Accepted Manuscript

Big Data analytics and Computational Intelligence for Cyber Physical Systems: Recent trends and state of the art applications

Rahat Iqbal, Faiyaz Doctor, Brian More, Shahid Mahmud, Usman Yousuf

PII: S0167-739X(17)32328-2
DOI: https://doi.org/10.1016/j.future.2017.10.021
Reference: FUTURE 3760

To appear in: *Future Generation Computer Systems*

Received date : 22 November 2016
Revised date : 7 October 2017
Accepted date : 12 October 2017

Please cite this article as: R. Iqbal, F. Doctor, B. More, S. Mahmud, U. Yousuf, Big Data analytics and Computational Intelligence for Cyber Physical Systems: Recent trends and state of the art applications, *Future Generation Computer Systems* (2017), https://doi.org/10.1016/j.future.2017.10.021

Big Data Analytics and Computational Intelligence for Cyber Physical Systems: Recent Trends and State of the Art Applications

Rahat Iqbal[1,2], Faiyaz Doctor[1,3], Brian More[1,4], Shahid Mahmud[1], Usman Yousuf[1]

[1]Interactive Coventry Limited,
Coventry University Technology Park,
Puma Way, Coventry, CV1 2TT,
United Kingdom,
Email: r.iqbal@interactivecoventry.com; dr.faiyaz@interactivecoventry.com; sm@interactivecoventry.com; usman.yousuf@interactivecoventry.com;

[2]Faculty of Engineering, Environment & Computing
School of Computing, Electronics and Mathematics
Coventry University
Priory Street
Coventry, CV1 5FB,
United Kingdom,
Email:r.iqbal@coventry.ac.uk

[3]School of Computer Science and Electronic Engineering,
University of Essex,
Wivenhoe Park,
Colchester CO4 3SQ,
United Kingdom,
Email: fdocto@essex.ac.uk

[4]Coventry University Enterprise Ltd
Coventry University Technology Park,
Puma Way, Coventry, CV1 2TT,
United Kingdom,
Email: BMore@cad.coventry.ac.uk

Big data is fueling the digital revolution in an increasingly knowledge driven and connected society by offering big data analytics and computational intelligence based solutions to reduce the complexity and cognitive burden on accessing and processing large volumes of data. In this paper, we discuss the importance of big data analytics and computational intelligence techniques applied to data produced from the myriad of pervasively connected machines and personalized devices offering embedded and distributed information processing capabilities. We provide a comprehensive survey of computational intelligence techniques appropriate for the effective processing and analysis of big data. We discuss a number of exemplar application areas that generate big data and can hence benefit from its effective processing. State of the art research and novel applications in health-care, intelligent transportation and social network sentiment analysis, are presented and discussed in the context of Big data, Cyber Physical Systems (CPS), and Computational Intelligence (CI). We present a data modelling methodology, which introduces a novel biologically inspired universal generative modelling approach called Hierarchical Spatial-Temporal State Machine (HSTSM). The HSTSM modelling approach incorporates a number of soft computing techniques such as: deep belief networks, auto-encoders, agglomerative hierarchical clustering and temporal sequence processing, in order to address the computational challenges arising from analyzing and processing large volumes of diverse data to provide an effective big data analytics tool for diverse application areas. A conceptual cyber physical architecture, which can accommodate and benefit from the proposed methodology, is further presented.

Keywords: Big Data, Big Data Analytics, Cyber Physical Systems, Computational Intelligence, CI and CPS applications, HSTSM

1. Introduction

The importance of big data in the information economy and to the modern way of life is widely acknowledged. This ever-growing impact can be summarized in the statement that "data is the new oil" or as IBM's Chief Executive Officer recently added: "Big Data is the new oil" (Hirsch, 2013). Just as oil has been instrumental in fueling the industrial revolution in the 20th century so big data is now fueling an ever evolving 21st century digital revolution.

Big data can be defined by the five Vs (see Figure 1): Volume, Velocity, Variety, Veracity and Value (Chang, 2015a).

- Volume refers to the vast amounts of data which are created and stored every second. The data can be in Zettabytes or Brontobytes. For example, big data created by social media, industrial production lines for manufacturing vehicle instrument clusters and so forth.
- Velocity refers to the speed at which big data are created, streamed and aggregated as well as the speed at which big data move around. For example, big data moves from and to social media in seconds and the speed of production lines for processing optic character recognition (OCR) or bank transactions in milliseconds.
- Variety refers to the various types of data collected. The created data can be structured, semi-structured or unstructured which is difficult to process using traditional approaches. It is not possible to categorize the data into regular relational databases for example, big data generated by social media (e.g., photos, text messages) and industrial production lines (e.g., sensory data).
- Veracity refers to the messiness or trustworthiness of the created data. Due to the variety and volume of data it can be messy and contain a lot of noise.
- Value refers to providing meaningful insight into big data. For example analysis of big data for automotive industries to discover patterns in data leading to faults.

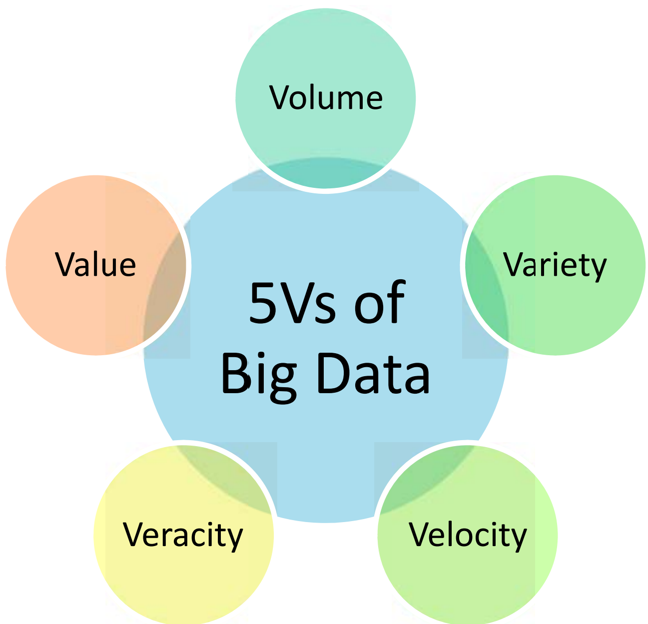


Figure 1: 5Vs of big data

Big Data analytics refers to the techniques used to examine and process Big Data so that hidden underlying patterns are revealed, relationships are identified, and other insights concerning the application context under investigation are exposed.

The original Big Data revolution started with fundamental physics experiments at CERN, and it has now evolved to developing complex data driven models for climate prediction, weather forecasting and seismology. Gone are the days when the computation time to model the atomic nucleus data took 24 hours on a cray supercomputer, it now takes minutes or even seconds on a laptop. Big Data is now transmitted via hundreds of operational satellites where global positional resolution is expected to be 40 cm in a few years' time.

Recent advances in hardware and software technologies have enabled big data acquisition. This data can be harvested from a large number of diverse sources including emails and online transactions, multimedia information such as audio, video and pictures, large databases containing health records and other information. In addition, information can be captured during a user's interaction with social media such as posts, status updates etc., data derived from search queries or click patterns of a user, physiological data such as heart rate, skin conductivity etc. as captured from wearable sensors, data derived and extracted from our interaction with our mobile devices, embedded artifacts inside a smart home, data from production machines and industrial robots, scientific research and other sources (Eaton, 2012). It is clear from the above that in the modern world data is being generated at an accelerated rate (Villars, 2011).

The potential utilization of this huge amount of information has catapulted Big Data and Big data analysis as a central focus for modern research communities, modern businesses, and governments (Hashem, 2015) working towards delivering the promise of a plethora of new application areas and opportunities which can emerge under completely diverse application contexts from smart cities (Hashem, 2016) to digital health care (Murdoch, 2013). As a result the benefits arising from this wealth of knowledge and information can affect research in numerous ways. These can include : promoting medical advances by providing evidence in the identification of symptoms and patterns concerning diseases, pandemics and modern health issues; or aiding in the creation of large ground truth databases for scientific fields such as sentiment analysis, which are in desperate need of vast amounts of data in order to successfully create models of human affect and effective behavior recognition techniques. The businesses that constitute the modern economy can also greatly benefit from Big Data and Big Data analytics since they can utilize the data generated from the interaction of users with social network (Yaqoob, 2016) or smart devices in order to identify the users' preferences towards a product, recognize dissatisfaction of modern clients, or understand the relationships between competitive and collaborating organizations, thus creating better and more appealing products and services , or improving existing ones.

Nowadays huge quantities of personalized and contextualized information are generated in platforms such as social networks extending to wearable devices, where millions of people interact and express their opinions and emotions. The development of advanced Big Data analytics and computational intelligence techniques enables the development of intelligent computerized solutions with the help of social and behavioral data based sentiment analysis.

Sentiment analysis aims to improve products and services, by automatically identifying the user's opinion, including their evaluations and affective state (Karyotis, 2017). Acquiring more nuanced insights of customer preferences and needs will provide modern organizations and businesses with a crucial advantage over their competition (Sagiroglu, 2013). For instance, big data received from electronic communications can be used to enable employees to become more emotionally resilient in the work place. As mentioned in (Hirsch, 2013) Big Data "is becoming a significant corporate asset, a vital economic input, and the foundation of new business models" (Hirsch, 2013). Businesses and organizations can benefit from Big Data through the deployment of technologies such as Cloud Computing services which would allow for meeting the storage and processing requirements of Big Data analysis (Chang, 2015b). In Chang and Will's work, a balanced approach to comparing non Cloud to Cloud storage was presented and realized by utilizing the appropriate experimental set up and metrics. From the team's experimental results a significant performance improvement was observed in relation to execution times, consistency between expected and actual execution times, and efficiency, when Cloud approaches were used compared to the case when non-Cloud systems were utilized (Chang and Wills, 2016b). Nowadays, Cyber-physical cloud systems (CPS) have emerged as state of the art cloud-based architectures, which are utilized in a wide range of applications. CPS can be defined as hierarchical architectures, where devices located in the physical layer need to communicate and transact securely with computing and communication resources in the cyber layer. These complex interactions include several Big Data related operations, such as sensing, storing, and processing large amounts of heterogeneous data. Therefore, it is extremely challenging to effectively handle these operations in terms of crucial aspects, such as security and energy management. Security is of outmost importance since in CPS environments a variety of sensitive transactions are performed (e.g. selling and buying of energy). Efficient energy management is also a widely acknowledged research challenge for CPS, which has a significant impact on provided services, but also the environment. CPS research can largely benefit from intelligent computational intelligence and Big Data analytics techniques in order to tackle these modern challenges.

