Flooding disaster resilience information framework for smart and connected communities

Sathish A. P. Kumar, Shaowu Bao, Vivek Singh & Jason Hallstrom

Journal of Reliable Intelligent Environments

ISSN 2199-4668
Volume 5
Number 1

J Reliable Intell Environ (2019) 5:3-15
DOI 10.1007/s40860-019-00073-2

Springer

Journal of Reliable Intelligent Environments (2019) 5:3–15
https://doi.org/10.1007/s40860-019-00073-2

ORIGINAL ARTICLE

Flooding disaster resilience information framework for smart and connected communities

Sathish A. P. Kumar[1] · Shaowu Bao[2] · Vivek Singh[3] · Jason Hallstrom[4]

Received: 21 October 2018 / Accepted: 11 February 2019 / Published online: 26 February 2019

Abstract

This paper presents the research challenges of designing a combined physical sensor- and social sensor-based information framework to collect heterogeneous flooding disaster data, and then to fuse those data and generate actionable understandings. Our overall objective is to improve the response preparedness of critical infrastructures, contributing to the goal of smart and connected communities. We propose methods to model physical and social sensors, and open demographic data integration with regional knowledge, and to leverage these fused data for understanding impending events and conditions deleterious to lives and properties. In addition, the proposed system will predict the disaster events and provide knowledge-based recommendations to inform emergency management personnel to enable the resilience of the smart and connected communities. Preliminary experiments for the framework are promising. Further work is needed to validate the framework in collaboration with the local emergency managers.

Keywords Sensors · Flooding disaster resilience · Smart and connected communities · Emergency management · Flood vulnerability index

1 Introduction and background

Although currently the coastal communities in United States get real-time notification from National Oceanic and Atmospheric Administration (NOAA) weather radio alerts, they also rely heavily on forecasts from local media for disaster mitigation efforts. With respect to situational information, local eyewitness reports and social media (such as Facebook and Twitter) are also used in an ad hoc manner. City and county governments also have mass notification systems, which allow them to make phone calls to large areas with a recorded message. However, there is no intelligent and automated information and communication framework in place to send the right message out quickly and to the majority of the affected people. Effectively combining the data from multiple sources, such as physical sensor data, social sensor data, and geo-model prediction products, will produce the most complete and high-quality information available. As a flooding condition develops, this information can go out as mass messages to everyone signed up for alerts though social media or phones, so they can respond appropriately to the emergency situation.

This paper explores the research challenges of designing an information framework that couples physical sensors with the social sensors to collect heterogeneous flooding disaster data, and then to fuse that data and generate actionable understandings. The goal is to improve the response preparedness of critical infrastructures, contributing to the goals of smart and connected communities.

In this paper, we explore methods to fuse social sensor data, physical sensor data, and open demographic data with

✉ Sathish A. P. Kumar
skumar@coastal.edu

Shaowu Bao
sbao@coastal.edu

Vivek Singh
vivek.k.singh@rutgers.edu

Jason Hallstrom
jhallstrom@fau.edu

1 Department of Computing Sciences, Coastal Carolina University, Conway, SC 29528, USA

2 Department of Coastal and Marine System Sciences, Coastal Carolina University, Conway, SC 29528, USA

3 School of Communication, Rutgers University, New Brunswick, NJ 08901, USA

4 College of Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

regional knowledge, and leverage the fused data for understanding and predicting impending disaster events. The proposed framework provides knowledge-based recommendations to inform county and city emergency managers. These recommendations will be validated by collaborating with emergency officials and participating in their functional exercises.

The proposed work contributes to the literature in the following ways: (a) it advances the research on design of computational frameworks for integrating heterogeneous streams of uncertain data (including social sensor data, weather data, and environmental sensor data) into actionable situations. (b) These actionable situations will serve in a generic modeling framework that allows community emergency management professionals to model situations of interest and raise alerts based on the vulnerability index generated in real time. (c) Numerical weather prediction (NWP) in coastal environments suffers a lack of adequate ground-truth observational data to realistically initialize the geo-models and validate their results. The physical sensor and social sensor modeling suite proposed in the framework allows for more accurate detection and prediction of disastrous weather events and their interactions occurring in the complex coastal environment. (d) The utility of this model suite will help advance the science of coastal and marine systems, especially with respect to the novel attempt to extract STTPoint (spatiotemporal-thematic-point) human activity information collected from social media to improve numerical geo-models.

Rest of the manuscript is organized as follows. Section 2 describes the related work in the literature. Section 3 describes the research objectives. The proposed framework is presented in Sect. 4. And Sect. 5 discusses the research challenges. Section 6 shows the preliminary experimental results. Section 7 recommends future works. Finally, the summary and conclusion are presented in the Sect. 8.

2 Related work

Work on understanding evolving situations has been undertaken in multiple areas including pervasive/ubiquitous computing [33], situation awareness [46], middleware systems [23, 24, 43–46], Geographic Information Systems (GIS) [36, 37, 43–45], sensor networks [40, 46–49], active databases/complex event processing [1], multimedia processing [35], data fusion [9], mobile information systems [19], web data mining [5, 18], social computing [10, 41], and mash-up computing [4, 12]. Each area, however, has its own perspective [11, 27]. For example, the field of GIS has created multiple sophisticated tools for geo-physical data analysis. However, temporal aspects and personalized control have been oversighted in these efforts. Active Database and Complex Event Processing efforts on the other hand focus on real-time streaming data (e.g. stock prices [1]) but neglect the role of geo-space. Rather than identifying the similarities and differences from each specific field, we focus on building upon and extending the efforts in all these areas. Specifically, our focus is on the aspects that have been rapidly changing in the last few years: (1) geo-temporal data; and (2) their real-time availability at a wide scale.

GIS and data science have been applied for remote sensing [32], weather prediction [20], disease surveillance [31, 42], and business location analysis [48]. Multiple sub-fields within GIS have also developed newer geo-spatial data models [15], geo-information retrieval techniques [29], spatial information processing techniques [43, 47], and integrated toolkits to visualize and analyze geo-temporal data [34]. However, processing real-time streaming data is still rare in this field, though expected to become more mainstream soon [28, 44]. There is also a growing trend to design tools and techniques to understand real-world situations. While mining search logs, weblogs [5, 45, 46] and microblogs [41] have been shown to be useful for predicting flu spread [18] and earthquakes [41], there is also a growing interest in defining generic frameworks that allow users to visualize and analyze their own data of interest [26, 39]. The proposed work advances the state-of-the-art in designing such generic frameworks with a specific focus on emergency recognition. Numerical geo-models play an important role in modern natural hazard forecasting. A coastal flooding is a complex weather event that involves the components of atmosphere, ocean, estuaries, and rivers. The following are examples of the widely used state-of-the-art geo-models that simulate and predict the above-mentioned components: (a) the atmospheric component of Weather Research and Forecasting (WRF) [73]. The WRF modeling is a numerical weather prediction and simulation non-hydrostatic primitive equation model with comprehensive atmospheric physics parameterization schemes that are used worldwide to simulate and forecast a wide range of weather events including hurricanes and severe rainfall events. (b) The oceanic of Regional Ocean Modeling System (ROMS) [72]. ROMS solves 3D primitive equations that are used widely to study how a given region of the ocean responds to physical forcing such as strong winds. This free surface model is suitable for simulating the SLR and hurricane caused the storm surge and inundation. (c) The surface wave model Simulating WAves Nearshore (SWAN) [8]. SWAN is a third-generation wave action model designed for coastal regions. It uses typical formulations for wave growth by wind, wave dissipation by white-capping, and four-wave nonlinear interactions (quadruplets or ''quads''). It also includes physical processes associated with shallow water such as bottom friction and depth-induced wave breaking. (d) Delft3D [13]. The Delft3D model simulates and predicts the flows, waves, sediment transports, morphological developments and ecological processes in estuaries, inlets, lakes, and rivers. Three-dimensional unsteady flow and transport phenomena resulting from tidal and meteorological forcing are simulated in Delft3D by solving the shallow-water hydrostatic pressure equations.

