Big data analytics: Computational intelligence techniques and application areas

Rahat Iqbal
Faculty of Engineering, Environment & Computing, Coventry University, United Kingdom
Interactive Coventry Ltd, Coventry University Technology Park, Puma Way, Coventry, United Kingdom

Faiyaz Doctor
Interactive Coventry Ltd, Coventry University Technology Park, Puma Way, Coventry, United Kingdom
School of Computer Science and Electronic Engineering, University of Essex, United Kingdom

Brian More
Interactive Coventry Ltd, Coventry University Technology Park, Puma Way, Coventry, United Kingdom
Coventry University Enterprises Ltd, Coventry, University Technology Park, Puma Way, Coventry, United Kingdom

Shahid Mahmu
Interactive Coventry Ltd, Coventry University Technology Park, Puma Way, Coventry, United Kingdom

Usman Yousuf
Interactive Coventry Ltd, Coventry University Technology Park, Puma Way, Coventry, United Kingdom

Abstract

Big Data has significant impact in developing functional smart cities and supporting modern societies. In this paper, we investigate the importance of Big Data in modern life and economy, and discuss challenges arising from Big Data utilization. Different computational intelligence techniques have been considered as tools for Big Data analytics. We also explore the powerful combination of Big Data and Computational Intelligence (CI) and identify a number of areas, where novel applications in real world smart city problems can be developed by utilizing these powerful tools and techniques. We present a case study for intelligent transportation in the context of a smart city, and a novel data modelling methodology based on a biologically inspired universal generative modelling approach called Hierarchical Spatial-Temporal State Machine (HSTSM). We further discuss various implications of policy, protection, valuation and commercialization related to Big Data, its applications and deployment.

1. Introduction

The importance of data in our increasingly information driven economy and society can be summarized in the statement that "big data is the new oil" as was recently quoted by IBM's chief Executive officer "(Hirsch, 2013). The analogy of the importance data to that of the world's reliance on this natural resource has been identified and highlighted by various studies demonstrating the power and impact of Big Data to modern life. This new digital resource can provide the main driving force for building the smart cities of tomorrow. Big data can be defined as huge sets of data with a structure of increasing variety and complexity. The inherent difficulty in dealing with these large amounts of data results in major challenges concerning their storage and analysis as well as the cost and time efficient delivery of results. Moreover, information should be delivered in an interpretable and easily visualized way (Sagiroglu and Sinanc, 2013). Big Data analytics refers to the techniques utilized in order to examine, process, discover and expose hidden underlying patterns, interesting relations and other insights concerning the application context under investigation.

As pointed out by Hashem et al. Big Data have three main characteristics. Firstly, as the name reveals the data themselves are numerous. Secondly, it is not possible to categorize the data into regular relational databases, and finally data streams are created, captured, and analyzed rapidly (Hashem et al., 2015). As Gerhard mentions, Big Data is a revolutionary leap forward from traditional analysis which possesses three main characteristics volume, variety, and velocity (Gerhardt et al., 2012). Volume refers to the amount of data, which are created and stored. Variety is related to the various types of data collected, and velocity can be defined as the speed of data generation, streaming and aggregation (Kaisler et al., 2013). In Kaissler et al.'s study, data value, and complexity are also proposed as Big Data characteristics. Data value is a measure of the usefulness of data in decision making processes, while complexity is a measure of the degree of interdependence and interconnectedness in Big Data structures.

Nowadays, in modern smart cities, there is a growth in the amount of data, which can be captured and utilized. Recent advances in hardware and software technologies, such as social media, the internet of things, wearable sensors, mobile technologies, data storage and cloud computing, data mining techniques and machine learning algorithms have resulted in the ability to easily acquire, analyze and store large amounts of data from different kinds of quantitative and qualitative domain specific data sources. This data can be harvested from a large number of diverse sources including emails and online transactions, multimedia information such as video audio and pictures, large databases containing health records and other information, information captured during a user's interaction with social media such as posts, status updates etc., data derived from the 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 or other devices inside a smart home, data from scientific research and others (Eaton et al., 2012). It can be easily concluded form the above that in the modern world, data is generated at a record rate (Villars et al., 2011). This fact has aided Big Data's emergence as a very important subject, which continually gained interest from both academia and industry.

The potential utilization of this huge amount of information to transform the way of life in modern smart cities, catapults Big Data and Big data analysis to the forefront of modern research communities, businesses, and governments (Hashem et al., 2015), delivering the promise of a countless new application areas and opportunities, which can emerge under completely diverse

application contexts from transportation (Hashem et al., 2016) to health care (Murdoch and Detsky, 2013). As a result, the benefits arising from this wealth of knowledge and information can affect research in numerous ways. This includes for example promoting medical advances by providing evidence in the identification of symptoms and patterns concerning diseases, pandemics or modern health issues, or aiding in the creation of large ground truth databases for scientific fields such as emotion research, and affective computing which are in desperate need of vast amounts of data in order to successfully create emotion models and effective emotion recognition techniques. Modern economy and businesses in the context of a smart city, 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 networks or smart devices in order to identify the users' preferences, or recognize unsatisfied needs of modern clients or understand the relationships between competitive and collaborating organizations, thus creating better and more appealing services and products, or improving already existing ones. Acquiring more nuanced insights of customer preferences and needs will provide modern organizations and businesses with a crucial advantage over competitors (Sagiroglu and Sinanc, 2013). 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). Fast evolution of ICT technologies, Big Data analytics, and energy efficient communication protocols has provided a new momentum to e-businesses and global connectivity (Qureshi et al., 2017). Big Data analytics can also facilitate government's and smart city local authorities' efforts towards delivering better services to their citizens. Big Data can aid governments in improving healthcare, public transport, education, and other fields of social life thus aiding in shaping a more efficient modern society. For example, data from traffic records can be utilized towards the improvement of public transport services delivered by the state to the population. Especially IoT related Big Data applications such as smart homes, combinations of wearable health devices, and smart cities are widely recognized as key factors which can contribute in the economic and social development of developing countries (Mital et al., 2017).

Utilizing Big Data poses a number of challenges such as security issues and robust handling and storage of data in emergency situations. Enterprises and organizations store, and analyze huge amounts of data, which should be processed in a secure way. This is a very challenging task in the modern environment. Traditional tools are not able to tackle this problem, and this fact is demonstrated in recent studies. This happens due to various reasons such as the centralised nature of Big Data stores or the enhanced capabilities of attackers to penetrate and nuteralise traditional security systems (Sagiroglu and Sinanc, 2013). In order to tackle these challenges companies should incorporate risk aware, contextual, and agile security models (Sagiroglu and Sinanc, 2013). As mentioned in Amin et al.'s paper, there is an identified need for lightweight intelligent security protocols that allows only legitimate users to access the sensitive information collected by various smart devices in the IoT environment (Amin et al., 2018).