Big Data analytics can also facilitate government's efforts towards delivering better services to their citizens. Big Data can aid governments in improving crucial sectors such as healthcare and public transport thus helping to shape a more efficient modern society. For example, Big Data analytics and computational intelligence techniques are able to provide intelligent solutions for challenging problems such as health shock prediction, or optimization of the public transport services delivered by the state to the population.

In order to capitalize on the advantages of Big Data analytics in an increasingly knowledge driven society there is a need to develop solutions that reduce the complexity and cognitive burden on accessing and processing these large volumes of data in both embedded hardware and software-based data analytics (Maniak, 2015) (Iqbal, 2015). Big challenges stem from the utilization of Big Data in the real world, since the implementation of real time applications is becoming increasingly complex. This complexity derives from a variety of data-related factors. One factor is the high dimensionality degree which a dataset may possess thus increasing the difficulty of processing and analyzing the data. The interactions, co-relations and causal effects of these high dimensional data parameters in relation to the behaviours and specific outcomes of these systems are often too complex to be analysed and understood by human users. Additionally, data can be accumulated from diverse sources and input channels, making online processing very demanding due to the variety of signal inputs which need to be synchronized, and diverse data types which need to be analyzed simultaneously. Furthermore, the collected data is often comprised of multiple types of inputs, which are also not always precise or complete due to various sources of imprecision, uncertainty, or missing data (e.g. malfunctioning or inaccurate sensors). Moreover, there is an inherent need in real life applications for high-speed storage, processing of data and retrieval of the corresponding analysis results. Another factor that should be taken into account is that the method utilized for Big Data analytics should extract knowledge from data in an interpretable way. The computational techniques deployed to perform this task should make the underlying patterns, which exist in the data, transparent to the person who wishes to utilize and understand them. Finally, there is a need for techniques performing online adaptation to incorporate contextual and user-specific elements in their design, and decision-making mechanism, in a user friendly and computationally feasible manner. All the above factors should be reflected in the computational and machine learning techniques utilized in order to process and analyze Big Data so that successful applications and models can be constructed (Suthaharan, 2014).

The rest of the paper is organized as follows. Section 2 discusses computation intelligence for big data analytics. Section 3 presents our novel methodology to provide solutions to data driven problems. Section 4 presents few examples of application areas in which the data driven methodology is applied. Section 5 concludes the paper.

2. Computational Intelligence for Big Data analytics

Machine learning (ML) approaches are used for modelling patterns and correlations in data in order to discover relationships and make predictions based on unseen data / events. ML approaches consist of supervised learning (learning from labelled data), unsupervised learning (discovering hidden patterns in data or extracting features) and reinforcement learning (goal oriented learning in dynamic situations) (Mitchell, 1997). As such, ML approaches can also be categorised into: regression techniques, clustering approaches, density estimation methods and dimensionality reduction approaches. Non-exhaustive examples of these approaches are: Decision tree learning, Associate rule learning, Artificial neural networks, deep learning support vector machines, clustering and Bayesian networks.

Computational Intelligence (CI) is a subclass of ML approaches where algorithms have been devised to imitate human information processing and reasoning mechanisms for processing complex and uncertain data sources. CI techniques form a set of nature-inspired computational methodologies and techniques which have been developed to address complex real-world data-driven problems for which mathematical and traditional modelling are unable to work due to: high complexity, uncertainty and stochastic nature of processes. Fuzzy Logic (FL), Evolutionary Algorithms (EA) and Artificial Neural Networks (ANN) form the trio of core CI approaches that have been developed to handle this growing class of real-world problems.

FL is an established methodology to deal with imprecise and uncertain data (Zadeh, 1965). FL provides an approach for approximate reasoning and modelling of qualitative data and adaptive control (Doctor, 2016) (Liu, 2014) based on the use of linguistic quantifiers (fuzzy sets) for representing uncertain real-word, data and user-defined concepts and human interpretable fuzzy rules that can be used for inference and decision-making. EAs are based on the process of natural selection for modelling stochastic systems (Whitley, 2001) and approaches such as genetic algorithms, genetic programming and swarm intelligence optimisation algorithms (Dreier, 2002) (Poi, 2008) (Parpinelli, 2011) can be used for optimising complex real-world systems and processes. Finally ANNs enable feature extraction and learning from experiential data (Haykin, 2009) and are based on imitating the parallel processing and data representation structure of neurons in animal and humans brains. A NN is an interconnected assembly of basic elements (artificial neurons) that broadly speaking resemble the neurons existing in our brains. The analyzing ability of neural networks is hidden in the values of the weights that connect these basic elements. These weights are acquired by adaptation, or by learning from training data (Gurney 1997).

A combination of CI techniques can be used to extract insight and meaning from the data, offering integrated solutions, which can be applied to a variety of application domains. Such solutions should be adapted to offline and online, hardware and software data processing and control requirements, which can be further optimised to domain dependent constraints and dynamics. Hence these approaches can be used to provide effective multipurpose intelligent data analysis and decision support systems for a variety of industrial and commercial applications characterized by large amounts of vague or complex information requiring analysis to support operational and cost effective decision-making (Doctor, 2013).

2.1 Deep learning for Big Data analytics

In Big Data analytics there is a growing need to accurately identify important features in the data affecting the outputs, and to determine the spatial co-relations between input variables at a given point in time as well as the causal or temporal co-relations between input parameters that change overtime. Effective modelling to identify patterns from these data sources can be employed to produce accurate predictions of how a system is supposed to behave under normal operational conditions, and enable the detection of abnormalities. Deep learning algorithms have attracted increasing attention by providing effective biologically inspired computational modelling techniques for addressing tasks such as speech perception and object recognition by extracting multiple levels of representation from various sensory inputs and signals (Bekkerman, 2011) (Bengio, 2009) (Hinton,2012) (Le, 2013) (Campo, 2014). These approaches can offer the means to model large-scale data with significant dimensionality as well as spatial and temporal correlations for sequence modelling tasks. Deep Learning (DL) approaches are based on the principle of using ANNs with multiple hidden layers as shown in Figure 1. This allows both unsupervised (bottom-up) training to generate higher level representations of sensory data which can then be used for training a classifier (top down) based on standard supervised training algorithms (Hinton, 2006). Feature learning methods are based on supervised approaches such as Deep NNs, Convolutional Neural Networks (CNNs), and Recurrent NNs along with unsupervised techniques such as Deep Belief Networks and CNNs and provide deep architecture that combine structural elements of local receptive fields, shared weights, and pooling that aims to imitate the processing of simple and complex cortical cells found in animal vision systems (Korekado, 2003).

The potential of utilizing deep learning techniques in Big Data analytics has been highlighted by recent review studies (Tolk, 2015) (Chen, 2015). In the work by Tolk et al. deep learning potential as a modelling approach and as a means to discover correlations from data is highlighted. Based on a thorough review of recent applications, the researchers argue that Big Data and deep learning have the potential to provide a new generation of modelling and simulation applications (Tolk, 2015). The ability of DL methods to handle cases where the amount of data is huge is also discussed in the work by Chen et al. (Chen, 2015). This work has demonstrated the key role of deep learning approaches for solving Big Data analytics problems.

There are a number of recent examples of research in applying optimized and enhanced deep learning techniques in order to analyze and process Big Data. More specifically in the work by Zaidi et al., the researchers present an algorithm for Deep Broad Learning which can be tuned in order to specify the depth of the mode, and this has achieved notable accuracy of performance for a large amount of data which can be considered competitive with other state of the art research (Zaidi, 2015). In (Alsheikh, 2016) the research addresses a very modern challenge arising from the large amount of data, which can be collected through mobile devices. The team explores deep learning as a technique for mobile Big Data analytics and presents a scalable learning framework for Apache Spark. From the experimental results, it can be seen that the team's framework achieves a significant increase in the speed of learning process for deep learning models, which are comprised of a large number of hidden layers and parameters. In the work by Lv et al., the researchers applied a deep learning technique that takes into account spatial and temporal correlations in order to utilize Big Data towards traffic flow prediction, with the purpose of achieving a high performance (Lv, 2015). In the paper by Chung et al. it is stated that deep neural networks demonstrate very high performance in pattern recognition tasks however they are computationally expensive as they require training time, which in some cases can reach an increase of a factor of 10 compared to other approaches (Chung, 2014). More specifically, research has investigated deep neural network training by utilizing the data parallel Hessian-free 2nd order optimization algorithm. The experimental results, which were calculated on speech tasks of a large scale, have demonstrated a significant performance increase without decreasing accuracy, thus allowing deep neural network training by utilizing Big Data in a suitable amount of time.

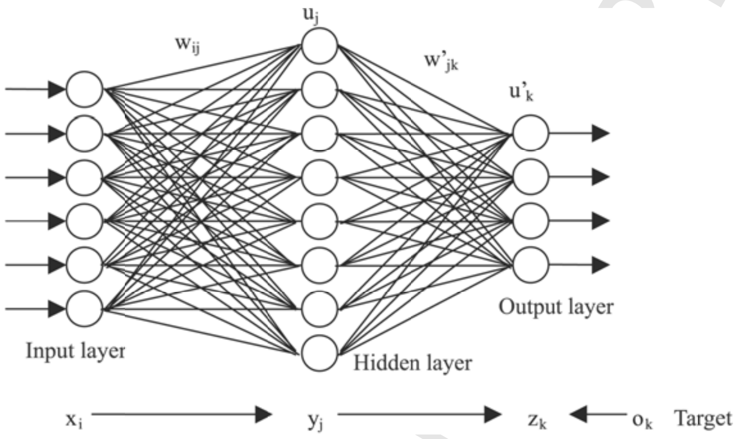


Figure 2. DNN structure.

2.2 Fuzzy Logic for Big Data analytics

Big Data analytics creates challenges concerning the nature of the data accumulated from diverse and potentially noise infected sources. These data sets are subject to high degrees of uncertainty and contain a large amount of noise and outlier artifacts. Previous research has demonstrated that fuzzy logic systems are CI applications, which can efficiently handle inherent uncertainties related to the data. For example creating models to predict the user's emotions usually relies on dealing with databases that contain a large amount of uncertainty, which is associated with the fuzzy nature of human emotion (Wu, 2012).