These models sample the state of the fluid (air and water) at present time (initial condition) and numerically integrate the governing equations of fluid dynamics and thermodynamics to predict the state of the fluid in a future time. An accurate initial condition is key to an accurate prediction. The process of incorporating observational data into the model initial condition is called data assimilation. Observational data routinely used by operational weather models include the radiosonde data (launched by weather balloons) and satellite data. In recent years, computing power has been increasing rapidly, but the observational data that can be used to initialize geo-models have lagged, especially for coastal areas, where the weather events are affected by the conditions of the ocean, whose observational stations are scarcer than on land. In this paper, we attempted to address this issue of inadequate observational data by adding the data extracted from a more diverse range of sources, including social sensors.

For situation awareness, Ye et al., described the nature and characteristics of different complex situations and discussed the challenges of situation identification [51]. In addition, they reviewed the techniques most widely used in modeling and inferring situations from sensor data [51]. In a related work, A. Coronato and G. De Pietro [52] proposed situation calculus-based approach for the situation awareness, whose outcome is to detect abnormal behaviors of cognitive impaired individuals when they are in situation-aware smart spaces. Similarly, McCarthy proposed a formal theory-based software, that can decide actions based on situations [53]. In his report, McCarthy gave several examples of how the goals can be achieved based on deduction from situation descriptions [53]. In a related work, Coronato et al., proposed formal models to study their appropriateness for specifying and verifying socio-technical collective adaptive systems (CAS) in the wake of CAS heterogeneous parts' complex and unexpected abnormal behavior [54]. Endsley proposed the concept of user-centered design in a book on situation awareness [55], which defines the concept of situation awareness (SA), how to determine SA requirements and how to automate it.

With respect to designing reflective neural networks (RNN) for metacognition and reflectivity, deep learning auto-encoders utilize a structure similar to the one presented in this paper. But their research application focuses on unsupervised learning of convolutional features for generic detection tasks [68, 71]. Anomaly detection has also been addressed through auto-encoders [69, 70]. One related work by Sun et al. proposes how metacognitive processes represented by the experimental data may be captured within a cognitive architecture [56]. Researchers from Nanyang Technological University, Singapore, have developed neural network-based metacognitive learning algorithms for the metacognitive learning framework [57, 59, 63, 66, 67]. Neural networks and metacognitive learning have been utilized for applications such as traffic flow prediction [58], improved perturbation tolerance [61], human action recognition [62], symbolic anomaly detection [64], autonomous robot control [65], and remaining useful life estimation [66]. Another related work by Yeh and Lo [60] present a neural network model that assesses automatically the learner metacognitive knowledge level by observing his/her online browsing behavior. As per our literature review, there is no particular work that applies RNN for metacognition from a disaster resilience application perspective.

3 Research objectives

Based on the knowledge gaps discussed in the previous section, our goal is to propose a framework to improve decision-making for emergency management scenarios with heterogeneous data sources via a prototype of the disaster resilience information framework. Our specific objectives are to:

1. Define and design a novel framework to collect, represent, and fuse uncertain heterogeneous real-time disaster-relevant data for situation recognition and prediction of future events.
2. Define and develop a current and near future Flood Vulnerability Index (FVI) for a given geographic area that is updated real time by the proposed framework.
3. Deploy, run pilot experiments, validate, and refine the above framework and FVI by working in close collaboration with the city and county emergency management department.

4 Proposed disaster resilience framework

Our proposed framework is as shown in Figs. 1 and 2. Following are some of the major components of the proposed framework.

Figure 1 shows the proposed framework for flooding event situation recognition. As shown in Fig. 1, the input to the framework comes from human sensor/actuator, device sensors and stored data. The core of the framework consists of spatiotemporal aggregation, situation detection operators and situation-based controllers. The output from the framework can be analyzed by analyst and the personalized response can be given back to human actuator.