In order to harvest 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 et al., 2015; Iqbal et al., 2015a, Iqbal et al., 2015b). Big challenges stem from the utilization of Big Data in 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 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 analyzed and understood by human users. Additionally, data can be accumulated from diverse sources and input channels, making the online processing very demanding due to the variety of signal input, which needs 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 that are also not always precise or complete due to various sources of imprecision, uncertainty together with 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 tries 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).

In the following section, we aim to explore the potential of applying Computational Intelligence techniques to Big Data analytics, and we present recent applications that utilize these techniques towards Big Data analytics tasks. In Section 3, we discuss potential areas where intelligent machine learning techniques can be applied on Big Data analytics resulting in novel and interesting smart city applications. In Section 4, we present our novel data modelling methodology approach. In Section 5, we consider different aspects of policy, protection, valuation and commercialization related to Big Data. Finally, in Section 6, we discuss the conclusions arising from this study.

2. Computational intelligence for big data analytics

Machine learning (ML) approaches offer a means for modelling patterns and correlations in data in order to discover relationships and make predictions based on unseen events. ML approaches consist of supervised learning (learning from labeled 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 triad 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 et al., 2016;

Liu et al., 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; Parpinelli and Lopes, 2011; Poli et al., 2008) can be used for optimizing complex real-world systems and processes. Finally ANNs enable feature extraction and learning from experiential data (Haykin et al., 2009) and are based on imitating the parallel processing and data representation structure of neurons in animal and humans brains.

In Big Data analytics, there is a growing need to accurately identify important features in the data affecting the outputs and determine the spatial co-relations between input variables at a given point in time as well as the causal or temporal co-relations between inputs 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 et al., 2011; Bengio, 2009; Campo et al., 2014; Hinton et al., 2012; Le, 2013). These approaches can offer 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, where training is both unsupervised (bottom-up) 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 and Salakhutdinov, 2006). Feature learning methods are based on supervised approaches such as Deep NNs, Convolutional NNs and Recurrent NNs along with unsupervised techniques such as Deep Belief Networks and Convolutional Neural Networks (CNNs) provide deep architecture that combine structural elements of local receptive fields, shared weights, and pooling that aim to imitate the processing of simple and complex cortical cells found in animal vision systems (Korekado et al., 2003).

The potential of utilizing deep learning techniques in Big Data analytics have been highlighted by recent review studies (Chen and Lin, 2014; Tolk, 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 Deep learning methods to handle cases where the sheer amount of data is huge is also discussed in the work by Chen et al. (Chen and Lin, 2014) where the key role which deep learning is coming to play in Big Data analytics solutions is demonstrated. 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 have presented an algorithm for Deep Broad Learning which can be tuned in order to specify the depth of the model and have achieved notable accuracy performance for large amount of data which can be considered competitive with other state of the art research (Zaidi et al., 2015). In (Alsheikh et al., 2016) the researcher's try to address 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 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 apply a deep learning technique, which takes into account spatial and temporal correlations in order to utilize Big Data towards traffic flow prediction purposes, achieving a high performance (Lv et al., 2015). In Chung et al.'s paper, it is stated that deep neural networks have a really high performance in pattern recognition tasks. However, they are computationally expensive as it concerns training time, which in some cases can reach an increase of a factor of 10 compared to other approaches (Chung et al., 2014). More specifically, the researchers have investigated deep neural network training by utilizing the data parallel Hessian-free 2nd order optimization algorithm. Their experimental results that were calculated on large scale speech tasks have demonstrated a significant performance increase without decreasing the accuracy thus allowing deep neural network training by utilizing Big Data in a logical amount of time.

Big Data analytics also bear challenges concerning the nature of the data accumulated from diverse and potentially noise infected sources. These data 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 which contain large amounts of uncertainty that is related to 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 with other more complex techniques, while at the same time retaining a satisfactory tradeoff between classification accuracy and time performance (Karyotis et al., 2015; Wu, 2010). 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 successful visualization of hidden relations existing in data thus allowing the users of applications or researchers searching for hidden patterns in data to easily visualize these underlying relations (Doctor and Iqbal, 2012). Finally, Fuzzy Logic systems and more specifically adaptive fuzzy logic Systems have demonstrated a very good potential concerning their ability to model and account for individual differences and contextual information with a very reasonable computational burden thus making them a very good choices for creating personalised and user-centered systems (Doctor et al., 2005; Karyotis et al., 2015; Karyotis et al., 2018).

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 excellent tradeoff between accuracy and interpretability (Behadada et al., 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 (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 show that this system has a very good prediction performance while at the same time having low processing times (Bing and Chan, 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. Their developed system has

demonstrated great flexibility and was able to handle data from various sources (Duggal et al., 2015).

Evolutionary algorithms (EA) are another CI technique, which can meet the requirements and challenges of Big Data analytics. EA are very good explorers of the search space which makes them very good candidates of Big Data analysis, since Big Data are subjects of a very high degree of dimensionality and sparseness (Bhattacharya et al., 2016). This problem was investigated in the paper by Bhattacharya et al., where the researchers have developed an EA with the ability to deal with both issues. Despite the fact that as it is acknowledged in their paper their application was not Big-Data ready, when it was evaluated for benchmark problems, it was shown that the developed approach had a very satisfying performance 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 et al., 2015), feature selection (Lin et al., 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 et al., 2015).

The 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 optimized 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 characterised by large amounts of vague or complex information requiring analysis for making operational and cost effective decisions (Doctor and Hagras, 2013).

3. 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 from a variety of different application areas, which can contribute to the creation of a truly intelligent digital environment in the context of a smart city.