Fuzzy systems have proven their ability to deal with this challenge, resulting in robust systems which have better or similar performance levels as other more complex techniques, while at the same time retaining a highly satisfactory tradeoff between classification accuracy and time performance (Wu,2010)(Karyotis, 2015). This is another crucial factor when dealing with huge amounts of data since it allows the delivery of classification results in a reasonable amount of time. Since Fuzzy Logic relies on the natural language fuzzy rules it allows for a successful visualization of hidden relations existing in data, thus enabling the users of applications or researchers searching for hidden patterns in data to easily visualize these underlying relations (Doctor, 2012). Developing applications with a high degree of visualization and interpretability contributes to improving the usability aspects of developed systems, which as recognized by research is a very important issue for successful commercial applications (Iqbal, 2011) (Iqbal, 2013). Finally, Fuzzy Logic systems and more specifically adaptive Fuzzy Logic systems have demonstrated considerable potential concerning their ability to model and account for individual differences and contextual information with a very reasonable computational burden, thus making them an excellent choice for creating personalised and user-friendly systems from huge amounts of data (Karyotis, 2016) (Karyotis, 2015) (Doctor, 2005).

Fuzzy Logic has been utilized as a basis for performing Big Data analytics in different application areas. Behadada et al. have presented a methodology for semi-automatically defining fuzzy partition rules. They have utilized large publicly available data sets, data from experts, and experimental data in order to construct a system for heart rate arrhythmia detection. Their results demonstrate an applicable tradeoff between accuracy and interpretability (Behadada, 2015). Fuzzy logic has also been used by a number of applications which utilize Big Data extracted from social networks for the analysis of public opinion (Glosh, 2016) (Bing, 2014). For example in the work by Bing et al., a matrix-based fuzzy system (FMM system) was developed in order to mine Twitter data. The experimental results showed that this system had a very good predicting performance whilst at the same time had low processing times (Bing, 2014). In another interesting application in the field of medicine, Duggal et al. utilized fuzzy logic based matching algorithms and MapReduce in order to perform Big Data analytics for clinical decision support. The developed system demonstrated a high level of flexibility and was able to handle data from various sources (Duggal, 2015). An example Fuzzy Logic system can be seen in figure 3, which consists of four main components: the fuzzifier, the inference engine, the rules, and the output processor. The fuzzifier is responsible for mapping a crisp value such as a sensory measurement into a fuzzy number. The inference engine is responsible for the way the fuzzy rules are fired and combined to map input fuzzy sets to output fuzzy sets. Rules are if-then statements, which are extracted by domain experts, or experimental data. Finally, the output processor is responsible for the defuzzification of the output fuzzy sets that refers to the generation of a crisp value for the produced fuzzy output.

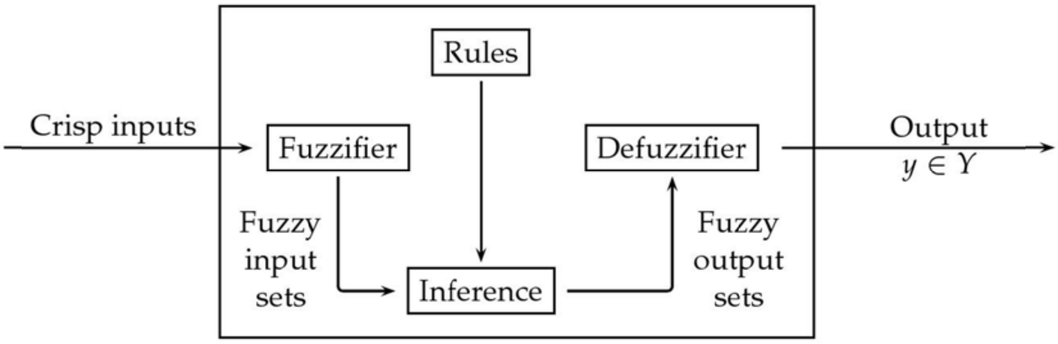


Figure 3. Fuzzy Logic System (Mendel 2001).

2.3 Evolutionary algorithms for Big Data analytics

Evolutionary algorithms (EA) are another CI technique, which can meet the requirements and challenges of Big Data analytics. Evolutionary algorithms mimic the evolution process to the discovery of the optimal solution of a specific problem by evolving a population of candidate solutions (Sun, 2005). EA are very good explorers of the search space which makes them excellent tools for Big Data analysis, since Big Data are subject to a very high degree of dimensionality and sparseness (Bhattacharya, 2016). This problem was investigated in the paper by Bhattacharya et al. where the researchers developed an EA with the ability to deal with both issues. Although the application was not Big-Data ready, as was acknowledged in the paper, the initial evaluation to identify benchmark problems showed that the approach performed very well compared to other modern techniques. Evolutionary algorithms are also proven techniques for a number of machine learning related problems which are utilized in Big Data Analysis such as clustering (Razavi, 2015), feature selection (Lin, 2016), and others. EA have also been used by recent research in conjunction with signal inputs such as the EEG signal, towards modern and demanding application areas such as multi-brain computing (Kattan, 2015). GA's can also be used in combination with other techniques in order to optimize several parameters of predictive models and facilitate the development of more accurate systems. For example, in the work by Karyotis et al., the researchers utilized a genetic algorithm to optimize the properties of the fuzzy sets used to develop a model for representing the user's emotions. The GA was used offline to tune the system and it was able to improve the accuracy of the proposed model (Karyotis 2017). GAs are widely used evolutionary algorithms based on the Darwinian theory of evolution, and a demonstrative example of their workflow can be observed in figure 4. A solution for a specific problem is represented by a chromosome comprising of genes that correspond to different values of parameters of the problem. In the beginning, an initial population is defined and for every chromosome, an objective function is calculated (such as values of an error function). This value would represent the fitness of the individuals. The chromosomes with the best fitness values are combined in order to generate new offspring. This process is repeated until a solution that satisfies a stop criterion to a given problem is achieved.

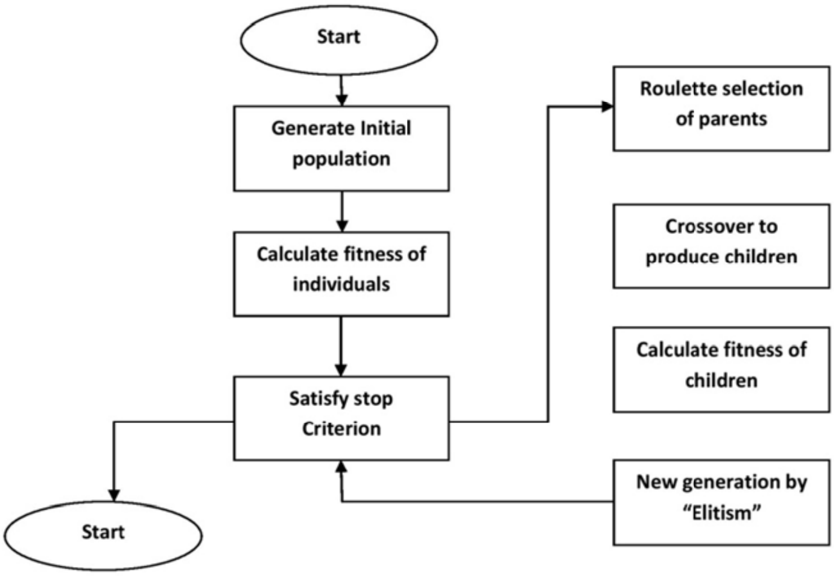


Figure 4. Genetic algorithm flowchart (Kang 2011).

3. Proposed data modelling methodology

In this section, we discuss a novel data modelling methodology that can be applied in diverse big data settings. This approach has already been used successfully towards the development of the taxi demand prediction application described in section **4**, and is able to provide solutions to other challenging real world big data problems. Based on the latest discoveries in the field of neuroscience the core part of our proposed data modelling methodology introduces a novel biologically inspired universal generative modelling approach called Hierarchical Spatial-Temporal State Machine (HSTSM) that has been developed on the understanding of the structure and functionality of the human brain. The proposed approach is based on a hybrid method incorporating several soft computing techniques such as: deep belief networks, auto-encoders, agglomerative hierarchical clustering and temporal sequence processing (Maniak, 2016). Our approach is able to handle high volumes of data characterised by complex correlations between input values and temporal consequences between different input states of the system (phrased in this work as spatial-temporal correlations). The approach is modelled to mimic the structure and functions of the mammalian brain based on a theory proposed by Jeff Hawkins (Hawkins, 2004) (Hawkins, 2009).

The main elements of our proposed methodology include data layer, input layer, data transformation layer, spatial pooling, temporal inferences, prediction model, optimization and finally the presentation of information or application layer as shown in Figure 5.

3.1 Data Layer

The data layer is responsible for capturing heterogeneous and homogeneous data from an existing infrastructure consisting of hardware and Internet of Things (IoT) data sources (digital and analogue device states and sensors signals); manual input data sources (operator actuations, checklists and survey data) and software data sources (structured and unstructured text, documents, images and audio inputs).

3.2 Input layer

As data is being derived from different sources there is a need for it to be effectively pre-processed by the application of data fusion techniques. Here sensor data fusion techniques are used to combine, align and interpolate over parameters from heterogeneous data streams that would have different sampling rates, missing information and variations in accuracies. It is crucial to acquire high-quality data, which is both precise and contains no or little noise. In order to make meaningful use of the data acquisition process, the data output needs to be encoded in a specific way.

3.3 Data Transformation Layer

Data transformation layer deals with transforming the pre-processed data into machine readable codes by the use of encoders. Input data is initially encoded into what is termed as a Sparse Distributed Representation (SDR) (Hawkins, 2009) which decomposes the representation of information over thousands of bits. Here at any point in time a small percentage of these bits may be activated, i.e. equal to 1's emulating the firing response patterns of neurons in the brain. Each bit represents a feature of information and a set of activated bits is therefore able to encode the semantic attributes of what is being represented. These bits are unlabeled and the associated semantics meanings are learnt. Hence if different SDRs have activate bits in the same location then they can be considered to spatially share the sematic attributes and we can use this to learn the spatial correlations between different input patterns as will be discussed in the following section.

3.4 Spatial Modelling

The discovery of correlations between individual inputs (bits) are determined through the spatial transformation of input space into a transformed feature space that is achieved through the use of deep belief networks, where this process is referred to as spatial pooling (Hawkins 2004). Hierarchical clustering is performed on the transformed features derived from the deep belief network, to extract a number of possible states of the modelled system. The main purpose of this operation is for reducing the input space to a fixed number of the most probable states of the underlying system being modelled.

