation agencies- or suggest potential risk events, such as social media data content monitoring. Although each data provider is rated initially, based on the source credibility, a continuous rating is maintained through validation mechanisms within the recommender system, in order to preserve the quality of the internal system and its data.

External data can contain risk information as well as contextual information, useful for extending the internal knowledge base, such as updated geographical data, climate information, human mobility patterns and localised events which increase the probability of certain risks.

8. Conclusions

Risk recommender systems have great both economic and social impact - it could save lives or important amount of time and money. Despite this recognized importance, only few such systems deals with a wide range of risks. Most of implementations are just interfaces of warning centers and are usually focused on a single type of natural hazard. This paper describes a new collaborative mobile recommender system for risks awareness, which enables a good management for a wide spectrum of risks. Its reliability is given by the fact that the recommendations are provided exclusively by the actor’s collaboration and also by the fact that the announcement of an event usually have to be confirmed by the users located closely to the event's zone and is based on users confidence. This certifies that the system is not polluted by other non-accurate information channels.

The requirements and the architecture of the mobile GIS recommender system are defined considering subscriber data and preferences, new risk data entry formats and specifications, and taking in consideration the data validation and reputation management abilities.

The paper pointed out also the utilization of priority on notification delivery based on the predictive geo-location of the subscribers. The infrastructure of a mobile risk management GIS recommender system is highly dependent on the data processing characteristics, data storage formats and requirements such as scalability and interoperability of other knowledge base systems.

The benefits of hybrid cloud infrastructure capabilities, combined with a predictive and self-regulated computing management facilitates the acute potential increase in computing resource needs. Interfaces with third party data providers, being continuously validated are an important input for a central knowledge base, assuring a higher level of data accuracy and predictive risk assessment.

Acknowledgements

This research was supported by the project „Bacau and Lugano – teaching Informatics for a Sustainable Society” - BLISS, co-financed by Switzerland through the Swiss-Romanian Cooperation Programme to reduce economic and social disparities within the enlarged European Union.

REFERENCES


IEEE Transaction on Knowledge and Data Engineering, vol. 17(6), 2005, pp. 734-749.


22. http://www.ubalert.com/about_us