3.1. Smart city projects

In the area of smart cities, several state of art research efforts have emerged demonstrating the potential of using computational intelligence techniques. In the work by Asaro et al., the researchers explored the deployment of computational intelligence techniques, and their potential to facilitate the task of understanding linguistic registers and improving citizen communication, which is a key part of a smart city's development (D'Asaro et al., 2017). Esposito et al. argued about the important role of intelligent techniques in smart city environments. In their work, the application of several machine-learning techniques in pervasive air quality monitoring was presented for the calibration of different components, towards pollutant concentration estimation (Esposito et al., 2017). Fuzzy Logic is used by many state of the art applications and research efforts. Costa and al. presented a Fuzzy-Logic based approach, which can be used in order to develop multi-systems smart city applications based on visual monitoring (Costa et al., 2017). Fuzzy Logic was also utilized as a computational base for producing alerts concerning urban flooding phenomena, in the work by Melo et al. The researchers proposed, and tested successfully, a fuzzy model for urban open data providing accurate flood alerts (Melo et al., 2016). Neural Networks have also been used

widely from recent research efforts in the smart city context. A prominent application area in this context, is the prediction of traffic flows. Habtie et al. proposed the deployment of a neural network based model in order to forecast traffic flows on urban road networks. Their experimental sessions relied on simulation and real world data, and highlighted the model's potential for producing accurate results (Habtie et al., 2016). Neural networks have also been used successfully in other smart city application areas. Lwoski et al. presented a state of the art regional detection system that utilized deep convolution networks to provide real-time pedestrian detection. Their approach achieved an accuracy of 95.7%, while at the same time managed to be simple and fast thus making it suitable for real time deployment (Lwowski et al., 2017). Genetic algorithms have also been used in the past for providing effective smart city solutions, and are the computational backbone of very recent research efforts. Fujdiak et al. utilized a genetic algorithm for optimizing municipal waste collection, a key and challenging area for a successful smart city. The proposed approach improved the logistic procedure of waste collection, by calculating optimal garbage-truck routes (Fujdiak et al., 2016). It is usual to combine the benefits of GA, with other computational intelligence techniques in order to provide optimized solutions for smart city problems. A demonstrative example is the work by Vlahogianni et al. In this work, the researchers used a genetically optimized multilayer perceptron in order to provide a solution to the parking availability problem. Their approach was able to predict with a reasonable accuracy region parking occupancy rates up to 30 min in the future, by utilizing data from the smart city of Santader in Spain (Vlahogianni et al., 2016). Gupta et al. proposed the utilization of a Hopfield Neural Network (HNN) in order to optimize traffic light sequence for an intersection, and calculated the optimal green time for the traffic lights by utilizing a GA. With the deployment of this method the traffic flow inside a smart city could be enhanced (Gupta et al., 2017). There are several smart city example projects in the United Kingdom. These projects include but are not limited to intelligent transportation applications. Demonstrative recent UK-based smart city project examples are Milton Keynes smart transportation system, Peterborough's virtual model, Cambridge intelligent technology for traffic management, air quality, energy network and health and social care and others (UK Smart Cities Index, 2016).

As demonstrated by the aforementioned recent research examples, intelligent transportation is an key application area which can greatly benefit from the deployment of CI and Big Data Analytics in the context of a modern smart city. 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. 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 et al., 2015; Vegni et al., 2013). These initiatives are supported by the development and deployment of technologies such as Vehicular Adhoc Networks (VANET) and the more recent Vehicular Content Centric Networks (VCCNs), which are designed to retrieve and distribute content fast and effectively in the vehicular environment (Wahid et al., 2017). A demonstrative smart automotive initiative taking advantage of modern ICT technologies is self-learning cars. Self-learning cars aim to enhance driving experience through the smart management of infotainment delivery, to pre-emptively handle 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. 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 unitization. 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 within urban locations while optimizing fleet routing and distribution to improve availability, waiting, journey times and fuel economy objectives in context of urban traffic conditions. This can be achieved by using intelligent approaches to predict hot spots relating to locations of where taxi's 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 cameras 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. There are numerous data sources and sensors, which can contribute to the development of effective smart transportation applications. These data sources and sensory equipment can be seen and are briefly described in Fig. 1. (See Fig. 2, Fig. 3.)
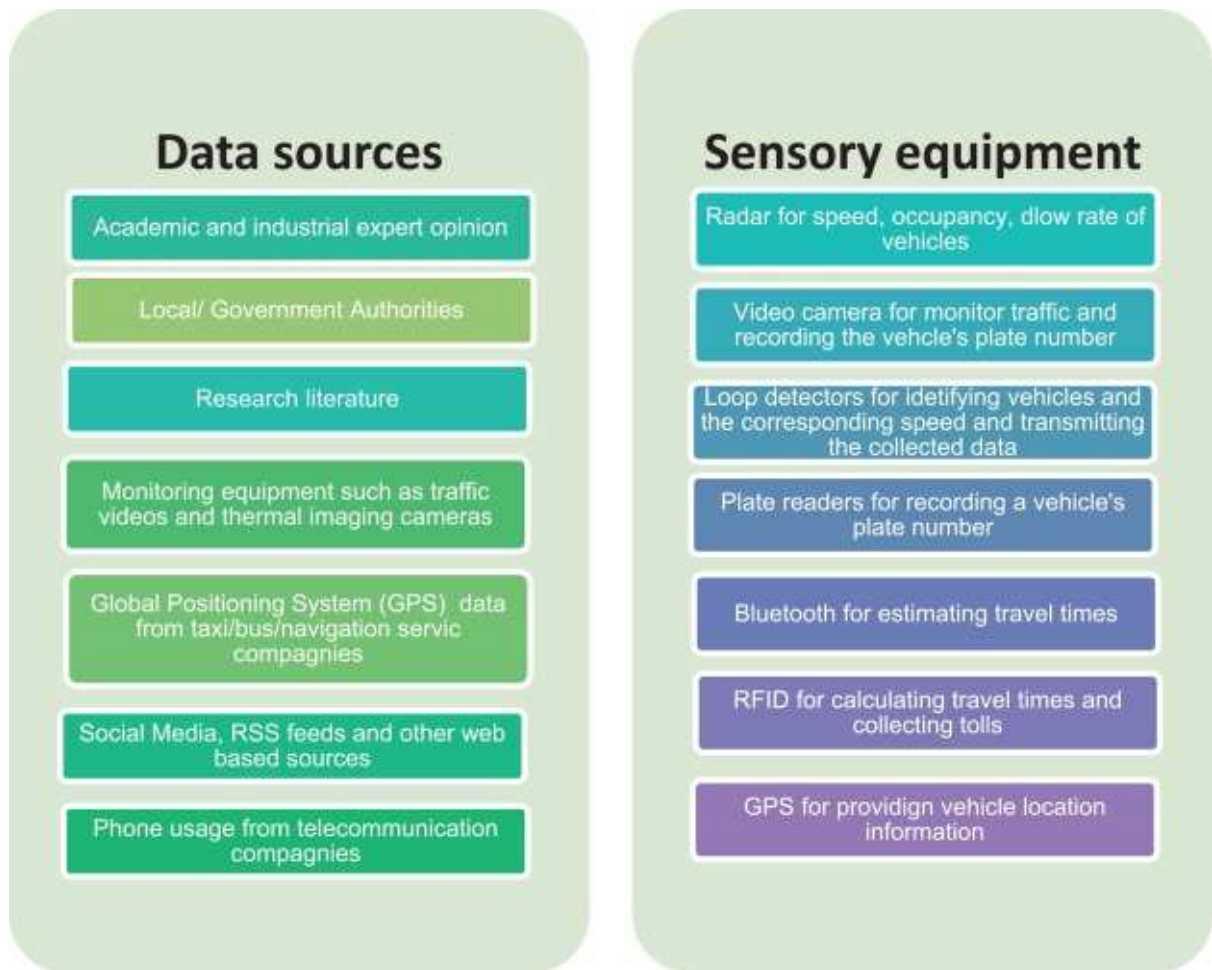
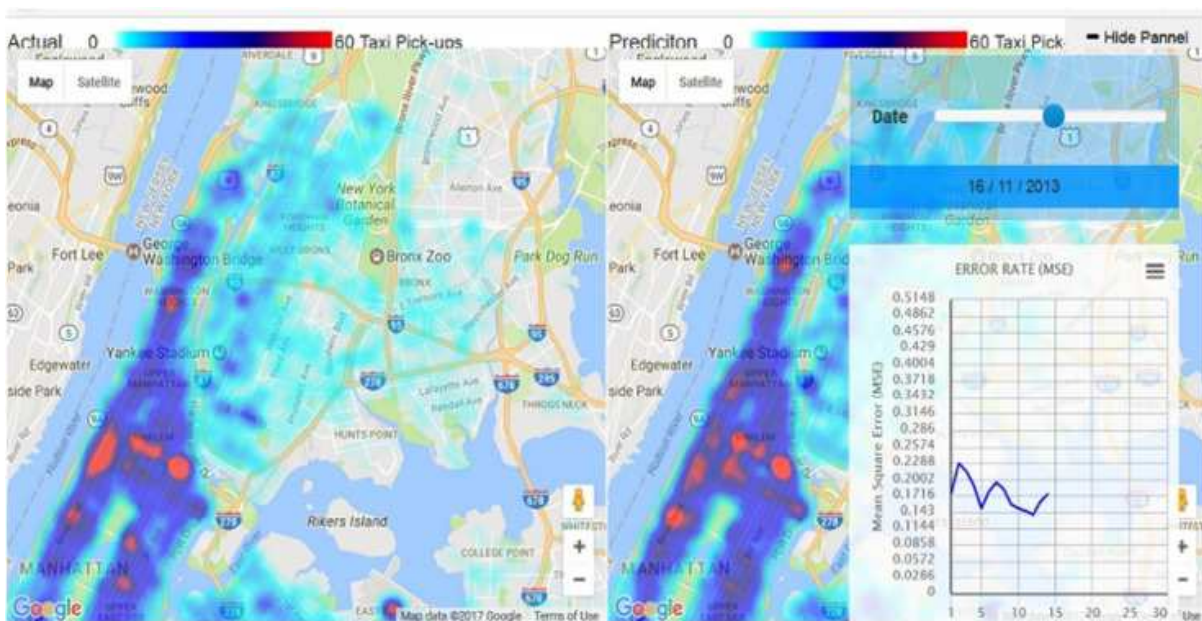Fig. 1. Data sources and sensory equipment for smart transport applications.