In the beginning of the modelling process, raw process data is generated by the relevant data sources and is transformed from its textual - human readable form, to binary sparsely distributed format. Data encoding aims to map the inputs defined for a specific modelling problem to discrete signals that can be understood by the HSTSM. The encoded vector is compressed, and an automatic process of feature extraction is performed with the use of deep belief neural networks (DBN). The trained restricted Boltzmann Machines (RBM) forming the DBN are used to initialize the deep auto-encoder. This unsupervised method of feature extraction enables us to acquire an improved and more compressed representation of the input space. Because of the unsupervised nature of the process, there is no need for labeling the data at this stage. Therefore, the algorithm is able to cope successfully with high volumes of data. The features extracted are used to automatically identify a set of possible patterns on the inputs to the model. This would correspond to a set of possible states of the monitored system. In order to achieve this we incorporate into the model agglomerative hierarchical analysis, which is an unsupervised method of cluster analysis. The basic metric used to conduct this analysis is defined to be the Euclidian distance. This process can be considered as spatial pooling, where the original binary input that occur close together in space are pooled together. This operation enables the extraction of a set of individual system states, and the analysis of the temporal sequence of input activities in terms of these states.

3.5 Temporal Modelling

The initial stage of the approach performs hierarchical organization of multiple levels of data abstraction to identify correlations between temporal sequences of input patterns. Temporal sequence learning is used to train the model on different temporal consequential relations between probable states of the system. This is used to infer the next predicted state of the inputs in comparison with the actual behaviour of the system, which is termed as temporal inference.

Temporal inference is performed on the identified states, and prediction of the next possible state of the system can be achieved with the utilization of an n-order Markov chain. Prediction acquired in this way can be consequently used to identify specific patterns of behavior of the modelling problem under investigation. At the end of this process, the predicted vector can be compared with the actual vector, generated at each moment in a specific application context, in order to identify specific patterns or irregular behavior. This can be achieved with the use of distance function or basic ML techniques, like a Multilayer Perceptron.

3.6 Prediction Model

The spatial pooling and temporal inference elements of the approach hence combine to produce a spatial temporal model of the operational behaviour of the system being modelled. The model can then be used in combination with prediction and classification approaches such as standard ANNs to predict future behaviour of the system under different operational conditions and detect deviations and changes in behaviour that might signify an underlying unknown effect or problem.

3.7 Optimisation

The prediction model can further provide inputs to an optimisation framework that is able to optimise processes based on quantitative and qualitative inputs from various sources. Here evolutionary systems based approaches such as genetic algorithms and swarm intelligence optimisation algorithms can be used to simulate what if scenarios through the simultaneous optimal estimation of unknown parameters for determining optimal states of the process, system or scenario being modelled. This approach can therefore be used as a means to determine behaviour changes and deviations of complex systems, which could be the result of environmental effects, human behaviour change or faults in equipment and devices.

3.8 Application Layer

The layer is responsible for displaying the results through the interface of an application or visualization which can provide stakeholder insight into the data being modelled. The visualization or interface can also be personalized in terms of the type and functional form of information representation in order to meet the individual users' needs.

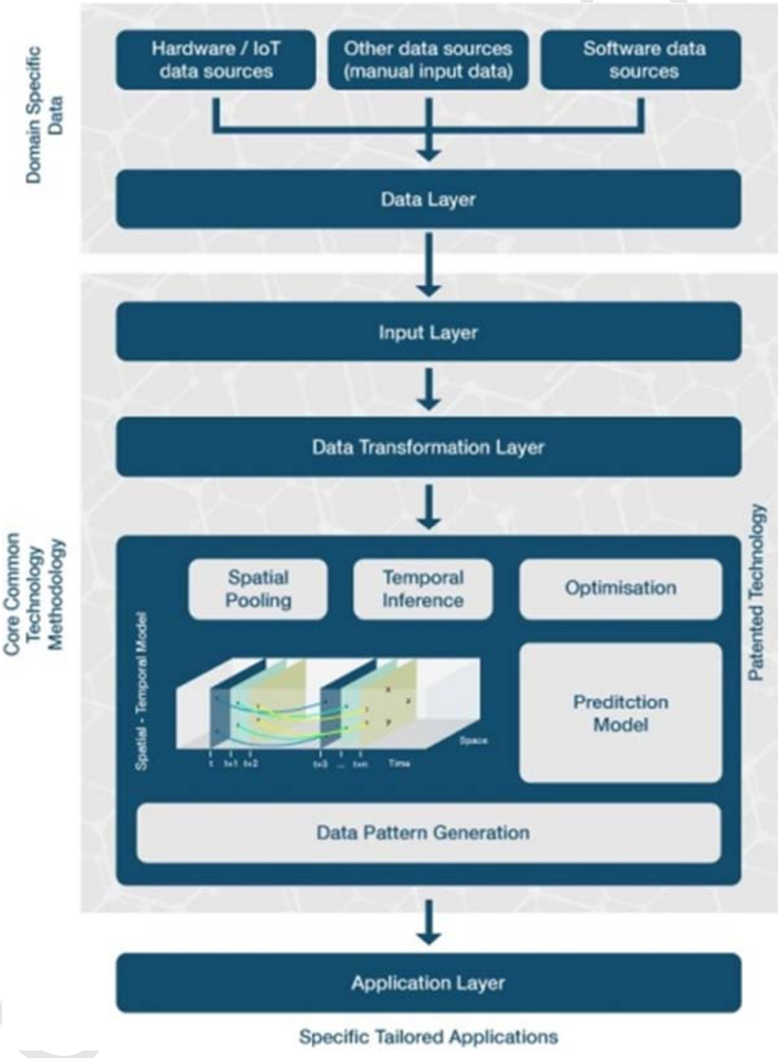


Figure 5. Big Data modelling methodology

A conceptual cyber physical architecture, which can accommodate and benefit from the proposed methodology, can be found in the work by Hossain et al. (Hossain, 2017) (See Figure 6). In this paradigm, data captured from sensors and actuators located in a smart phone, are used to monitor the user. The aforementioned data, along with data related to location information, are sent from physical space to cyber-space. This data can be processed in order to provide intelligent recommendations, and support to the user via the smart phone. In this case, the user's phone is the connecting interface between cyber and physical world, via a range of communication technologies (e.g., Bluetooth, GPS, Ultra Wide Band, WiFi, RFID) (Hossain, 2017). The proposed data modeling methodology can be accommodated in this and similar architectures that interface with edge level embedded systems, and contribute to the efficient manipulation and exploitation of data, as well as facilitate the development of successful machine learning-based research and commercial applications.

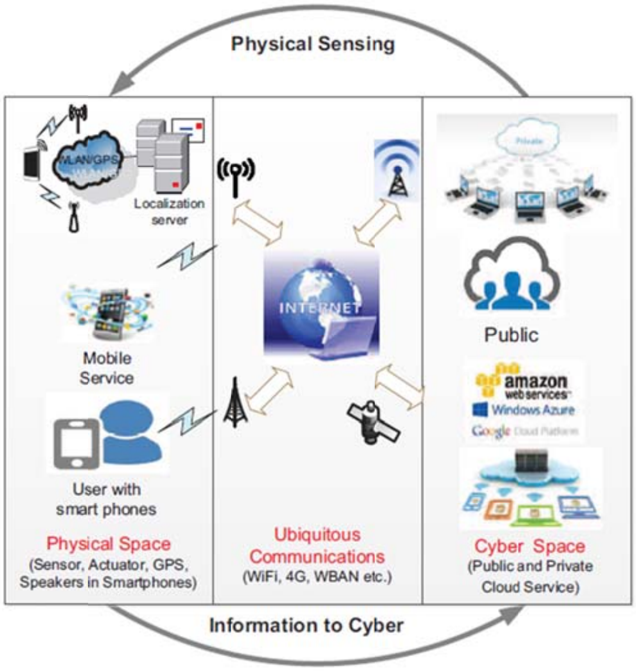


Figure 6. Health shock prediction research methodology (Hossain, 2017).

4. Potential Application areas

Big Data and CI can be utilized in order to provide novel applications with scientific and commercial value. In this section, we provide examples of opportunities for improved health services, optimized transportation, social network sentiment analysis, and present novel applications that tackle modern challenges in these domains.

4.1 Health Services and Health Shock Prediction

In recent years, we have witnessed the introduction of advanced cyber-physical systems (CPS) in various healthcare applications domains. Modern health CPSs are able to replace traditional health-care delivery practices and services. In order to achieve their goals CPSs should be equipped with powerful algorithms able to tackle the analysis, synthesis, processing of data, and the integration of technology challenges of the new age (Ahmed, 2013). CPSs face Big Data problems since large amounts of high velocity data are produced from various devices in different formats, and this data needs to be stored and processed timely and effectively. Nowadays, given the right tools are provided intelligent technologies can benefit from a variety of data sources containing rich information, in order to deliver their services. This data include: research data such as clinical trial, and screening data, clinical gene or protein data from scientific institutions; medical billing and insurance data, which can be used to analyse the cost of medical services, stored in geographically dispersed databases; clinical data collected for diagnosis such as medical image and electronic medical records (EMR); individual activity and emotion data collected by standard wearable devices and other forms(Zhang, 2017). CPS technologies should possess data analytics and modeling tools, based on the methodology described in this paper, in order to deal with the diverse nature and complexity of the aforementioned data sources, demanding time and computational requirements.
An example of recent research in health CPSs is the paper by Hossain et al.. Here researchers have introduced a monitoring system that utilizes a multimodal approach combining voice and EEG signals. The team have presented a cloud supported CPS able to use smart-phones, cloud computing and CPS, in order to support healthcare monitoring (Hossain, 2017). In Zhang et al.'s work a CPS called Health-CPS that was based on cloud and big data analytics technologies, was presented. This system was designed for patient-centric healthcare applications and services and consisted of a data collection layer with a unified standard, a data management layer for distributed storage and parallel computing, and a

data-oriented service layer (Zhang, 2017). In the paper by Wang et al. a cyber physical architecture consisting of a communication core, a computation core, and resource scheduling and management core was presented, and the authors provided examples of its application to healthcare monitoring and decision support systems (Wang, 2011).

Health services can benefit patients and clinicians by providing optimised health information and recommendations, based on individual and population-based profiling, and using advanced Big Data analytics algorithms. Gaining a better insight into an individual's healthcare needs is of central importance in order to provide tailored treatment and formulate therapy intervention recommendations. Moreover, this extends to fitness, lifestyle, social care and wellbeing monitoring, based on personalised preferences and goals that can be used to promote positive health, support behaviour change related to diet, exercise, and reduction of stress, and effective management of changing care needs such as for the elderly. Monitoring and delivering these services though personal (Harding, 2015), cloud based (Sultan,2014), mHealth (Free, 2013) and Internet of Things (IoT) applications (Pang, 2015) will help to empower people to manage their health and life style more effectively, owing to reduced healthcare costs. Computational health informatics can be used on large population based data through the development of interpretable decision support models for promoting effective health policy, and intervention planning for crisis management related to disease epidemics and famine.