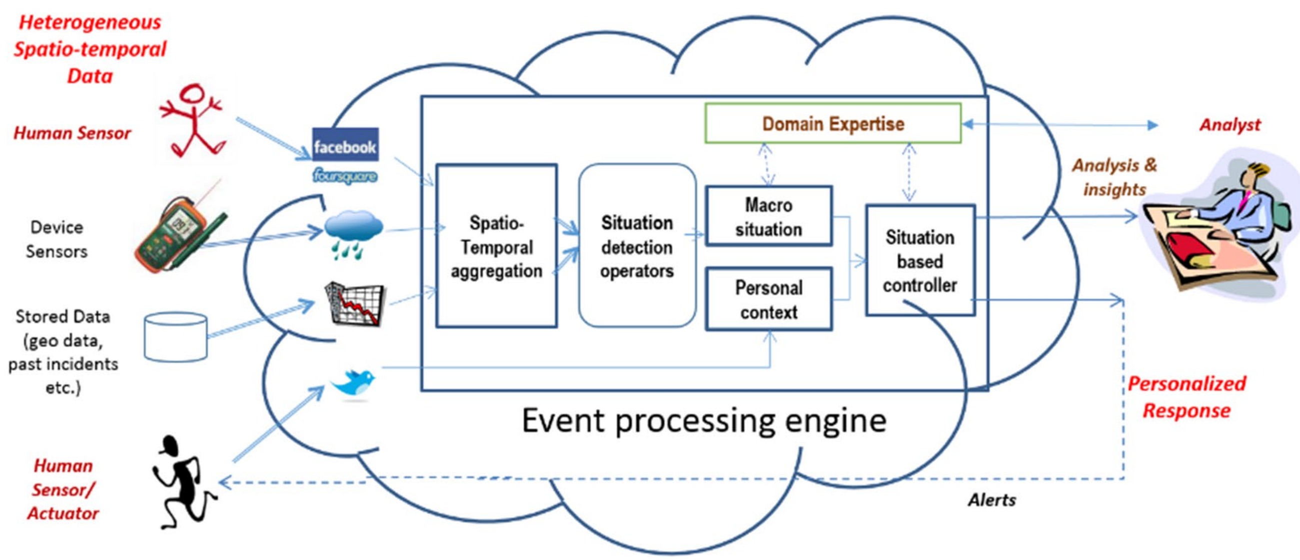

Fig. 1 Proposed framework for situation (flooding event) recognition

Fig. 2 Workflow for situation recognition

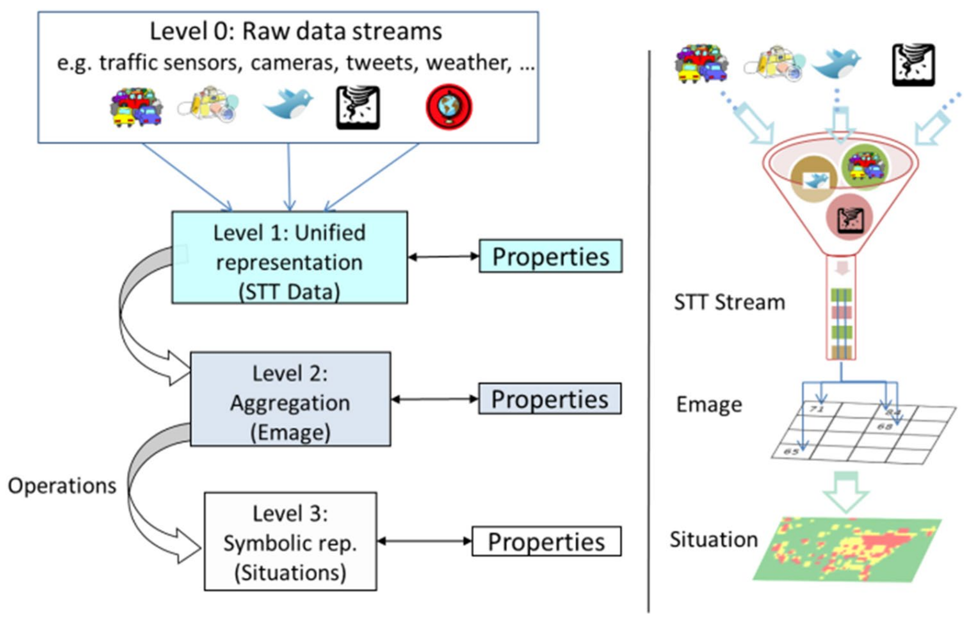

Figure 2 shows the workflow for situation recognition. As seen in Fig. 2, several raw data streams are aggregated to form STT data, which are then transformed to Emage.

The operations are then performed on Emage to define situations through symbolic representation [21].

4.1 Input data from multiple sensors

Building toward generic cyber-physical situation awareness requires identification of different types of relevant data. Different data representation categories [50] have been proposed in the past, each relevant to the usage context. We propose the following data types required for real-time geophysical situation recognition:

1. Stored data: this includes textual data, databases, audio, video content, sensor readings, and curated geographic data such as maps and cartographs.
2. Live sensor readings:

   (a) Regional remote-sensing sensors: these are the sensor data whose coverage area is best represented as a geographic region such as satellite and radar observation.

   (b) Point sensors:

I. In situ device: these are the physical sensor data whose coverage area is best represented as a geographic point such as weather station.
II. Human: these are 'abstract sensors' who report observations as a single point. Human reports, such as tweets, tend to be more semantic and biased.

4.2 Data representation

A data representation needs to be identified that can integrate information coming from different information sources, each of which might have been created by a different agency with a different original purpose. The framework focuses on data representation with space, time and theme, which is grounded in the real world. Space and time are considered as fundamental axes for designing different aspects of the framework (including data representation, operators, and building blocks). This draws upon the basic nature of the physical world and builds on the concepts from the fields of GIS, cartography, satellite sensing, and location-based computing [38]. All incoming data of interest can be converted to and represented in a common STT (space, time, theme) format: STTPoint = <geo-coordinate, time-coordinate, (theme, value)>. All point sensors observe a particular value for an application-relevant attribute (theme) at a certain spatiotemporal coordinate. This representation can be used to capture data ranging from traffic speeds, to air quality level, twitter mentions, and crime reports. Region sensors record values for a theme from a spatiotemporal region, but they can be considered as a collection of STTPoints observed 'in a batch'. For example, a satellite observed geo-image can be considered as a set of pixels (STTPoints) observed together. Hence, the regional observations can be used to derive values for points within its coverage.

The framework considers E-mages [47, 48] and E-mage Streams [17] as its data model. An E-mage is a grid-like data structure, each cell of which captures a value associated with a particular application theme (e.g. temperature, number of flu cases) at a spatiotemporal coordinate. The use of a grid is based on the understanding that grids are the fundamental data structure used by humans to understand and analyze spatial data (e.g. maps, satellite images). They also capture the semantics and the notion of spatial neighborhood very elegantly, and geography-driven-joints [2, 3] between data streams reduce to simple overlaying of grids. The 2D data grid in an E-mage is akin to an image, and this allows for repurposing of a rich collection of image and video processing operators such as segmentation and aggregation for analyzing the spatiotemporal data [47, 49]. Such a representation also aids easy visualization and provides an intuitive query and mental model. A flow of E-mages forms an E-mage Stream, which serves as the primary data model or data representation in the proposed framework. For computational purposes, continuous and ordinal values can also be normalized into a stream of numeric values.

4.3 Recognizing the current situation

Computationally, the translation of raw data to actionable insights requires the use of various concept recognition techniques. In the past, different approaches such as object recognition, scene recognition, and event detection have shown to be useful for different problems (e.g. detecting a 'tree' within an image, or a 'packet drop' within a network channel). However, they focused only on a single media object, media type, observation location, or timeframe. Today, with the growth in the internet of things, social media and planetary-scale sensing, there is an inherent need to assimilate spatiotemporally distributed heterogeneous data streams into actionable information. Consequently, the concepts such as objects, scenes, and events need to be extended to recognize situations (e.g. epidemics, flashfloods). Based on a survey of prevalent definitions across multiple domains [47–49], and our focus on spatiotemporal data, we define a situation as: "an actionable abstraction of observed spatiotemporal descriptors." In the above definition, the expression descriptors illustrate the emphasis on a computationally recognizable concept, e.g. via statistical analysis of some attributes. Similarly, spatiotemporal reiterates our focus on real-world observations. As a computational concept, situations focus only on observable (via human/device sensors) aspects of the world, and need to be actionable abstractions, i.e. those that are explicitly defined by human domain experts to support decisions and actions. Examples of such situations include beautiful days/hurricanes/wildfires/flooding, traffic (jams/smooth/normal), economic recessions/booms. Heterogeneous sensor streams can be combined to derive actionable situations as shown in Fig. 3. The unified STT format employed (level 1) records the data originating from any spatiotemporal bounding box using its numeric value. Aggregating such data results in two-dimensional data grids (level 2). At each level, the data can also be characterized for analytics. The situational descriptor (level 3) is defined by the user as a function of different spatiotemporal characteristics.

4.4 Prediction of future events

The fused data from the previous steps with the unified STT format can be fed to our existing prediction geo-models as their initial conditions. This prediction output along with the current situational data can be used by the city planners and emergency management professionals in a planning context. The unique fused data collected from a diverse range