Fig. 2. Taxi demand forecasting application (actual pickups left and predicted pickups right).
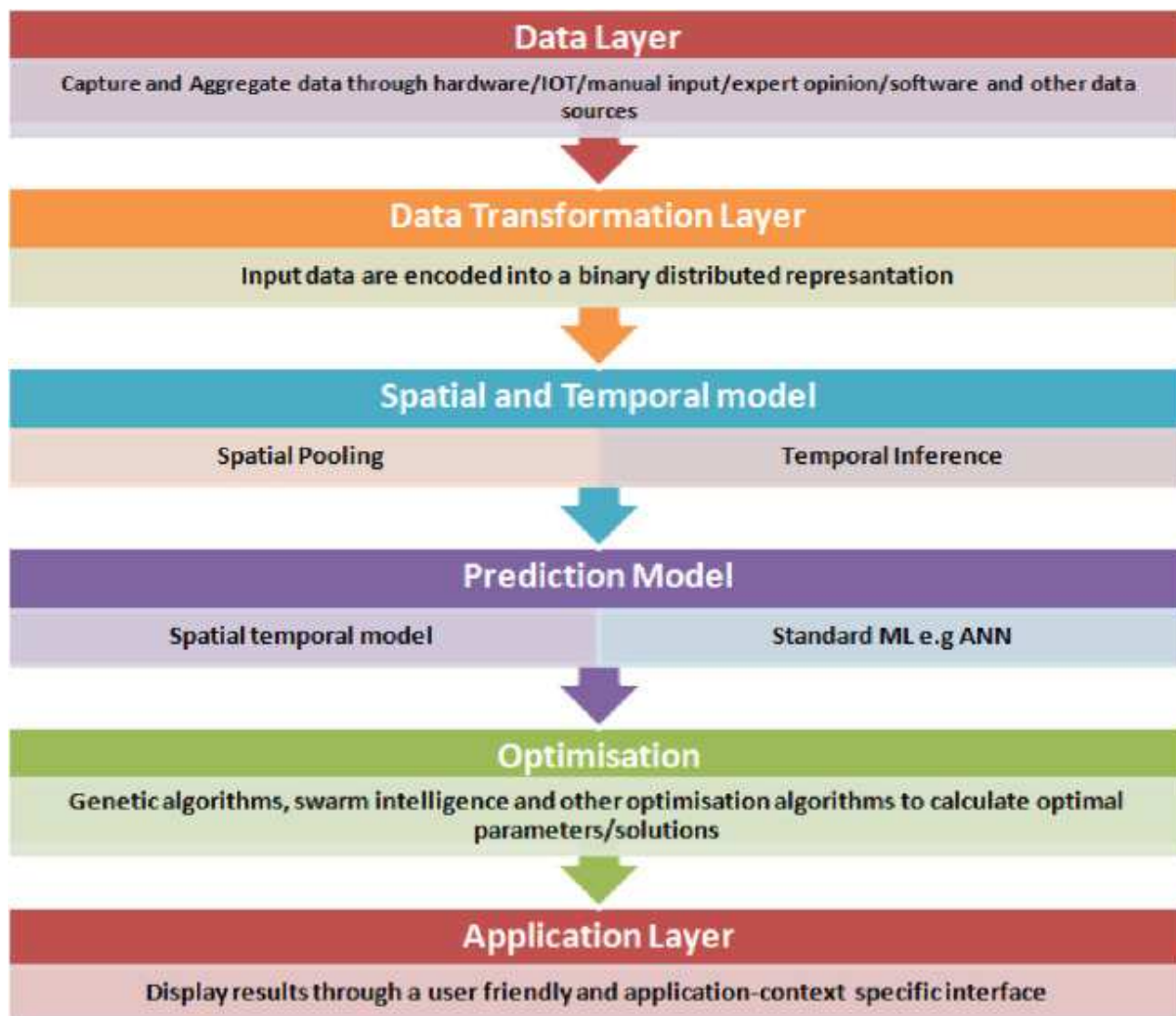
Fig. 3. Big data modelling methodology.