Dementia is an age related, progressive, neurodegenerative condition, also considered to be one of the biggest global public health challenges that the current generation needs to face. The growing prevalence of diseases such as Alzheimer's, and their impact on a society benefiting from greater longevity, is a critical health challenge of national importance. Managing patients with dementia calls for improvements in effectively monitoring the progress of their condition, adjusting therapy interventions, and adapting to changing care needs (Doctor, 2014). Advanced deep learning approaches can be used to enable the development of context aware dementia monitoring, and predictive care recommendation solutions that can intelligently forecast behaviour changes of individual patients, from utilising contextual heterogeneous data, while handling uncertainties associated with incorporating qualitative data from stakeholders. Human interpretable care recommendation decisions can provide care staff with the means of implementing evolving care and therapy plans, in response to the changing needs of patients due to the effects of cognitive decline.

Big data analytics and computational intelligence techniques are also able to facilitate governments and organizations provide solutions to challenging problems. Health shock is a health related event with a heavy impact on a household because of the cost of treatment, or the absence from work caused by the health related event. Health shock especially affects individuals in developing and underdeveloped countries. Government, organizations, policy makers, and individuals could largely benefit from the development of intelligent Big Data analytics techniques to battle health shock by facilitating accurate prediction and by highlighting the factors contributing to this phenomenon. These techniques will enable effective and timely mitigation and management of health shock events. In order to address this problem and contribute in this direction a Big Data analytics and visualization framework based on fuzzy logic was presented, which utilized a large-scale dataset towards health shock prediction purposes (Mahmud, 2016). The researchers utilized cloud Amazon web services integrated with Geographical Information Systems (GIS) to facilitate the collection, storage, indexing, and visualization of data for different stakeholders using smart devices, and develop their framework (Mahmud 2016). The dataset utilized comprised of data from questionnaires, and an online system where each data sample comprised of 47 features. Data were gathered from 1000 households belonging to 29 villages in rural areas of Pakistan. The collected data where pre-processed and analyzed by utilizing expert-opinion to extract meaningful measures related to living standards, health risk, accessibility to health facilities and income allocation labeled with a level of health shock incurred. A fuzzy rule summarization technique was utilized in order to develop a health-shock prediction model. The proposed predictive fuzzy model was evaluated by utilizing k-fold cross validation. An overview of the research methodology can be seen in figure 7. The model achieved a very high performance of 89%, while at the same time retained a very high interpretability degree with the help of the extracted natural language fuzzy rules. These rules can enable all relevant stakeholders to understand the causal factors affecting this phenomenon, and make informed and effective decisions. Moreover as demonstrated by the team's experimental results the proposed framework was also able to deliver results in competitive times with a low computational burden thus making it suitable for real-time and Big Data settings. Furthermore, data were mapped to the iron triangle model (Kissick, 1994) under the socio-economic, geographical, and cultural norms, and factors that lead to health-shock.

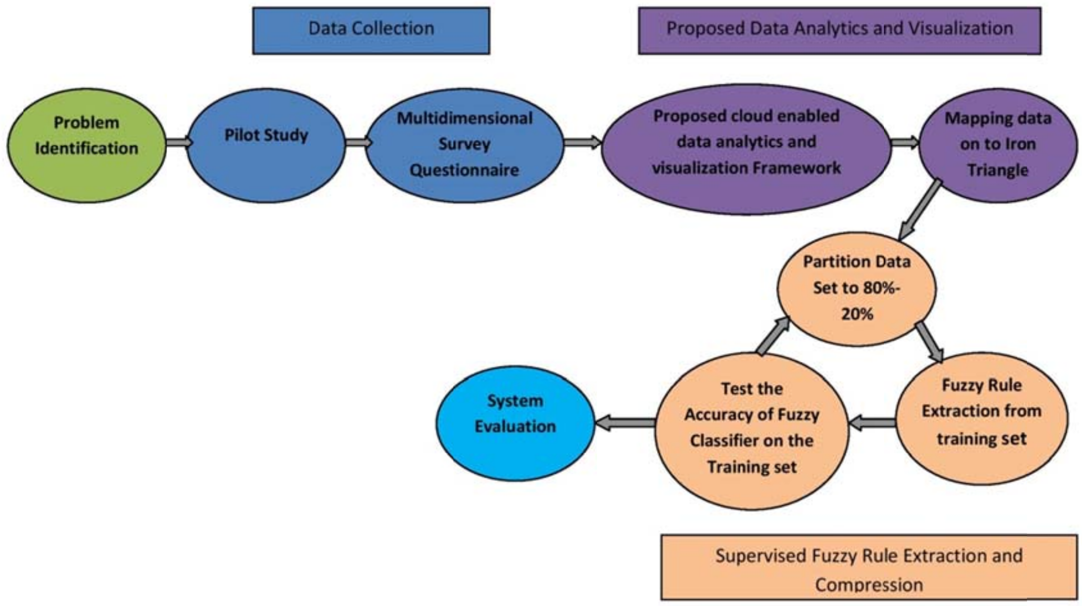


Figure 7. Health shock prediction research methodology (Mahmud 2016).

4.2 Intelligent Transportation

The growth of modern urbanised and increasingly connected rural environments requires the development of efficient transportation infrastructures to better support the needs of visitors and commuters. The boast in wireless communication techniques, computational intelligence and data analytics capabilities, mobile cloud computing, and context-aware technologies, gave rise to novel research and commercial applications in the transportation field. This includes the upgrade of current vehicular network technologies to cloud-assisted context-aware vehicular CPS. A demonstrative example is the work by Wan et al., where a multi-layered context-aware architecture was presented. The authors investigated an application scenario concerning dynamic parking services, and discussed challenges and solutions regarding context aware hazard prediction, dynamic vehicle routing, and vehicular clouds (Wan, 2014). One of the most prominent application areas of CPS is traffic guidance and forecasting. Accurate traffic prediction could provide improved route planning and movement of goods and people, and more informed and efficient decision making for traffic management or development and maintenance of transport infrastructure. Very recent work by Chen et al. provided a forecasting model for short-term traffic conditions in vehicular CPS within urban environments. As it was demonstrated by the team's experimental results, their proposed fuzzy Markov prediction model can be used in order to effectively forecast short-term traffic (Chen, 2017). It is also predicted that CPSs will affect the design of aircrafts, air traffic management systems, and aviation safety, thus impacting upon the development of next generation air transportation systems. As identified in Baheti et al.'s work new research areas consist of: new processes to achieve greater capacity, improved safety, as well as optimize the tradeoff among these goals; integrated flight deck systems, and autonomous systems; enhanced vehicle check, health monitoring, management; and research related to aircraft control systems (Baheti, 2011). Besselink et al. investigated the potential of ICT technology to reinforce CPS systems with an integrated logistic system to organize the movement of trucks travelling together. Their results demonstrated that optimized interaction between the tracks has clear benefits such as improved fuel consumption (Besselink, 2016).

A number of initiatives have been implemented to satisfy personalised and contextualised user defined objectives in relation to improving user mobility, utility, and satisfaction while helping to avoid congestion (Djahel, 2015) (Vegni, 2013). Self-learning cars (Jaguar Land Rover Development, 2016) aim to address this through the smart management of infotainment delivery to pre-emptively manage the delivery of personalised content based on availability of WiFi connections, smart anticipatory download, and cashing of content for on demand services. This focuses a lot on approaches for information retrieval and its delivery based on ad hoc communication networks (Alhabashneh, 2017) (Iqbal, 2015) (Qureshi, 2017). Understanding each driver's behaviour and information needs in terms of their intentions can be used to provide relevant information and services as they are needed which can be used to improve user satisfaction, vehicle efficiency, and energy utilization.

Another problem is effectively managing transport networks in urban areas (Tirachini, 2013), for instance how to pick up passengers more efficiently based on identifying demand for taxi services. This can be achieved by using intelligent approaches to predict hot spots relating to locations of where taxis pick up and set down people in an urban area over the course of the day based on historical and real-time geo-spatial data. Data collected on high and low predicted taxi demands over the urban area together with contextual information pertaining to traffic conditions, geospatial distribution of the fleet and vehicle telematics can then be used to provide recommendations to taxi operators for the distribution and optimization of taxi services. User behavior modelling can further be used to recommend real-time re-routings to satisfy personal objectives while relieving congestion. The problem of optimally managing the distribution of taxi services can also be tackled at large transportation hubs such as railway stations and airports to meet passenger demands. For example, an airport Taxi stand passenger queue tracking system can be implemented from real time CCTV camera feeds and the application of vision processing algorithms for identifying and counting individuals waiting in a queue to estimate the number of people entering/exiting and queuing throughput over time. The system can be used to measure length, growth rate and predict the wait time for each queue, which can be visualized in real-time and used to send notifications / alerts based on operator triggers and thresholds.

Our ongoing work is focused on optimizing taxi fleet distribution and routing in context of urban traffic conditions in order to enhance availability, reduce waiting and journey times, and promote fuel economy. In order to achieve the research objectives a novel deep learning based spatial modeling technique was developed and applied which was able to predict hot spots relating to locations of where taxis pick up and set down people in an urban area over the course of the day. The structure and ideas of the computational modelling approach is described in detail in section 4. The data used to train the predictive model were acquired from NYC Open Data webpage. This data include the 2013 Green taxi trip data which contain trip records from all trips completed in green taxis in NYC in 2013. The data records include fields such as capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The data used were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Livery Passenger Enhancement Program (LPEP). An example data sample can be seen in figure 8. The proposed model was utilized in the development of a commercial software application, and achieved above 95% prediction accuracy.

Field	Value
pick up_datetime	12/06/2013 12:11:05 PM
dropoff_datetime	12/06/2013 12:14:06 PM
store_and_fwd_flag	N
rate_code	1
Dropoff_latitute	40.811981201171875
Passenger_count	1
Trip_distance	0.5
Fare_amount	4
Extra	0
MTA_tax	0.5
Tip_amount	0
Tolls_amount	0
Ehail_fee	
Total_amount	4.5
Payment_type	2
Trip_type	
Pickup_longitude	-73.9650650024414
Pickup_latitude	40.8061408996582
Dropoff_longitude	-73.962234497707031

Vendor_id	1

Figure 8. Data sample.