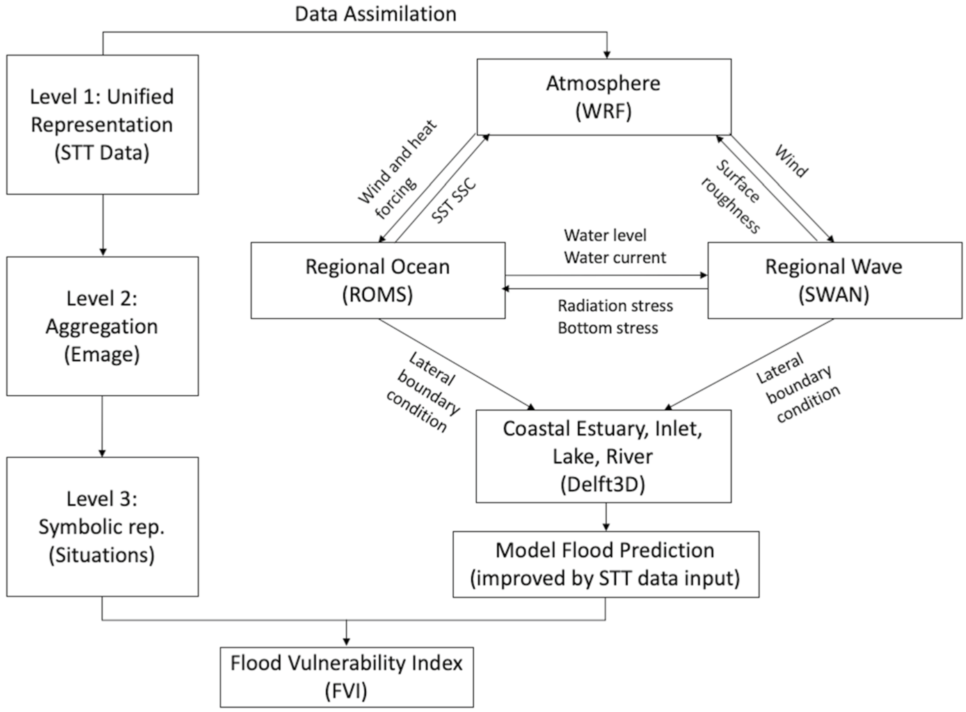

Fig. 3 Diagram of the coupled geo-model. STT data are assimilated into the coupled geo-model and help improve the disastrous weather and flood prediction, which, in turn, help increase people's awareness and preparedness, thus to reduce the susceptibility

of sources, including physical sensor and social sensor data, will help the model more accurately describe the initial state of the situation, therefore, allowing the geo-models to generate more accurate flood prediction [6].

The framework allows for detection of abnormal situations based on models of situations of interest, which are encoded into the system. Following [17], we propose to work with domain experts to define what an abnormal situation (e.g. flood level 3) means in terms of the different data sources available. The high-level situation of interest construct (e.g. flood level 3) is broken down into smaller, more directly assessable features (e.g. water level, number of people affected) in an iterative manner until each of the features identified can be measured directly via an available data source (e.g. water sensor, census data). Once the situation of interest has been modeled it can be converted into a "standing query" that runs over the incoming data streams using an implementation engine such as EventShop described in [17]. An alternate formulation of the above approach is to encode a normal situation (e.g. no flooding) as the standing query and any time the output changes beyond a range an alert can raised indicating an abnormal situation. Last, it is possible to encode the standing query in terms of dynamic properties which are indicative of the impending change in the situation to allow for future predictions. For instance, if the rate of water accumulation per minute exceeds a threshold than future flooding situations can also be detected using the system.

The existing prediction modeling framework is an atmosphere–ocean–wave–river-coupled geo-model to predict hazardous weather events affecting a coastal community. Data sources for these models are the real-time data obtained from the sensors, either physical sensors or human sensors, related to rainfall-inducing weather events. The flooding events considered include those induced by tropical and extratropical cyclones. These coastal weather events and their impacts interact with each other to form a complex system. We have developed a flexible, efficient, and powerful parallel coupler [7]. The coupler has been used to couple WRF, ROMS, SWAN and Delft3D models to form a powerful geo-model for coastal hazardous flooding events. The coupled geo-model framework is shown in Fig. 3. As shown in Fig. 3, in this coupled geo-model, WRF provides the atmospheric forcing to ROMS and SWAN; in return, ROMS and SWAN feedback the upper ocean states to WRF as a part of its surface condition. The ROMS and SWAN domain grids cover a relatively large area to allow more realistic representation of the ocean circulation patterns. A higher resolution Delft3D model grid is nested in ROMS and SWAN domain grid, receiving its lateral boundary condition (LBC) from ROMS and SWAN. Delft3D uses a flexible finite element mesh grid to more realistically represent the complex topography and bathometry in estuaries, inlets, lakes, and rivers. The coastal flooding prediction will be produced by this coupled geo-model. Some fused data, (e.g. pressure, humidity, temperature, wind speed and direction), in STTPoint format, will be assimilated into the geo-model to improve its initial condition's accuracy, using data assimilation methods such as three/four-dimensional variational assimilation method (3D/4D-VAR) or Ensemble

Kalman Filter (ENKF). Other data (e.g. flooded area, flood water level, rainfall, cloudiness, etc.) can be used to early validate the geo-model prediction. The fused social sensor data and the geo-model prediction are used to calculate the FVI index. Vulnerability is defined here as the extent of harm that can be expected under certain conditions of exposure, susceptibility, and resilience. More specifically, Vulnerability = Exposure + Susceptibility − Resilience [16, 25]. The FVI we calculate is dynamical in the sense that it varies for different locations and for different weather events. Exposure is mostly the coastal infrastructure, property, and people that could be affected by the impending flood events; susceptibility is the exposure's likelihood of being harmed by the floods, including the people's awareness and preparedness for the impending floods. Resilience is the coastal community's capability of enduring the flood events. In this study, the FVI is calculated based on the: (1) geo-model's prediction of where and when the floods may occur (therefore, the infrastructure, property and people that may become exposed to the impending flooding events) and (2) fused data from the Level 2 (Emage) and Level 3 (Situations) showing the exposed's awareness of and preparedness for the impending flooding. The coupled geo-model will be used mainly to address exposure and susceptibility by: (1) providing more accurate and timelier flood prediction to identify the exposure, and (2) increasing people's awareness and preparedness, thus reducing the susceptibility. The FVI can be incorporated into the disaster information framework and made available in real time to the decision-makers, emergency managers, and public to improve preparedness and help mitigate the disasters, such as flooding.

5 Research challenges

The proposed approach opens up a number of newer research challenges. Following are few of them.

5.1 Handling data uncertainty

Incoming data may have associated uncertainty with it, which, if ignored, can lead to imprecise or invalid conclusions. The data structures thus need to support data representation with uncertainty and define relevant spatiotemporal operations. To tackle these challenges, we need to take inspiration from the literature in statistics, pattern recognition, and machine learning, which often represent uncertain data as vectors whose univariate and multivariate distributions are defined to model different types of uncertainties. In many applications, owing to the central limit theorem [30], a Gaussian distribution is a reasonable distribution to model uncertainties such as measurement errors or spatiotemporal correlations in data. Operations such as sampling, evaluation, conditioning, and marginalization have also been defined for such Gaussian distributions. Similarly, matrix transforms on multivariate Gaussian's have well-defined analytical forms. We need to bring these well-established concepts from the field of pattern recognition, and extend and adapt them in a spatiotemporal data processing framework. We propose to define a distribution Emage or dEmage to represent a spatiotemporal Gaussian distribution, i.e. we define dEmage as (e;Σe), where e is an underlying Emage with the mean values and Σe is the covariance matrix associated with e. Here, e captures the spatial and temporal structures, whereas Σe captures the uncertainty information. Linear spatiotemporal operations such as sum, mean, and weighted mean can be expressed as matrix transformations on the underlying multivariate Gaussian distribution [22]. Having the ability to represent uncertain data, we can implement several complex operations such as Kalman filtering and other data assimilation approaches where uncertain representations are needed. This allows us the ability not only to determine present situations from observations but also to make predictions for future.