3.1.1. Smart city transportation case study-taxi demand prediction

In this section, we present a case study describing a smart city intelligent transportation system, focused on taxi demand forecasting. This system aims to optimize taxi fleet distribution in the context of a smart city, resulting in several benefits for a modern smart city, such as taxi-availability, reduced waiting and journey times for customers, better fuel management, environmental benefits and others. The authors utilized the publicly available dataset from the New York City Open Data webpage containing taxi trip data from 2013. The dataset included the following fields: pick-up and drop-off dates/times/locations, journey distances, fares, rate types, payment types, and driver-reported passenger counts.

The proposed system was trained to predict the locations on the map, where taxis picked up and set down customers in urban areas, over the course of a day. The system features user-friendly interface that enables the user to see the actual taxi demand, and the taxi demand predicted by the system, as well as the error of the prediction, calculated in terms of the Mean Square Error (MSE). The dataset utilized to train and test the taxi demand system ranged from January 2012 to November 2013. The data were separated to a training and testing datasets, in order to objectively evaluate the predictive model (December 2013 was excluded from the analysis due to additional

noise in the data, caused by the holiday season). The system utilized the methodology describe in Section 4 in order to produce the forecasting results, achieving an accuracy above 95% on the testing data. In terms of MSE, the system achieved approximately 0.18, meaning that on average if each point on the map is considered, the predictions could be wrong by 0.18 taxi pickups (we consider each taxi pick up location on the map to have an accuracy of up to 5 m).

The forecasting results produced by this system can be used as inputs by standard evolutionary algorithms and other optimization techniques to enable a city authority, or a taxi service provider to optimize their taxi fleet distribution based on the predicted demand and improve other aspects of taxi usage in a smart city such as minimizing CO_2 emissions.

3.2. Personalised health services for smart cities

In the context of a smart city, delivering effective digital health services is crucial for both governments and the public. Personalised health services can benefit patients and clinicians by providing optimized health information and recommendations, based on individual and population-based profiling, using advanced Big Data analytics algorithms. For example, in the very recent study by Chang et al., the researchers explored factors influencing the population group of older adults to adopt hearing aids in the context of a smart city. They encouraged this population group to utilize smart hospitals to consult audiologists about hearing loss and hearing aids rehabilitation, thus promoting their standard of living and ultimately their quality of life (Chang et al., 2018). Gaining a better insight into an individual's, or a group's healthcare needs is of great importance in order to provide tailored treatment and therapy intervention recommendations. Moreover, this extends to fitness, lifestyle 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. Monitoring and delivering these services though personal (Harding et al., 2015), cloud based (Sultan, 2014), m-Health (Free et al., 2013) and Internet of Things applications (Pang et al., 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 the changing care needs (Doctor et al., 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 utilizing contextual heterogeneous data, while handling uncertainties associated with incorporating qualitative data from stakeholders. Human interpretable care recommendation decisions can provide care staff the means of implementing adaptive care and therapy plans, in response to changing needs of patients due to the effects of cognitive decline.

Personalised health applications are very context and user specific, and they should be able to provide friendly human-machine interfaces allowing the user to visualize the results of Big Data analysis in order to draw useful conclusions and reveal hidden relations in the data. Therefore, tools and methodologies can be developed to provide personalisation and contextualisation for domain specific data visualization and interpretation, which include information retrieval, recommender systems and contextual information filtration through relevance feedback-based approaches. Context and task-oriented presentation of retrieved content can be based on understanding user information needs in the health and other application domains, for instance, optimal visualization of dynamic information, such as air traffic data and stock market information, to reduce cognitive burden on the user. Decision modelling will be used to intelligently tune and optimize recommendation of user and context relevant information. Context-aware visualization will help to reduce the complexity and cognitive burden on interpreting large volumes of data by providing relevant information to different stakeholders. Fuzzy Logic based techniques are able to provide this kind of frameworks. This can be seen in the recent work by Mahmud et al. where a data analytics and visualization framework for health shocks prediction was proposed, utilizing cloud computing services on Amazon, integrated with geographical information systems able to support Big Data requirements and visualization through smart devices. The researchers have also collected data from a large number of households in different regions of Pakistan and have developed a fuzzy predictive model, which comprised of interpretable fuzzy rules which provided stakeholders with a clear view of the factors influencing health shocks (Mahmud et al., 2016). Another framework for cloud based Big Data analytics and visualization was presented in Lu et al. (Lu et al., 2011). The researches have applied their framework to climate studies and have demonstrated that their framework was able to perform data analysis and visualization in competitive times. The proposed framework was able aid to the integration of data and tools and support interactive visualization on the fly thus allowing end users to perform analysis in a simple and flexible manner (Lu et al., 2011). 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 an important issue to address for successful commercial applications (Iqbal et al., 2011; Iqbal et al., 2013).

3.3. Biometrics and surveillance for smart cities

Smart cities around the world are getting more intelligent, but they also need to be designed with security in mind, thus providing a safe environment for their citizens. Biometrics security is related to the identification of humans based on their unique biometrical and behavioural traits. These include physical traits such as fingerprint, DNA, Iris recognition, hand geometry, face detection, and recognition, and behavioural traits such as voice, gait and typing rhythm, including digital signature (Chirillo and Blaul, 2003; Ngo et al., 2015). Recent applications of biometrics security measures are of critical importance due to growing security threats in public areas. Recent events such as the Paris attacks in 13/11/2015, and Brussels airport bombing in 22/3/2016, suggest that increasing measures need to be taken to understand and interpret intelligence data accumulated from various sources through the use of intelligent data analytics approaches. For example, social media, sensors, and surveillance cameras are embedded and used in the environment ubiquitously. Surveillance cameras are used to monitor traffic flow, threats, and suspicious activities in order to support government and law-enforcement agencies. They play a vital role in security issues, as they are widely deployed in public places to prevent or investigate crimes and to detect a suspicious person in order to make the environment more secure (Iqbal et al., 2015a, Iqbal et al., 2015b). Computational intelligence methods can help to classify physiological and behavioural traits to