4.3 Social network sentiment analysis

Another interesting application field, which can largely benefit from Big Data research, is sentiment analysis. Sentiment analysis aims at facilitating the processing of high volume data in order to automatically identify the user's evaluations and feelings towards specific products and services (Pang, 2008). Modern Facebook and Twitter users generate huge amounts of data through their posts, comments, and status updates. These data are very rich sources of emotional and cognitive information and can be utilized towards creating very interesting and beneficial sentiment analysis applications. Among the various contexts in which social network sentiment analysis can be beneficial is education. Big Data analytics can be applied towards sentiment analysis purposes on users of e-learning, and computer assisted learning environments in order to enhance the learning experience and promote student's wellbeing (Karyotis, 2017). Understanding the student's feelings and attitude towards the learning process can provide guidelines towards successful adaptation of the taught material or modification of the pedagogy utilized in order to deliver this material. Evidence of the effectiveness of this approach can be found in the work by Tan et al. The results from the team's experimental sessions have highlighted that by incorporating information from social networks, sentiment classification can be notably improved (Tan, 2011). Recent examples in the field of sentiment analysis in e-learning environments can be found in the works by Ortigosa et al. (Ortigosa, 2014) and Martin et al. (Martin, 2012).

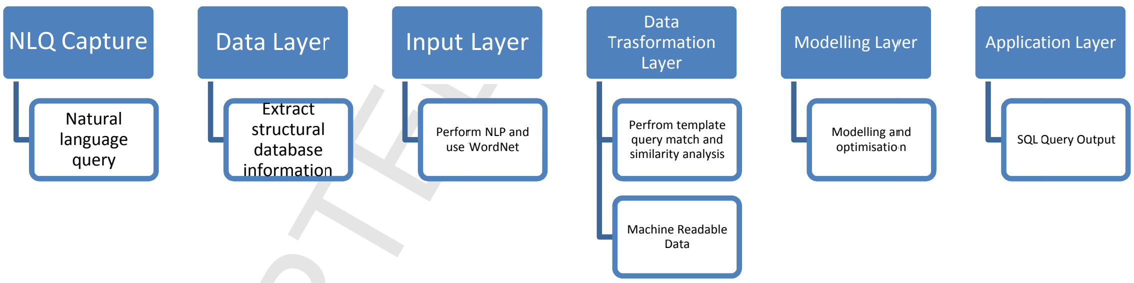
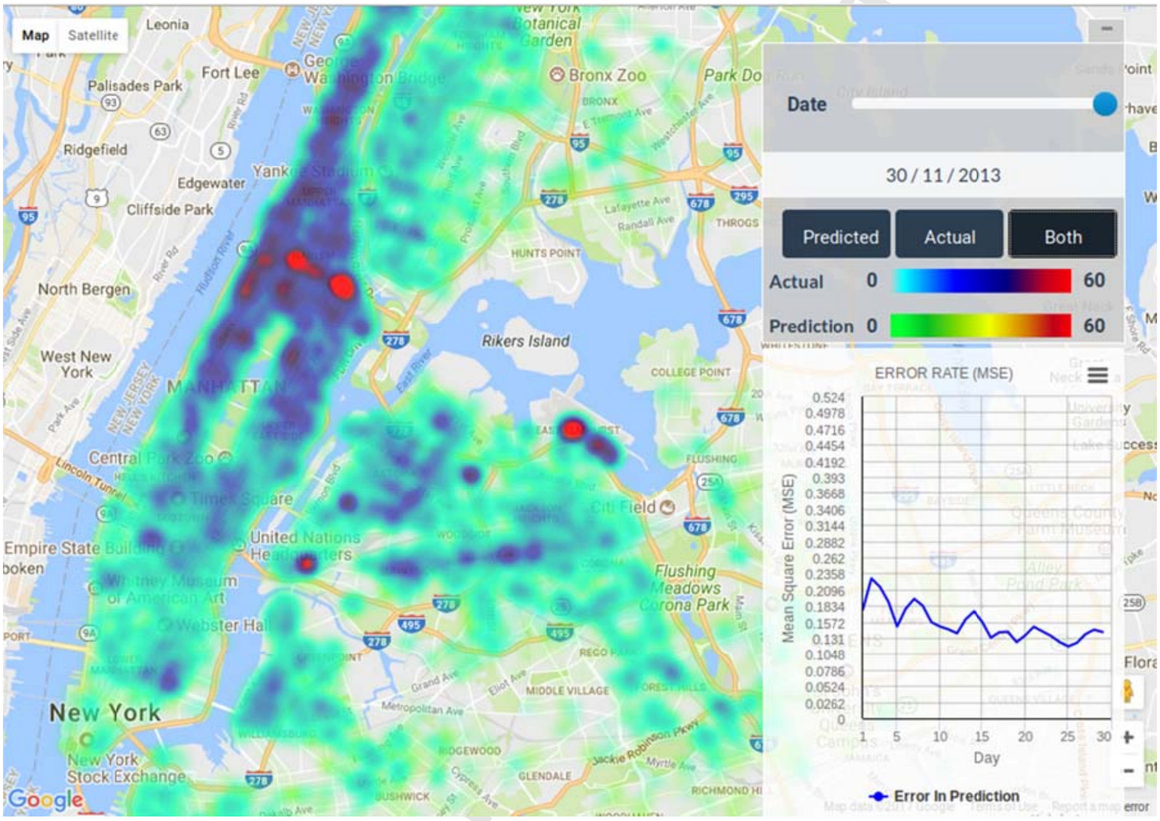


Figure 10. A high-level diagram of the parsing process.

In order to contribute to the goals and objectives of social network sentiment analysis we developed a human-computer interaction platform for natural language processing which can be applied to social media posts and reveal the recipient's sentiment. Social media posts are subject to demanding challenges related to the processing of unstructured and natural text. Therefore, we constructed a powerful semantic, syntactic, and synonym mapper able to translate the text from social media posts into a concept related to the user's sentiment. This concept is a word present in a database containing a large vocabulary, which can be used to describe effectively the user's sentiment and cognitive affective state. Upon entering the post as a natural language query, the proposed system parses the semantic meaning of its components to determine what part of speech they each belong to, and thus what part of the resulting query they are likely to occupy. It achieves this by using a trained model of hundreds of thousands of ‘corpora’ (texts), which are specifically labeled, for the parts of speech tokens (words in a sentence for example). Then, a complete, up-to-date schema is extracted from the database in real-time, in order to find out which tables, columns, subjects etc. the query may be present both in the database and the natural language query. This is done in a variety of ways, like machine semantic analysis using an IEEE-endorsed algorithm, analysing the actual meaning of the query, synonym substitution, and spelling correction. A high-level diagram of the parsing process is shown in figure 10. Following this

Figure 9. Taxi demand prediction application interface.



process, the analysis engine builds a complex relational graph to determine the various relationships between identified parts of the query, including complex, multi-table relationships, from which an SQL query is constructed, once all required relationships are identified.

Gaining an insight of the user's/client's affective state, through the utilization of social media data, can fuel the generation and diversity of intelligent mechanisms able to deliver affective feedback in a wide spectrum of application areas. This feedback could extend from the digital to the physical world, by using standard hardware devices (e.g. smart phones, home appliances, lighting) in the context of intelligent environments. Intelligent cars could modify their engine and cabin configuration to meet their driver's needs and preferences. E-learning systems could provide tailored learning material and tailored encouragements to the students. Automated manufacturing systems could take into account the client's sentiment when designing and developing new products, in order to better satisfy their customer's needs. Tele-home healthcare, robot companions, patient monitoring could be enriched by taking into account the patient's context, behaviors and emtions in order to deliver more efficient patient specific services. Homes with the ability to modify their environment (lighting, temperature etc.) in order to suit their resident's mood.

5. Conclusions

In this paper, we have discussed the importance of big data analytics and computational intelligence techniques. We provided a comprehensive survey of computational intelligence techniques appropriate for the effective processing and analysis of big data. We have presented a data modelling methodology, which introduces a novel biologically inspired universal generative modelling approach called Hierarchical Spatial-Temporal State Machine (HSTSM). We investigated the benefits arising from the utilization of computational techniques namely deep learning neural networks, evolutionary algorithms, and fuzzy logic in Big Data analytics. We identified and highlighted potential novel real life CPS applications arising from the vast amount of information on offer by modern high tech societies, the deployment of intelligent computational techniques, and state of the art solutions to address challenges in these application areas. In this work, a novel approach for Big Data modelling is presented. The proposed methodology relies on a hybrid method, which is based on the structure and functions of the mammalian brain. It incorporates different soft computing techniques and it has the potential to deal with large amounts of data, which are characterized by spatial-temporal correlations. This approach can tackle the high requirements and maximize the potential of dealing with Big Data and therefore can be considered as a state of the art tool for Big Data analytics. The potential benefits arising from this research are numerous and span over a large spectrum of application areas. Utilizing this novel methodology to exploit Big Data's potential can lead to applications with significant impact to knowledge, society, economy, and individuals. Scientific knowledge and research may benefit from revealing hidden patterns in Big Data or by delivering Big Data analysis results in ways, which can be easily visualized and interpreted. Society could profit from the delivery of applications, which promote improved public transportations and health services. E-businesses and organizations could also be assisted through sentiment analysis tools, which contribute at the delivery of products and services which meet their customers' needs. Finally, an individual may benefit through the development of personalized and contextualized products and services, which are able to account effectively for complex related notions, such as their cognitive/affective state. Future work will involve the utilization of the proposed methodology to different application areas in order to create novel models and applications with significant commercial and scientific value and the further improvement of the developed systems.