5.2 Modeling situations working with domain experts

The first step toward detecting situations is to represent them. Modeling situations before building applications in hardware and software is important as it allows application designers to focus on 'the big-picture' without getting into implementation details, encourages goal-driven (rather than availability driven) thinking [14], promotes reuse of components across applications, and renders explicit requirement list (e.g. sensors/data sources needed, computation rate required, spatiotemporal resolution required). Hence, the framework needs to provide a generic way for different domain experts to externalize their mental model of the situation of interest [47, 48]. While situations of interest (SOI) can vary significantly, there exist certain common elements that are required to model all the situations discussed in the previous section. Generically, the situations can be recognized by transforming heterogeneous data into a common representation, extracting features from it, and using those features to classify it. This allows us to identify the generic operations required for situation recognition. Similarly, while situations can be vague and ill-defined to start with (e.g. what constitutes an 'epidemic'), all computational concepts can be made more concrete by following a simple methodology: iteratively splitting down a vague concept into its more tractable constituents. As shown in Fig. 4, the value of a modeling approach lies in its expressive power and usability in designing multiple applications. We build upon our prior work on situation modeling [45, 46], and

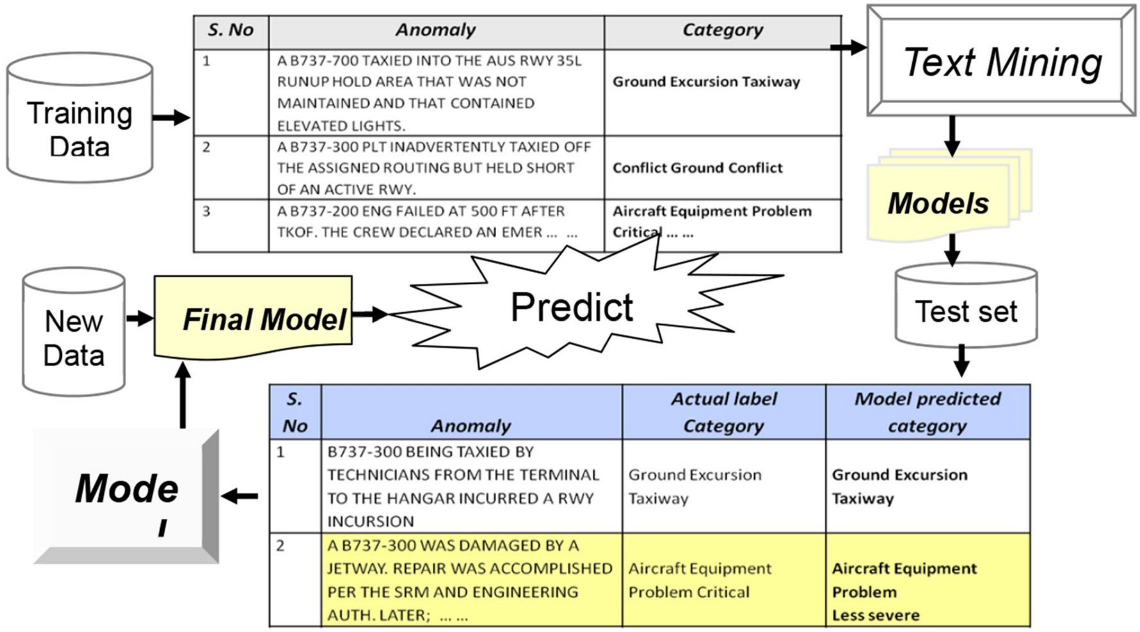

Fig. 4 Workflow of situation modeling approach

validate and refine the modeling approach by working with domain experts.

5.3 Assimilation of the fused data into a geo-physical prediction model

It is a challenge to assimilate fused data into the geo-model to improve its initial condition. The social sensor data, for example, those extracted from tweets, may have a wide range of uncertainties and qualities. How to address the uncertainty and quality through quality control procedures may become a major challenge in the process of assimilating these types of data into the coupled geo-model framework. While the findings from Research Challenge 1 guide us in dealing with uncertainty, the raw data should also be filtered and quality controlled. To tackle these issues, we first need to identify and eliminate the outliers via unsupervised machine-learning techniques. Similarly, conflicting data would be identified by evaluating their statistical distributions to predict their quality and usefulness. Second, vast amounts of data, especially those extracted from social media, may be non-numerical types of data. Because the 3D/4D-VAR and ENKF methods can only take numerical data input, these non-numerical social media data need to be converted into numerical STTPoint format before being used as an input to the weather prediction model, which is challenging. Because most of the social media posts such as twitter have imbedded GIS information, they can be filtered according to their geo-locations. And the information contained in the text or graphics in a social media post can thus be associated with the geo-physical process for that location. For example, when a tweet posted from New York city contains the text of "floods" or photos of flooded streets, it can be used as a binary numerical data "flooded = True" to initialize or early validate the geo-physical forecast model for that location. It is recognized that the colossal volume of the social media data, in millions per minute, makes it impossible to process in real time by traditional methods. However, in recent years, advances in artificial intelligence (AI), machine learning (ML) and natural language processing (NLP) start to provide a promising new approach to tackle this challenge. For instance, Jongman et al. [74] found that the flood-related Twitter activity, when combined with disaster response organizations and the Global Flood Detection System (GFDS) satellite flood signal, can help disaster relief organizations "gain a quicker understanding of the location, the timing, as well as the causes and impacts of floods." To leverage these new advances, United Nations Office for the Coordination of Humanitarian Affairs developed the Artificial Intelligence for Digital Response (AIDR) (http://aidr.qcri.org/) to uses supervised machine learning and artificial intelligence to tag thousands of social media messages per minute. These structured data are then ready for use in dashboards, maps, or other analytics programs to help generate useful information during crisis hazard events from images and comments contained in social media data.

6 Experimental results

In our experimentation, we collected the tweets related to Hurricane Irma and Fig. 5 shows the spatial distribution heat map of the collected Hurricane Irma-related tweets.

The coupled geo-model prediction framework has been used to forecast a wide variety of disastrous weather events and their impacts in the South Carolina coastal regions. The good agreement, as shown in Fig. 6, between the modeled and the observed water level changes during tropical storm Ana (2015) suggest the effectiveness of the coupled geo-model in predicting disastrous weather events and their impacts on the coastal community. While the preliminary experimental results for the existing coupled geo-model framework are promising, our objective in this paper is to improve its predictive accuracy by adding the fused data, including those from the social sensors, as an input to be assimilated into the model's initial conditions, such that the emergency responders are better positioned to provide emergency relief.

We have been working on developing EventShop as a tool for integrating heterogeneous data for situation recognition. This tool was put to test during hurricane Sandy, which affected large parts of the United States and the Caribbean during Oct 2012. We used a combination of forecasted hurricane path, census, and Red Cross shelters to detect the macro-risk level for each location in the United States. Twitter messages mentioning certain hashtags as proxies for personal concern and the system sent back messages to users in risky situations, informing them of the nearest open Red Cross shelter. Some of the tweets were re-tweeted, and replied to by the users, indicating an interest in receiving and sharing such information. Working with county officials, we plan to extend the existing framework and results from the preliminary results to generate real-time Flood Vulnerability Index maps and send back personalized alerts to people in the most vulnerable situations.