effectively identity suspicious and abnormal behaviour for various applications such as vehicle theft, crowd surveillance through monitoring and analysis of multiple real-time data streams. Physiological information can also be used to improve safety and wellbeing of working population, involved in repetitive, long duration or late night working conditions, such as drivers involved in moving shipping containers around the clock. The drowsiness levels of drivers, for example, can be detected through feature classification algorithms able to detect the face, its orientation together with eyelid closure, blink rate, and gaze, which can be used to indicate drowsiness levels. The study conducted by Borghini et al. have demonstrated the close relation of various signal such as the Electroencephalography, Electrooculography and heart rate which can be captured by modern unobtrusive sensors with the transition of drivers from a normal driving state, to a state of mental fatigue and drowsiness. Their findings also supported that drowsiness is closely related to high blink rates and low heart rate values. Automatic detection of this state achieves 90% accuracy in the case of offline analysis (Borghini et al., 2014). However, it is important to mention that online detection requires robust techniques, able to handle reliably high velocity data from various input sources.

3.4. Business and government economic strategy recommendation for smart cities

Economic development of smart cities and modern societies as well as effective decision making by local and government authorities can be supported by intelligent business and government economic strategy recommendation systems. Big Data have an increasing significance in shaping the strategies and operations of businesses and recent research focused on providing guidelines in order to maximize and optimize this positive impact to modern organizations (Rehman et al., 2016). As it is highlighted in the work by Larson et al. Big Data has affected business intelligence and the utilization of information. In order for this Big Data to be utilized effectively new challenges arise concerning fast analytics and data science as part of business intelligence (Larson and Chang, 2016). The utilization of computational techniques in stock market and other financial factors can provide an insight to businesses in order for them to optimize their strategies and adjust their decision making to meet the demands of market conditions and the general economic environment. For instance, the shaping of effective e-marketing strategies enables e-businesses to generate more traffic and sales thus maximizing profit (Grzywaczewski et al., 2010). The combination of deep learning and fuzzy techniques are able to contribute in the processing of diverse data sources in order to result in effective business models which are able to account successfully for the inherent uncertainty related to market drivers such as consumer demand, government policy etc. Moreover, financial models can also be built in order to forecast and warn for potentially hazardous situations. In 2015, Nagaraj and Sridhar (2015) proposed a bankruptcy prediction system to categorize companies based on their extent of bankruptcy risk in order for it to be used as a decision support system. The team's predictive system is an SVM based approach, which predicts bankruptcy for a customer dataset. With the utilization of Big Data, which reflect the surrounding financial environment, and more advanced machine learning techniques, decision support systems with similar aims can be developed in order to facilitate shaping strategies for large enterprises or even governments in order to avoid potentially disastrous situations.

3.5. Population displacement models for smart cities

An application field related to the modern global socioeconomic norms is the utilization of Big Data analytics and machine learning techniques in order to understand, model and predict social phenomena such as migration flows, and population displacements. Recent research attempts for modelling population movements have been performed by Gulden et al. and Sokolowski et al. Sokolowski created an agent based model which can predict, measure and evaluate future

population movements in Syria. In their paper the researchers provided a methodology in order to craft the environment and agents, to represent the city of Aleppo in Syria. The team's model provided considerably accurate results concerning the displacement of population, which has happened in 2013 (Sokolowski et al., 2014). In the work by Gulden et al., a stylized model for representing internally displaced population dynamics in East Africa was presented. The researches discussed different frameworks for modelling this kind of phenomenon. Their models included a local, pruned, and hybrid interaction model (Gulden et al., 2011). Nowadays Europe is currently under a crisis due to massive migration and refugee flows due to people is fleeing their homelands in order to escape the horrors of war, or to avoid harsh economic conditions existing in a number of countries. This phenomenon has led to a large socio-economic crisis especially for reception countries, which found themselves utterly unprepared. Utilizing Big Data, and advanced Big Data analytics techniques, would allow for the development of computational models to represent this phenomenon. These models could shed light in the factors influencing this kind of phenomena, such as migration, and allow for accurate prediction concerning the evolution of the phenomenon and predicting future large migration flows. Towards achieving the goals described above, a variety of data falling under the category of Big Data can be utilized in order to construct effective models to represent this phenomenon. These models can be built based on data generated through the interactions of users with social networks, or popular searches related to specific events, or areas affected by natural or manmade disasters, or by utilizing publicly available databases and statistics available through international organization such as the UN. These models can be used to simulate population displacement phenomena, and as recommendation tools to aid in the design of strategies to address these issues.

3.6. Sentiment analysis for smart cities

Another interesting application field deriving from Big Data research is Sentiment analysis. Sentiment analysis aims at facilitating the analysis of high volume data in order to automatically identify the user's evaluations and feelings towards specific products and services (Pang and Lee, 2008). This process will lead to the designing of more effective and innovative commercial products, thus strengthening the economic environment of a smart city or a country. Modern Facebook and Twitter users generate huge amounts of data through their posts, comments and status updates. This 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. 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 et al., 2011). Recent examples in the field of sentiment analysis in e-learning environments can be found in the works by Ortigosa et al. (Ortigosa et al., 2014) and Martin et al. (Martin et al., 2012).

3.7. Affective computing for smart cities

Affective computing (AC) is "computing that relates to, arises from, or deliberately influences emotion". The aim of AC is the development of computer systems able to recognize the emotional

state of their users; systems which possess the ability to imitate human emotions; and ultimately virtual agents and systems which are able to feel their own emotions (Picard, 1995, Picard, 1997). AC delivers the promise of a higher level of human machine interaction by incorporating emotion in the design of intelligent computers systems. As it is stated in Calvo et al.'s work, AC struggles to bridge the gap of the emotional human with the emotionally challenged computer (Calvo and D'Mello, 2010). In order for an AC application to be effective it should be able to recognize the user's emotional state, model the relation of the user's affect with the surrounding environment and finally produce the necessary output signals in order to influence the user towards a more beneficial emotional state (Wu, 2010). In order to achieve this goal an AC application should use the necessary computational intelligence and emotion theory in order to produce accurate predictions concerning their user's affective state. However, most of AC applications utilize models which are developed and trained with limited data sources derived from small scale experiments. This is a big obstacle towards the successful design of such applications since the developed computational models fail to reflect sufficiently general populations consisting of people with diverse cultural or psychological profiles. Big Data and Big Data analytics utilizing deep learning techniques are able to overcome this limitation. People interact with other people and with their environment through smart devices and social networks and while they do so they produce emotions and are influenced by their emotions, thus producing vast numbers of personalised and contextualised affect related information. This data can be utilized to train intelligent systems by applying advanced deep learning techniques classify and predict the emotional and behavioural states of users. In the future, these Big Data sources can become the new large scale laboratories, where new emotion models can be proposed and their computational representations can be developed, tested and applied. A smart city is fertile ground for developing successful affective computing applications. These applications will be able to improve the quality of life of the population by taking into account a crucial factor, namely human emotion related to peoples activities and interaction with various smart city services.