References

1. Ahmed, S. H., Kim, G., & Kim, D. (2013, November). Cyber Physical System: Architecture, applications and research challenges. In Wireless Days (WD), 2013 IFIP (pp. 1-5). IEEE.
2. Alhabashneh, O., Iqbal, R., Doctor, F., & James, A. (2017). Fuzzy Rule Based Profiling Approach For Enterprise Information Seeking and Retrieval. *Information Sciences, 394–395, July 2017, 18-37.*
3. Alsheikh, M. A., Niyato, D., Lin, S., Tan, H. P., & Han, Z. (2016). Mobile Big Data Analytics Using Deep Learning and Apache Spark. arXiv preprint arXiv:1602.07031

4. Baheti, R., & Gill, H. (2011). Cyber-physical systems. The impact of control technology, 12, 161-166.
5. Behadada, O., Trovati, M., Chikh, M. A., & Bessis, N. (2015). Big data-based extraction of fuzzy partition rules for heart arrhythmia detection: a semi-automated approach. Concurrency and Computation: Practice and Experience.
6. Bekkerman, R., Bilenko, M., & Langford, J. (Eds.). (2011). Scaling up machine learning: Parallel and distributed approaches. Cambridge University Press.
7. Bengio, Y. Learning Deep Architectures for AI. Foundations Trends Machine Learning, vol. 2 (1).
8. Besselink, B., Turri, V., van de Hoef, S. H., Liang, K. Y., Alam, A., Mårtensson, J., & Johansson, K. H. (2016). Cyber–physical control of road freight transport. Proceedings of the IEEE, 104(5), 1128-1141.
9. Bhattacharya, M., Islam, R., & Abawajy, J. (2016). Evolutionary optimization: a big data perspective. *Journal of Network and Computer Applications*, *59*, 416-426.
10. Bing, L., & Chan, K. C. (2014, December). A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data. In *Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing* (pp. 652-657). IEEE Computer Society.
11. Campo, I., Finker, R., Martinez, M., Echanobe, J., & Doctor, F. (2014, July). A real-time driver identification system based on artificial neural networks and cepstral analysis. In Neural Networks (IJCNN), 2014 International Joint Conference on (pp. 1848-1855). IEEE.
12. Chang, V., Kuo, Y. H., & Ramachandran, M. (2016). Cloud computing adoption framework: A security framework for business clouds. Future Generation Computer Systems, 57, 24-41.
13. Chang, V., & Wills, G. (2016b). A model to compare cloud and non-cloud storage of Big Data. Future Generation Computer Systems, 57, 56-76.
14. Chang, V. (2015,a). Towards a Big Data system disaster recovery in a Private Cloud. Ad Hoc Networks, 35, 65-82.
15. Chang, V. (2015b). A proposed framework for Cloud Computing adoption. *International Journal of Web Services Research*.
16. Chen, C., Liu, X., Qiu, T., & Sangaiah, A. K. (2017). A short-term traffic prediction model in the vehicular cyber–physical systems. Future Generation Computer Systems.
17. Chen, X. W., & Lin, X. (2014). Big data deep learning: challenges and perspectives. *Access, IEEE*, *2*, 514-525.
18. Chung, I. H., Sainath, T. N., Ramabhadran, B., Pichen, M., Gunnels, J., Austel, V., ... & Kingsbury, B. (2014, November). Parallel deep neural network training for big data on blue gene/q. In *High Performance Computing, Networking, Storage and Analysis, SC14: International Conference for* (pp. 745-753). IEEE.
19. Djahel, S., Jabeur, N., Barrett, R., & Murphy, J. (2015, May). Toward V2I communication technology-based solution for reducing road traffic congestion in smart cities. In Networks, Computers and Communications (ISNCC), 2015 International Symposium on (pp. 1-6). IEEE.
20. F. Doctor and R. Iqbal "An intelligent framework for monitoring student performance using fuzzy rule-based linguistic summarisation," In H. Abbass, D. Essam and R. Sarker, Eds., Fuzzy Systems (FUZZ-IEEE), IEEE International Conference on. 10-15 June 2012, pp. 1-8.
21. F.Doctor, H. Hagras, V. Callaghan, 'A Fuzzy Embedded Agent-Based Approach for Realizing Ambient Intelligence in Intelligent Inhabited Environments," IEE Transactions on Systems, Man and Cybernatics- Part A: Systems and Humans, vol. 35, no. 1, January 2005, pp. 55-56.
22. Doctor, F., Syue, C. H., Liu, Y. X., Shieh, J. S., & Iqbal, R. (2016). Type-2 fuzzy sets applied to multivariable self-organizing fuzzy logic controllers for regulating anesthesia. *Applied Soft Computing*, *38*, 872-889.
23. Doctor, F., Iqbal, R., & Naguib, R. N. (2014). A fuzzy ambient intelligent agents approach for monitoring disease progression of dementia patients. Journal of Ambient Intelligence and Humanized Computing, 5(1), 147-158.
24. Doctor, F., & Hagras, H. (2013). *U.S. Patent No. 8,515,884*. Washington, DC: U.S. Patent and Trademark Office.

25. Dreier, F. (2002). Genetic Algorithm Tutorial..
26. Duggal, R., Khatri, S. K., & Shukla, B. (2015, September). Improving patient matching: Single patient view for Clinical Decision Support using Big Data analytics. In *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on* (pp. 1-6). IEEE.
27. C. Eaton, D. Deroos, T. Deutsch, G. Lapis and P.C. Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Mc Graw-Hill Companies, 978-0-07-179053-6, 2012
28. Free, C., Phillips, G., Watson, L., Galli, L., Felix, L., Edwards, P., ... & Haines, A. (2013). The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med, 10(1), e1001363.
29. B.Gerhardt, K. Griffin and R. Klemann, "Unlocking Value in the Fragmented World of Big Data Analytics", Cisco Internet Business Solutions Group, June 2012,http://www.cisco.com/web/about/ac79/docs/sp/Information-Infomediaries.pdf
30. Gurney, K. (1997) An introduction to neural networks. London: UCL press
31. Harding, C., Ilves, P., & Wilson, S. (2015). Digital mental health services in general practice. Br J Gen Pract, 65(631), 58-59.
32. Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. Information Systems, 47, 98-115.
33. Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... & Chiroma, H. (2016). The role of big data in smart city. International Journal of Information Management, 36(5), 748-758.
34. Haykin, S. S., Haykin, S. S., Haykin, S. S., & Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River: Pearson Education.
35. Hawkins, J. 2004, On intelligence, Times Books, New York.
36. Hawkins, J. & George, D. (2006), Hierarchical Temporal Memory: Concepts, Theory, and Terminology.
37. Hawkins, J. George, D. Niemasik, J. (2009), Sequence memory for prediction, inference and behavior, Philosophical Transactions of the Royal Society, Published 30 March 2009.DOI: 10.1098/rstb.2008.0322
38. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504-507.
39. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE*, *29*(6), 82-97.
40. Hirsch, D. D. (2013). Glass House Effect: Big Data, the New Oil, and the Power of Analogy, The. Me. L. Rev., 66, 373.
41. Hossain, M. S. (2017). Cloud-supported cyber–physical localization framework for patients monitoring. IEEE Systems Journal, 11(1), 118-127.
42. Iqbal, R., Grzywaczewski, A., Halloran, J., Doctor F., Iqbal, K., (2015): "Design implications for task-specific search utilities for retrieval and reengineering of code", Enterprise Information Systems, Taylor and Francis, pp 1751-7575. doi: 10.1080/17517575.2015.1086494
43. Iqbal, K, Odetayo M. James, A., Iqbal, R., Kumar, N., Bama S.,(2015): "An efficient image retrieval scheme for colour enhancement of embedded and distributed surveillance images", Journal of Neurocomputing, Vol 174, Part 2, pp. 413-430
44. Iqbal, R., Shah, N., James, A., & Duursma, J. (2011). ARREST: From work practices to redesign for usability. Expert Systems with Applications, 38(2), 1182-1192.
45. Iqbal, R., Shah, N., James, A., & Cichowicz, T. (2013). Integration, optimization and usability of enterprise applications. Journal of Network and Computer Applications, 36(6), 1480-1488.

46. Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In System Sciences (HICSS), 2013 46th Hawaii International Conference on (pp. 995-1004). IEEE.
47. Kang, C.C., Chuang, Y.J., Tung, K.C., Chao, C.C., Tang, C.Y., Peng, S.C., and Wong, D.S.H. (2011) 'A genetic algorithm-based boolean delay model of intracellular signal transduction in inflammation'. BMC bioinformatics 12(1), S17
48. Karyotis, C., Doctor, F., Iqbal, R., James, A., & Chang, V. (2017). A fuzzy computational model of emotion for cloud based sentiment analysis. Journal of Information Sciences.
49. Karyotis, C., Doctor, F., Iqbal, R., James, A., & Chang, V. (2016). A Fuzzy Modelling Approach of Emotion for Affective Computing Systems.
50. Karyotis, C., Doctor, F., Iqbal, R., and James, A. (2015) An Intelligent Framework for Monitoring Students Affective Trajectories Using Adaptive Fuzzy Systems. IEEE International Conference on Fuzzy Systems, 2-5 August 2015, Istanbul, Turkey
51. Kattan, A., Doctor, F., & Arif, M. (2015, October). Two Brains Guided Interactive Evolution. In Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on (pp. 3203-3208). IEEE.
52. Kissick, W. L. (1994). Medicine's dilemmas: infinite needs versus finite resources. Yale University Press.
53. Korekado, K., Morie, T., Nomura, O., Ando, H., Nakano, T., Matsugu, M., & Iwata, A. (2003, September). A convolutional neural network VLSI for image recognition using merged/mixed analog-digital architecture. In *Knowledge-Based Intelligent Information and Engineering Systems* (pp. 169-176). Springer Berlin Heidelberg.
54. http://www.landrover.com/experiences/news/self-driving-car.html. As accessed online on 24 May 2016.
55. Le, Q. V. (2013, May). Building high-level features using large scale unsupervised learning. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 8595-8598). IEEE.
56. Lin, K. C., Zhang, K. Y., Huang, Y. H., Hung, J. C., & Yen, N. (2016). Feature selection based on an improved cat swarm optimization algorithm for big data classification. *The Journal of Supercomputing*, 1-12.
57. Liu, Y. X., Doctor, F., Fan, S. Z., & Shieh, J. S. (2014). Performance Analysis of Extracted Rule-Base Multivariable Type-2 Self-Organizing Fuzzy Logic Controller Applied to Anesthesia. BioMed research international, 2014.
58. Lu, S., Li, R. M., Tjhi, W. C., Lee, K. K., Wang, L., Li, X., & Ma, D. (2011, November). A framework for cloud-based large-scale data analytics and visualization: Case study on multiscale climate data. In Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on (pp. 618-622). IEEE.
59. Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2015). Traffic flow prediction with big data: A deep learning approach. *Intelligent Transportation Systems, IEEE Transactions on*, *16*(2), 865-873.
60. Mahmud, S., Iqbal, R., & Doctor, F. (2016). Cloud enabled data analytics and visualization framework for health-shocks prediction. Future Generation Computer Systems.
61. Maniak, T., Iqbal, R., Doctor, F. (2016): "*A Method for Monitoring the Operational State of a System*", Interactive Coventry Ltd Pending Patent Reference No: 1607820.6
62. Maniak, T., Iqbal, R., Doctor F., Jayne, C., (2013): "Automated Sound Signalling Device Quality Assurance Tool For Embedded Industrial Control Applications", Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC), pp., 4812-4818.
63. Maniak, T., Jayne, C., Iqbal, R., & Doctor, F. (2015). Automated intelligent system for sound signalling device quality assurance. Information Sciences,294, 600-611.