The ultimate product from this proposed framework is the Flood Vulnerability Index (FVI) calculated based on "situations", which are derived from the sensors' spatiotemporal-thematic-point (STT) data and their aggregated "Emage", and geo-physical model's flood prediction, which also incorporated the STT data to enhance the model's initial representation. A prototype Flood Risk Index has been developed and validated [75], in which the key factors that affect the hurricane-induced storm surge and inundation in coastal areas are identified as $\Delta\eta = \frac{\rho_a C_d U_m^2 L}{\rho_w h g}$, where $\Delta\eta$ is the storm surge height, ρ_a air density, ρ_w seawater density, C_d drag coefficient, U_m hurricane maximum wind speed, L the fetch of the wind field, h water depth, and g is gravity = 9.8 ms^{-2}.

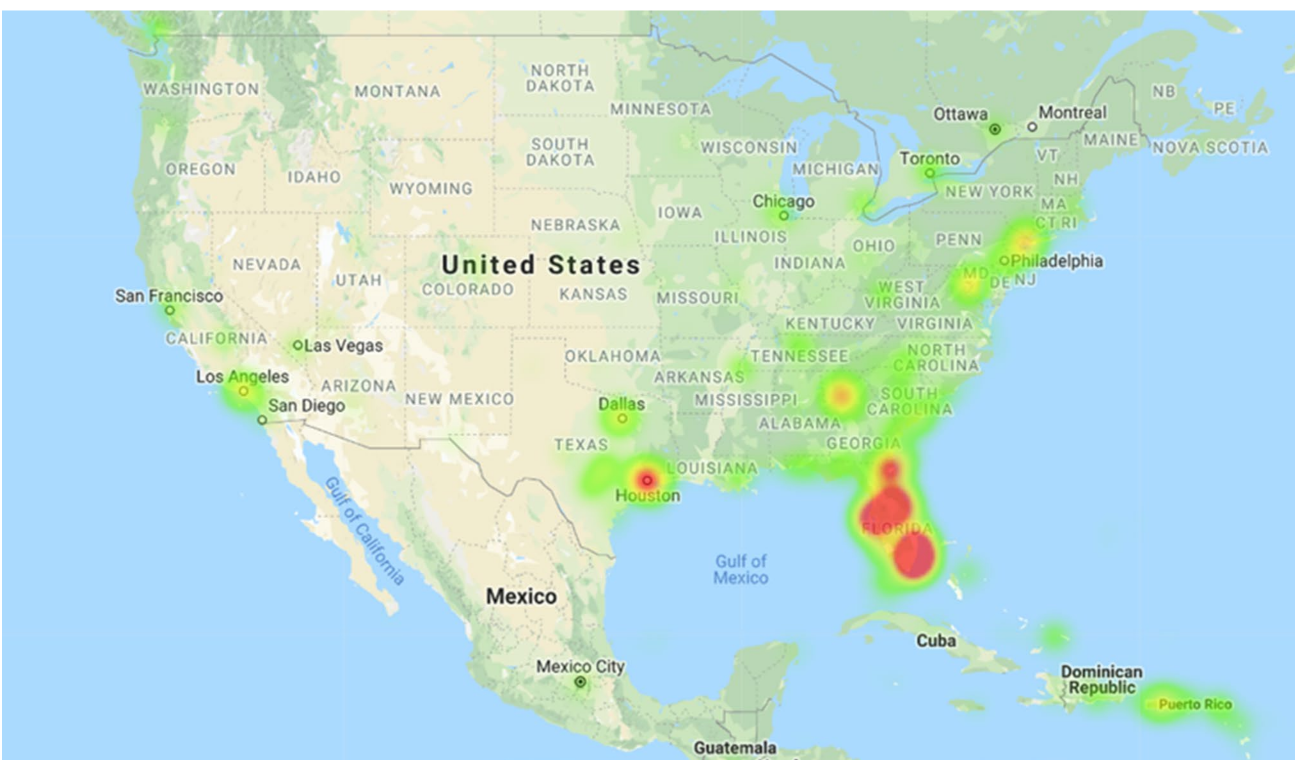

Fig. 5 Heat map showing the spatial distribution of the collected Hurricane Irma-related tweets

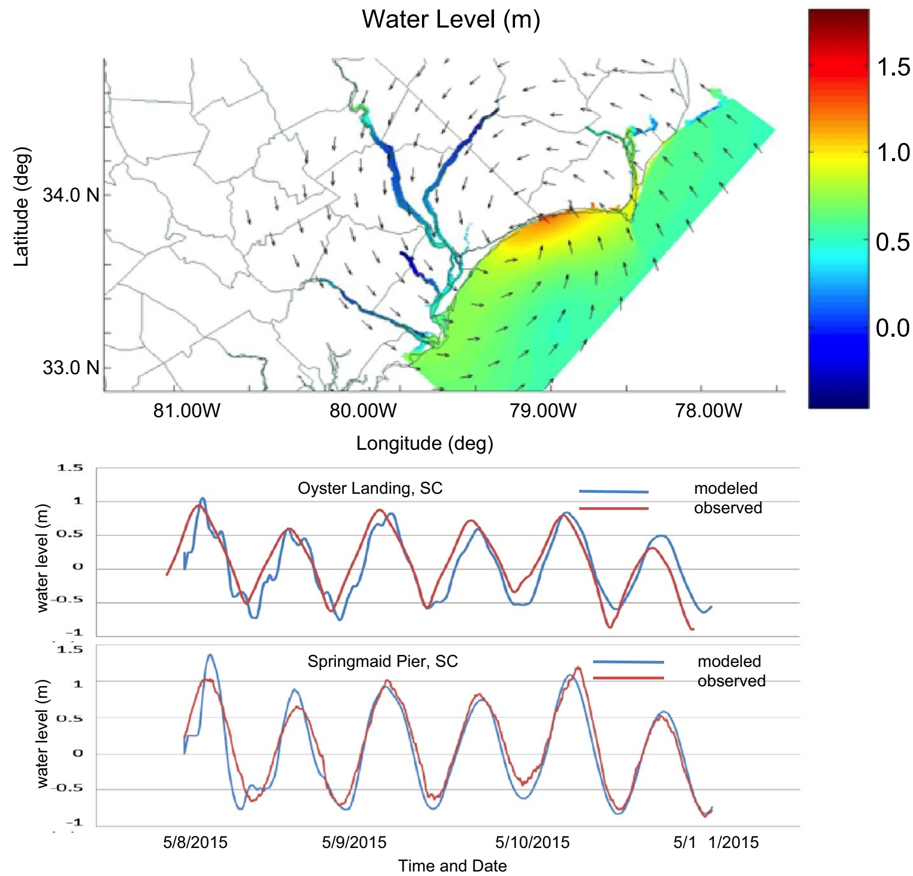

Fig. 6 Example of a prediction of the water level changes (including both tidal and storm surge effects) (upper panel) during the landfall of tropical storm Ana in May 2015 and its validation with NOAA's observational data at two-gauge stations (lower panel)

After the storm surge height estimate is obtained, it can be used to calculate the inundation area. The distance that the water can inundate is given by $I = \frac{\Delta\eta}{\alpha}$, where α is the slope $\frac{\Delta y}{\Delta x}$ of the coastal topography. The inundation area, assuming a semi-circle shape, is given by $A = I^2 \frac{\pi}{2}$. This prototype FVI has the advantage of easy calculation. When the above-mentioned variables are obtained from the physical and social sensors as STT and Emage, they can be used to quickly estimate the situations, representing the storm surge height and inundation extent in this framework, using the FVI formula.