3.8. Manufacturing and fault detection for smart cities

Fault detection in complex manufacturing domains may require a combination of CI techniques for modelling and classifying various types of faults (hardware/software) while contending with environmental uncertainties and using these models to further optimize production processes to reduce scrap production, cost, machine down-time and human effort. Developing robust and effective manufacturing lines for the industry sector, is a key factor in developing a functional and prosperous smart city. In an example of recent research in the area Maniac et al. has presented a deep learning neural network based approach which utilized unlabeled sound data and labeled historical data accumulated by experts, in order to automatically detect and recognize faulty audio signalling devices (Maniak et al., 2013; Maniak et al., 2015). The proposed system eliminated the need for manual inspection and t achieved high recognition accuracy. The system's success demonstrates the potential of applying similar machine learning approaches to data of various sources monitoring a production and manufacturing line such as CCTV etc. ML techniques and especially adaptive techniques, which are able to train from past observations are able to handle such data. This data as stated in the work of Monostori et al. are subject to large amounts of complexity and uncertainty, and the deployment of these kind of ML techniques can be very beneficial to the performance of modern manufacturing companies and enhance the effectiveness of conventional control and scheduling approaches (Monostori, 2003).

4. Proposed data modelling methodology

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 a number of soft computing techniques such as: deep belief networks, auto-encoders, agglomerative hierarchical clustering and temporal sequence processing. 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 and George, 2006).

The main elements of this method are as follows. Input data is initially encoded into what is termed as a sparse distributed representation (Hawkins and George, 2006). The initial stage of the approach performs hierarchical organization of multiple levels of data abstraction to identify correlations between temporal sequences of input patterns. The discovery of correlations between individual inputs 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 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. 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. The spatial pooling and temporal inference elements of the approach are combined 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. The prediction model can further provide inputs to optimisation framework or fuzzy decision model that is able to optimize processes based on uncertain inputs from various sources. 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 and faults in equipment and devices.

The (HSTSM) modelling approach can be integrated as part of our proposed data modelling methodology, which is illustrated in Fig. 1. Here we show the core part of the methodology comprising of the data input layer, data transformation layer, modelling layer, prediction layer, optimization and application layer. This forms a common framework, which can be applied in various application contexts as the ones identified in the previous section.

5. Policy, protection, valuation and commercialization

The possible miss-use for Big Data sets, data to power the Internet of things, personal data held and large corporate commercialization currently features as one of the biggest issues in global commerce. As outlined in this paper controlling the accessibility of Big Data is of critical importance to the global economy and countries and territories face a huge challenge to legislate access to these data sources given differences of opinion, ethics and laws governing the distribution and

security of information. Commerce thrives on the access and availability of meaningful data, whether this be for ethically sound medical purposes, or for the selling of goods and services, personal security or life style information. Hence irrespective of the purpose of the data, there is a need to be sure that such data is used in a responsible way.

Most data of general importance to the public is supplied in open source formats, governed by a myriad of software licenses, permitting its, access and use. Legal use of, infringement and enforcement of database rights and data sets is covered by country specific or international laws. Cases which are currently of international importance relate to the supply of personal data held by one corporation outside the jurisdiction of the requesting nation, where this data is of critical importance to the requesting nation's security i.e. the prevention of terrorist attacks. Specifically, an example of this can be when state sponsored security agencies such as the CIA force a corporation holding data in a European member state to release that data. Here an underlying question is where would this forced jurisdiction over an individual corporation's and another country's sovereign rights end? As a result, governance and legal systems have to make a decision on the collective national and international interest over the individual rights to keep confidential personal or commercially sensitive data. It will be some time before the courts in countries such as USA and Europe determine a set of universally acceptable laws and policies on information accessibility that holistically considers forced privacy and data protection infringement issues in a more open and inclusive manner. However, it not clear how countries like Russia, China and India would respond to the such proposals, which also have an implication on strategies for trying to combat potential government backed cyberwarfare initiatives.

The use of anonymized Big Data analytics in the health and safety regimes are not usually contested for ethical or social reasons, the issues tend to revolve around the consequences of ubiquitous access to personal data, and how this then affects the individual, for instance the use of DNA to diagnose illness is considered ethical however the use of the same data to increase life insurance premiums may not be. In the same way data on driver's style and ability and can be used for reducing the insurance payments for "safe" drivers whilst increasing the premiums for "erratic" drivers. The public should have more transparency the ability to influence governments and policy on how our data is used. New EU legislation on personal data usage has been strengthened, and presents the opportunity for the public to view what data is held on them by utility companies, and corporations, and how this is shared for other purposes. One observation on these policies is whether they are legally binding, and who abides by them. Taking the international bioethics guidelines as an example which covers use and ownership of IP for naturally occurring flora and fauna. These guidelines are generally ignored by countries and cooperation.

Big data sets and the analytics behind the manipulation of data is big business, and worth billions of dollars per annum to the holders of such data. Ultimately the debate on ethics, policies and law will reside with different nations and valuation will always be dictated by the price industry will pay for access to this data. As a general principle it is hoped that and anonymized datasets and databases can be made freely available to society to help drive the digital and smart city revolution while being sensitive to ethics, privacy and security. Advances in technologies that combine AI with distributed ledger systems may provide the answer for intelligently distributing and verifying access to both data and smart devices.

Cloud computing adds another level of IP and regulatory issues, especially when combined with Open Source Software [OSS]. If OSS is used by a data analytics company the licensing regime stems a spectrum of possible types from the permissive [FreeBSB, Apache] through hybrid [Mozilla, LGPL] to reciprocal [AGPL, GPL]. Software vendors needs to consider the scope of their liability and construct warranty and indemnity clauses to effectively cover conformation of specification free from viruses media free from defects, non-infringement of IP and the warranty period.

Cloud service providers have to consider how to work under multiple national laws, data protection and security, international transfer, data breach notification, compliance with international frameworks and data retention laws. These issues are as important to the owners of these services as to how their liability related to the technology they provide is legally protected.