64. Martin, J. M., Ortigosa, A., & Carro, R. M. (2012, October). SentBuk: Sentiment analysis for e-learning environments. In Computers in Education (SIIE), 2012 International Symposium on (pp. 1-6). IEEE.
65. Mendel, J. M. (2001). Uncertain rule-based fuzzy logic systems: introduction and new directions.
66. Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. Jama, 309(13), 1351-1352.
67. A., Ortigosa, J. M., Martín and R. M. Carro. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, 31, 527-541.
68. Pang, Z., Zheng, L., Tian, J., Kao-Walter, S., Dubrova, E., & Chen, Q. (2015). Design of a terminal solution for integration of in-home health care devices and services towards the Internet-of-Things. Enterprise Information Systems, 9(1), 86-116.
69. B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1–135, 1 2008.
70. Parpinelli, R. S., & Lopes, H. S. (2011). New inspirations in swarm intelligence: a survey. *International Journal of Bio-Inspired Computation*, *3*(1), 1-16.
71. Poli, R., Langdon, W. B., McPhee, N. F., & Koza, J. R. (2008). *A field guide to genetic programming*.
72. Qureshi, F. F., Iqbal, R., & Asghar, M. N. (2017). Energy Efficient Wireless Communication Technique Based on Cognitive Radio for Internet of Things. Journal of Network and Computer Applications.
73. Razavi, S. H., Ebadati, E. O. M., Asadi, S., & Kaur, H. (2015). An efficient grouping genetic algorithm for data clustering and big data analysis. In *Computational Intelligence for Big Data Analysis* (pp. 119-142). Springer International Publishing.
74. Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In Collaboration Technologies and Systems (CTS), 2013 International Conference on (pp. 42-47). IEEE.
75. Sultan, N. (2014). Making use of cloud computing for healthcare provision: Opportunities and challenges. International Journal of Information Management,34(2), 177-184.
76. Sun, J., Zhang, Q., & Tsang, E. P. (2005). DE/EDA: A new evolutionary algorithm for global optimization. Information Sciences, 169(3), 249-262.
77. Suthaharan, S. (2014). Big data classification: Problems and challenges in network intrusion prediction with machine learning. ACM SIGMETRICS Performance Evaluation Review, 41(4), 70-73.
78. Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., & Li, P. (2011, August). User-level sentiment analysis incorporating social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1397-1405). ACM.
79. Tirachini, A. (2013). Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services. Transportation Research Part C: Emerging Technologies, 30, 239-256.
80. Tolk, A. (2015, July). The next generation of modeling & simulation: integrating big data and deep learning. In *Proceedings of the Conference on Summer Computer Simulation* (pp. 1-8). Society for Computer Simulation International.
81. Vegni, A. M., Biagi, M., & Cusani, R. (2013). Smart vehicles, technologies and main applications in vehicular ad hoc networks. INTECH Open Access Publisher.
82. Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. White Paper, IDC.
83. Wan, J., Zhang, D., Zhao, S., Yang, L., & Lloret, J. (2014). Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutions. IEEE Communications Magazine, 52(8), 106-113.
84. Wang, J., Abid, H., Lee, S., Shu, L., & Xia, F. (2011). A secured health care application architecture for cyber-physical systems. arXiv preprint arXiv:1201.0213.
85. Whitley, D. (2001). An overview of evolutionary algorithms: practical issues and common pitfalls. *Information and software technology*, *43*(14), 817-831.

86. D. Wu, Using Physiological Signals: Virtual Reality Stroop Task, IEEE Transactions on Affective Computing vol. 1, no. 2, (2010), pp. 109–118.
87. D. Wu, “Fuzzy sets and systems in building closed-loop affective computing systems for human-computer interaction: Advances and new directions,” IEEE World Congress on Computational Intelligence, Brisbane, Australia, June 10-15, 2012.
88. Yaqoob, I., Chang, V., Gani, A., Mokhtar, S., Hashem, I. A. T., Ahmed, E., ... & Khan, S. U. (2016). Information fusion in social big data: Foundations, state-of-the-art, applications, challenges, and future research directions. International Journal of Information Management.
89. Zadeh, L.A. (1965). ‘Fuzzy sets’. Information and Control 8 (3): 338–353
90. Zaidi, N. A., Webb, G. I., Carman, M. J., & Petitjean, F. (2015). Deep broad learning-Big models for Big data. arXiv preprint arXiv:1509.01346
91. Zhang, Y., Qiu, M., Tsai, C. W., Hassan, M. M., & Alamri, A. (2017). Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. IEEE Systems Journal, 11(1), 88-95.

.

Big Data Analytics: computational intelligence techniques and application areas

Profiles

Dr Rahat Iqbal is Managing Director of Interactive Coventry Ltd and a Reader/Associate Professor in the Faculty of Engineering, Environment and Computing at Coventry University. He has a track record of project management and leadership of industrial projects funded by EPSRC, TSB, ERDF and local industries (e.g. Jaguar Land Rover Ltd, Trinity Expert Systems Ltd). He was involved in the project management and development of the EU FP7 project CHIL (Computers in Human Interaction Loop) at the Technical University of Eindhoven , Netherlands. Recently, he has successfully led a project in collaboration with Jaguar Land Rover on self-learning car for predicting driver's behaviour for personalisation of telematics and optimisation of route planning. He has managed many industrial projects, in Intelligent Systems, Predictive Modelling, User Behaviour, Information Retrieval and Fault Detection. He has published more than 100 papers in peer-reviewed journals and reputable conferences and workshops. Dr Iqbal is on the programme committee of several international conferences and workshops. He is also a fellow of the UK Higher Education Academy (HEA). Dr Iqbal has also edited several special issues of international journals within the field of Information Retrieval and User Supportive systems.

Dr Faiyaz Doctor is Research Director of Interactive Coventry Ltd and a Senior Lecturer in the Faculty of Engineering, Environment and Computing at Coventry University. He has worked in industry and academia to develop Novel Computational Intelligence Solutions addressing real world problems related to smart environments, Energy Optimisation, Predictive Analytics and Collaborative Decision Support. His work has resulted in high profile innovation awards (best KTP regional finalist 2011, Lord Stafford award for Innovation) and an international patent on improved approaches for Data Analysis and Decision-Making using Hybrid Neuro-Fuzzy and Type-2 Fuzzy Systems: wo/2009/141631. He has led and co-led projects funded through the Newton Fund and Conacyt (Mexico), the TSB in collaboration with local industry (e.g. Jaguar Land Rover Ltd) and though collaborative consultancy with international partners (e.g. Ministry of Labour, Saudi Arabia). Dr Doctor has published over 50 papers in peer reviewed international journals, conferences and workshops. He regularly serves on organisation and programme committees of several international conferences and workshops in the field of Computational Intelligence. He is also a member of the IEEE and IEEE Computational Intelligence Society.

Dr Brian More is representing Coventry University Enterprise Ltd as Director of Intellectual Property (IP) and serves on the Board of Directors of Interactive Coventry Ltd. He works as Director for Intellectual Property at Coventry University with responsibility for policy, protection, valuation and commercialisation of all forms of IP. He manages a portfolio of 20 patent families, trademarks, designs and copyright. He has had 25 years experience working with Intellectual Property, is an inventor on 6 patents and jointly owns 3 trade marks. Dr More has been active in starting 15 companies using IP and attracting investment into them. He is passionate about education and training in the IP arena and developed accredited courses for academic and industry use. He has published 20 peer reviewed papers on research, innovation and entrepreneurship. He worked at CEA, NPL and BNFL's Company Research Laboratory. He is a Director of 2 companies and sits on 3 national advisory panels. He sits on the Board of Trustees of the Institute of Nanotechnology prior to which he was Chairman of the Steering Group to the Institute. He has worked for private contractors on assessment of development proposals in the field of Nanotechnology and worked on EU Framework projects as commercialisation consultant.

Dr Shahid Mahmud is Chairman and Chief Executive Officer of Interactive Group. He has more than 31 years of professional experience in the field of ICT. He has served on various federal committees of the Government of Pakistan addressing the formulation and implementation of the National Telecom and IT policies. Dr Mahmud is 2016 Distinguished Eisenhower Fellow. He is also a Senior Fellow, Global Think Tank Network (GTTN) and Co Chair for ICT on the Corporate Advisory Council of the National University of Science and Technology (NUST). He is active in several philanthropic activities, working with youth-oriented and community service projects such as Buraq Planetary Society, TRUCE, Begum Mehmooda Welfare Trust and Zubaida Khaliq Memorial Free Hospital. He has been the founder director and shareholder of Paktel Limited, Indus Vision, Pak Globalstar (Pvt) Limited, SHOA (Pvt) Limited, and Shaheen Pay TV (Pvt) Limited. He has also served as a Director of Askari Bank for over six years. In recognition for having spent his entire career in promoting IT in Pakistan, Dr Mahmud was given the Lifetime Achievement Award: at the 12th Teradata National IT Excellence Awards for 2014.

Mr. Usman Yousuf is Chief Executive Officer UAE at Interactive Group. His primary responsibilities include developing and implementing business strategy for the Middle East operation, forging key partnerships and alliances, and directing technical and project teams. In addition to being a certified Project Management Professional (PMP), he is also certified in logistics and The Open Group Architecture Framework 9 (TOGAF 9 Level 2). He has completed training in Strategic visioning, time management and negotiation and attended numerous workshops and trainings over the course of his career. He has worked with the United Nations Office for Project Services (UNOPS) in his previous role in business development at a regional IT firm. He is actively involved with the Buraq Planetary Society (www.buraqsociety.org) as Vice Chairman and volunteers his time for the development of Pakistan's youth. He has lived in the UAE for over 12 years and is fluent in English, Urdu and Arabic.

Big Data Analytics: computational intelligence techniques and application areas

Highlights

1. We highlighted the importance of Big Data in modern life and economy.
2. We investigated the benefits of computational intelligence techniques namely deep learning neural networks, evolutionary algorithms and fuzzy logic in big data analytics.
3. We presented a novel data modelling methodology which introduces a novel biologically inspired universal generative modelling approach called Hierarchical Spatial-Temporal State Machine (HSTSM).
4. We explored the potential of the powerful combination of Big Data and Computational intelligence and identified a number of areas where novel applications in real world problems can be developed.
5. We have also discussed various aspects of policy, protection, valuation and commercialization related to big data.