7 Future work recommendations

In the future, this work can be extended to measure and understand the impacts of the extreme events through mining of social and economic media, and to correlate these "soft" impacts with physical impact measures. Understanding social and economic impacts will inform local and state governments, emergency management teams, and decision-makers in planning for and responding to weather events. As an example, the Flood Vulnerability Index described in Section V needs to incorporate social-economic factors such as population and economy development level. The combined application of artificial intelligence (AI), machine learning (ML) and natural language process (NLP) will be a key strategy to extract the flood related information from the vast amount of real-time social media sensors.

The proposed framework can be extended with the design and application of a social media analytics framework designed for assessing post-storm social and economic response. The framework can consist of multiple layers. For example, the first layer could ingest physical storm impact data from existing sensors and modeling sources. The second layer could apply natural language processing to mine social media (e.g., Twitter, Facebook) for social sentiment. The third layer is analogous, applying similar techniques to mine economic media sources (e.g., NYSE, NASDAQ) to evaluate economic impacts. In each case, the datasets would be geo-tagged to enable geo-spatial comparisons across physical, social, and economic layers. Whether persistent relationships exist, and whether those relationships can be leveraged for prediction can be explored.

8 Summary and conclusion

To address the resilience from the disaster situations, we have proposed Flooding Disaster Resilience Information Framework. The proposed research work builds an innovative toolbox for emergency managers and increase the emergency preparedness and resilience of the community. The multiple sources of information utilized by the framework will lead to more accurate emergency situation detection and the most appropriate response actions in a timely manner, which could save more lives, resources, and critical infrastructures. While the preliminary results are encouraging, the proposed framework also improves the efficacy of the community officials in preparing for impending disasters, as they will be able to model, test, and respond to situations using multiple-source uncertain data streams without the need for programming skills.

Acknowledgements The work is partially supported by the National Science Foundation grant CNS-1763294. Authors would also like to acknowledge the support of Dr. Dong Wang, Dr. Brian Xu, and Mr. Ike Vayansky for their help in generating methodology diagram and experimental results. Authors would also like to thank Dr. Zhenlong Li for providing the social media data for analysis.

References

1. Adi A, Etzion O (2004) Amit—the situation manager. VLDB J 13:177–203
2. Anselin L, Syabri I, Kho Y (2005) GeoDa: an introduction to spatial data analysis. Geogr Anal 38:5–22
3. Arctur D, Zeiler M (2004) Designing geodatabases: case studies in GIS data modeling. ESRI Press, Redlands
4. Baheti R, Gill H (2011) Cyber-physical systems. In: The impact of control technology, pp 161–166
5. Bansal N, Koudas N (2007) Blogscope: spatio-temporal analysis of the blogosphere. In: Proceedings of the 16th international conference on World Wide Web, pp 1269–1270
6. Balica S, Wright NG (2009) A network of knowledge on applying an indicator-based methodology for minimizing flood vulnerability. Hydrol Process 23:2983–2986
7. Bao S, Yu Z, Xu J, Yan T, Pietrafesa L, Gayes P (2016) An easy to implement method to couple weather research and forecast (WRF) with other geophysical models and its application on simulating a northwest pacific typhoon. Environ Model Softw
8. Booij N, Holthuijsen LH, Ris RC (1996) The "SWAN" wave model for shallow water. In: Coast engineering proceeding
9. Beach A, Gartrell M, Xing X, Han R, Lv Q, Mishra S, Seada K (2010) Fusing mobile, sensor, and social data to fully enable context-aware computing. In: Proceedings of the eleventh workshop on mobile computing systems & applications, pp 60–65
10. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2:1–8
11. Chae H, Kim DH, Jeong D, In H (2006) A situation aware personalization in ubiquitous mobile computing environments. In: Emerging directions in embedded and ubiquitous computing, pp 213–223
12. Daniel F, Casati F, Soi S, Fox J, Zancarli D, Shan MC (2009) Hosted universal integration on the web: the mashart platform. In: Service-oriented computing, pp 647–648
13. Delft, Hydraulics (2006) Delft3D-FLOW user manual. Delft, the Netherlands
14. Dey AK, Abowd GD, Salber D (2001) A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Hum Comput Interact 16:97–166
15. Feng YH, Teng TH, Tan AH (2009) Modelling situation awareness for context-aware decision support. Expert Syst Appl 36:455–463
16. Fuchs S, Kuhlicke C, Meyer V (2011) Editorial for the special issue: vulnerability to natural hazards—the challenge of integration. Nat Hazards 58:609–619
17. Gao M, Singh VK, Jain R (2012) EventShop: From heterogeneous web streams to personalized situation detection and control. In: Proceedings of the ACM international conference on web science
18. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2008) Detecting influenza epidemics using search engine query data. Nature 457:1012–1014
19. Girolami M, Lenzi S, Furfari F, Chessa S (2008) Sail: A sensor abstraction and integration layer for context awareness. In: Software engineering and advanced applications, 2008. SEAA'08. 34th Euromicro conference, pp 374–381
20. Goodchild MF, Parks BO, Steyaert LT (1993) Environmental modeling with GIS. Oxford University Press, New York
21. Gray J, Liu DT, Nieto-Santisteban M, Szalay A, DeWitt DJ, Heber G (2005) Scientific data management in the coming decade. ACM SIGMOD Rec 34:34–41
22. Harville DA (2008) Matrix algebra from a statistician's perspective. Springer, Berlin
23. Henricksen K, Indulska J, McFadden T, Balasubramaniam S (2005) Middleware for distributed context-aware systems. In: On the move to meaningful internet systems 2005: CoopIS, DOA, and ODBASE, pp 846–863
24. Hore B, Jafarpour H, Jain R, Ji S, Massaguer D, Mehrotra S, Venkatasubramanian N, Westermann U (2007) Design and implementation of a middleware for sentient spaces. Intell Secur Inf IEEE 2007:137–144
25. Hufschmidt G (2011) A comparative analysis of several vulnerability concepts. Nat Hazards 58:621–643
26. Jadhav A, Purohit H, Kapanipathi P, Ananthram P, Ranabahu A, Nguyen V, Mendes PN, Smith AG, Cooney M, Sheth A (2010) Twitris 2.0: Semantically empowered system for understanding perceptions from social data. In: Proceedings of the semantic web challenge 2010
27. Jakobson JB, Lewis L (2006) A framework of cognitive situation modeling and recognition. In: Military communications conference, 2006. MILCOM 2006. IEEE, 2006, pp 1–7
28. Jia L, Lin L, Wei C (2009) "The research and application on streaming data of GIS data mining," in Database Technology and Applications. First Int Workshop 2009:209–212
29. Jones CB, Purves R, Ruas A, Sanderson M, Sester M, Van Kreveld M, Weibel R (2002) Spatial information retrieval and geographical ontologies an overview of the SPIRIT project. In: Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval, 2002, pp 387–388
30. Kallenberg O (2002) Foundations of modern probability. Springer, Berlin
31. Kulldorff M (2001) Prospective time periodic geographical disease surveillance using a scan statistic. J R Stat Soc Ser A (Stat Soc) 164:61–72
32. Lillesand TM, Kiefer RW, Chipman JW (2004) Remote sensing and image interpretation. Wiley, Oxford

33. Loke SW (2004) Representing and reasoning with situations for context-aware pervasive computing: a logic programming perspective. Knowl Eng Rev 19:213–234
34. Mitchell T (2005) Web mapping illustrated: using open source GIS toolkits. O'Reilly Media, Newton
35. Nazari Shirehjini A (2006) Situation modelling: a domain analysis and user study. In: Intelligent environments, 2006. IE 06. 2nd IET international conference on, 2006, pp 193–199
36. Oloufa AA, Ikeda M, Oda H (2003) Situational awareness of construction equipment using GPS, wireless and web technologies. Autom Constr 12:737–748
37. Popovich VV, Pankin AV, Voronin MN, Sokolova LA (2006) Intelligent situation awareness on a GIS basis. In: Military communications conference, 2006. IEEE, pp 1–7
38. Rao B, Minakakis L (2003) Evolution of mobile location-based services. Commun ACM 46:61–65
39. Ratkiewicz J, Michael C, Mark M, Bruno G, Snehal P, Alessandro F, Filippo M (2011) Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th international conference companion on World Wide Web, ACM, pp 249–252
40. Roman R, Lopez J, Gritzalis S (2008) Situation awareness mechanisms for wireless sensor networks. Commun Mag IEEE 46:102–107
41. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: Real-time event detection by social sensors. In: Proceedings of the 19th international conference on World Wide Web, 2010, pp 851–860
42. Khoukhi L, Cherkaoui S, Gaïti D (2009) Managing rescue and relief operations using wireless mobile ad hoc technology, the best way?. IEEE LCN On-MOVE, Switzerland
43. Xu J, Wang S, Su S, Kumar SAP, Wu C (2016) Latent interest and topic mining on user-item bipartite networks. In: 2016 IEEE international conference on services computing (SCC), pp 778–781. IEEE
44. Xu B, Kumar SA (2015) Big data analytics framework for system health monitoring. In: Big data (BigData Congress), 2015 IEEE international congress on, pp 401–408. IEEE
45. Xu B, Kumar S (2015) A text mining classification framework and its experiments using aviation datasets
46. Alampalayam SP, Kumar A (2004) Predictive security model using data mining. In: Global telecommunications conference, 2004. GLOBECOM'04. IEEE, vol 4, pp 2208–2212. IEEE
47. Alampalayam SP, Kumar A (2003) Security model for routing attacks in mobile ad hoc networks. In: Vehicular technology conference, 2003. VTC 2003-Fall. 2003 IEEE 58th, vol 3, pp 2122–2126. IEEE
48. Kumar S, Kumar A, Srinivasan S (2007) Statistical based intrusion detection framework using six sigma technique. IJCSNS 7(10):333
49. Alampalayam S, Natsheh EF (2008) Multivariate fuzzy analysis for Mobile ad hoc Network threat Detection. IJBDCN 4(3):1–30
50. Duran J, Kumar SAP (2011) CUDA based multi-objective parallel genetic algorithms: adapting evolutionary algorithms for document searches. In: Proceedings of the international conference on information and knowledge engineering (IKE). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), p 1
51. Ye J, Dobson S, McKeever S (2012) Situation identification techniques in pervasive computing: a review. Pervas Mob Comput 8(1):36–66
52. Coronato A, De Pietro G (2013) Situation awareness in applications of ambient assisted living for cognitive impaired people. Mob Netw Appl 18(3):444–453
53. McCarthy J (1963) Situations, actions, and causal laws. No. AI-MEMO-2. Stanford Univ CA Dept of Computer Science
54. Coronato A, De Florio V, Bakhouya M, Serugendo G (2012) Formal modeling of socio-technical collective adaptive systems. In: Self-adaptive and self-organizing systems workshops (SASOW), 2012 IEEE sixth international conference on, pp 187–192. IEEE
55. Endsley MR (2016) Designing for situation awareness: an approach to user-centered design. CRC, Oxford
56. Sun R, Zhang X, Mathews R (2006) Modeling meta-cognition in a cognitive architecture. Cognit Syst Res 7(4):327–338
57. Subramanian K, Savitha R, Suresh S (2014) A complex-valued neuro-fuzzy inference system and its learning mechanism. Neurocomputing 123(10):110–120
58. Asaithambi K, Nagoor Gani A (2016) Performance of self-organized and metacognitive neuro fuzzy system for traffic flow prediction. Int J Pure Appl Math 107(4):1025–1036
59. Sateesh Babu G, Suresh S (2013) Meta-cognitive RBF network and its projection based learning algorithm for classification problems. Appl Soft Comput 13(1):654–666
60. Yeh S, Lo J (2005) Assessing metacognitive knowledge in web-based CALL: a neural network approach. Comput Educ 44:97–113
61. Anderson ML et al (2006) The metacognitive loop I: enhancing reinforcement learning with metacognitive monitoring and control for improved perturbation tolerance. J Exp Theor Artif Intell 18(3):387–411
62. Subramanian K, Suresh K (2012) Human action recognition using meta-cognitive neuro-fuzzy inference system. Int J Neural Syst 22(06):1250028
63. Savitha R, Suresh S, Sundararajan N (2012) Meta-cognitive learning in a fully complex-valued radial basis function neural network. Neural Comput 24(5):1297–1328
64. Paisner M, Perlis D, Cox MT (2013) Symbolic anomaly detection and assessment using growing neural gas. In: Proc. of 25th IEEE international conference on tools with artificial intelligence (ICTAI)
65. Löffler A, Klahold J, Rückert U (1999) Artificial neural networks for autonomous robot control: reflective navigation and adaptive sensor calibration. In: Proc. of the 6th intl conf on neural information processing, Perth, Australia, pp 667–672
66. Sateesh Babu G, Xiao-Li L, Suresh S (2016) Meta-cognitive regression neural network for function approximation: application to remaining useful life estimation. In: Proc. of 2016 international joint conference on neural networks (IJCNN)
67. Subramanian K, Das A, Suresh S, Savitha R (2014) A meta-cognitive interval type-2 fuzzy inference system and its projection based learning algorithm. Evolv Syst 5:219
68. Masci J et al (2011) Stacked convolutional auto-encoders for hierarchical feature extraction. In: Proc. of international conference on artificial neural networks, pp 52–59
69. Sakurada M, Takehisa Y (2014) Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proc. of the 2nd workshop on machine learning for sensory data analysis, MLSDA, p 4
70. Sabokrou M et al (2015) Real-time anomaly detection and localization in crowded scenes. In: Proc. of IEEE conference on computer vision and pattern recognition workshops, pp 56–62
71. Feng Q et al (2016) Anomaly detection of spectrum in wireless communication via deep autoencoder. In: Proc. of international conference on computer science and its applications, pp 259–265
72. Shchepetkin AF, McWilliams JC (2005) The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Model 9:347–404
73. Skamarock WC et al (2008) A description of the advanced research WRF Version 3. NCAR Technical Note NCAR/TN-475 + STR, p 113

74. Jongman B, Wagemaker J, Romero BR, De Perez EC (2015) Early flood detection for rapid humanitarian response: harnessing near real-time satellite and twitter signals. ISPRS Int J Geo-Inf 4:2246–2266
75. Bao S, Gayes P, Pietrafesa L (2018) The need and rationale for a coastal flood risk index. In: OCEANS 2018 MTS/IEEE Charleston, IEEE, pp 1–4
76. Landsea C, Franklin J, Beven J (2015) The revised Atlantic hurricane dataset (HURDAT2). NOAA National Hurricane Center, USA