6. Conclusions

In this paper, we explored Big Data and Big Data analytics potential to support the development of modern smart cities, and investigated the benefits arising from the utilization of computational techniques namely deep learning neural networks, evolutionary algorithms and fuzzy logic in data analytics. We identified and highlighted potential novel smart city applications arising from the vast amount of information offered by modern high-tech societies, and from the deployment of intelligent computational techniques, and discussed various aspects of policy, protection, valuation and commercialization issues related to Big Data.

Big data and CI have the potential to influence modern societies, businesses and the public. Deployment of these technologies in the context of a modern smart city, can produce tangible benefits such as: achieve considerable cost reductions, support decision making, and motivate the development of new products and services in a very broad spectrum of application areas. These novel technologies are able to facilitate the efforts of governments and local authorities and enhance economic development, ultimately leading to a considerable improvement of the quality of life for citizens. This fact is demonstrated by the taxi demand prediction case study discussed in this paper. The accurate classification results provided by the proposed system can be used by the relevant smart city authorities, to optimize the real-time distribution of taxi's based on the predicted demand, resulting in improving taxi availability, reducing waiting times and journey times, even achieving environmental benefits related to minimizing the CO_2 emissions of the taxi-fleet without compromising the quality of its provided services.

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 brains. It incorporates different soft computing techniques and it has the potential to deal with the large amounts of data generated in the context of a smart city, characterised by complex spatial-temporal correlations. This methodology can provide the computational base for developing intelligent Big Data solutions. By exploiting the capabilities of the methodology to process data in competitive times, reveal hidden patterns and provide accurate classification and forecasting results, these solutions will have a significant impact on knowledge, society, the economy, and individuals. Scientists and researchers can use this methodology to discover hidden knowledge in the data, governments can process their data to support real time decision making and provide improved services to their citizens, the economy can be supported by the development of personalised and contextualised products and services utilizing this methodology, and the public will benefit from the incorporation of intelligent technology in their everyday lives.

Future work will involve the utilization of the proposed methodology to different smart city application areas such as the ones identified in this review study (fault detection, emotion modelling, sentiment analysis, population displacements, data visualization, economic strategy recommendation, personalised health services, intelligent transportation and biometrics and surveillance) in order to create novel models and applications with significant commercial and scientific value.

References

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).

Amin, R., Kumar, N., Biswas, G.P., Iqbal, R., Chang, V., 2018. A light weight authentication protocol for IoT-enabled devices in distributed cloud computing environment. Futur. Gener. Comput. Syst. 78, 1005–1019.

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. In: Concurrency and Computation: Practice and Experience.

Bekkerman, R., Bilenko, M., Langford, J., 2011. Scaling Up Machine Learning: Parallel and Distributed Approaches. Cambridge University Press.

Bengio, Y., 2009. Learning deep architectures for AI. Found. Trends Mach. Learn., vol. 2 (1).

Bhattacharya, M., Islam, R., Abawajy, J., 2016. Evolutionary optimization: a big data perspective. J. Netw. Comput. Appl. 59, 416–426.

Bing, L., Chan, K.C., 2014, December. In: A Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data. Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing. IEEE Computer Society, pp. 652–657.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev. 44, 58–75.

Calvo, R.A., D'Mello, S., 2010. Affect detection: an interdisciplinary review of models, methods, and their applications. Affect. Comput. IEEE Trans. 1 (1), 18–37.

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. IEEE, pp. 1848–1855.

Chang, V., Wang, Y., Wills, G., 2018. Research investigations on the use or non-use of hearing aids in the smart cities. Technol. Forecast. Soc. Chang. http://dx.doi.org/10. 1016/j.techfore.2018.03.002. (in press).

Chen, X.W., Lin, X., 2014. Big data deep learning: challenges and perspectives. Access IEEE 2, 514–525.

Chirillo, J., Blaul, S., 2003. Implementing Biometric Security. Hungry Minds, Incorporated.

Chung, I.H., Sainath, T.N., Ramabhadran, B., Pichen, M., Gunnels, J., Austel, V., ... Kingsbury, B., 2014. Parallel deep neural network training for big data on blue gene/ q. In: High Performance Computing, Networking, Storage and Analysis, SC14: International Conference. IEEE, pp. 745–753.

Costa, D.G., Collotta, M., Pau, G., Duran-Faundez, C., 2017. A fuzzy-based approach for sensing, coding and transmission configuration of visual sensors in smart city applications. Sensors 17 (1), 93.

D'Asaro, F.A., Di Gangi, M.A., Perticone, V., Tabacchi, M.E., 2017. Computational intelligence and citizen communication in the Smart City. Informatik-Spektrum 40 (1), 25–34.

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. IEEE, pp. 1–6.

Doctor, F., Hagras, H., 2013. U.S. Patent No. 8,515,884. U.S. Patent and Trademark Office, Washington, DC.

Doctor, F., Iqbal, R., 2012. An intelligent framework for monitoring student performance using fuzzy rule-based linguistic summarisation. In: Abbass, H., Essam, D., Sarker, R. (Eds.), Fuzzy Systems (FUZZ-IEEE). IEEE International Conference on. 10–15 June 2012. pp. 1–8.

Doctor, F., Hagras, H., Callaghan, V., January 2005. A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments. IEEE Trans. Syst. Man Cybern. Syst. Hum. 35 (1), 55–56.

Doctor, F., Iqbal, R., Naguib, R.N., 2014. A fuzzy ambient intelligent agents approach for monitoring disease progression of dementia patients. J. Ambient. Intell. Humaniz. Comput. 5 (1), 147–158.

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. Appl. Soft Comput. 38, 872–889.

Dreier, F., 2002. Genetic Algorithm Tutorial.

Duggal, R., Khatri, S.K., Shukla, B., 2015. 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. IEEE, pp. 1–6.

Eaton, C., Deroos, D., Deutsch, T., Lapis, G., Zikopoulos, P.C., 2012. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Companies.

Esposito, E., De Vito, S., Salvato, M., Fattoruso, G., & Di Francia, G. (2017). Computational Intelligence for Smart Air Quality Monitors Calibration. In International Conference on Computational Science and its Applications (pp. 443–454). Springer, Cham. 978-0-07-179053-6, 2012.

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.

Fujdiak, R., Masek, P., Mlynek, P., Misurec, J., Olshannikova, E., 2016, July. Using genetic algorithm for advanced municipal waste collection in Smart City. In: Communication Systems, Networks and Digital Signal Processing (CSNDSP). 2016 10th International Symposium. IEEE, pp. 1–6.

Gerhardt, B., Griffin, K., Klemann, R., 2012. "Unlocking Value in the Fragmented World of Big Data Analytics", Cisco Internet Business Solutions Group. http://www.cisco.com/web/about/ac79/docs/sp/Information-Infomediaries.pdf.

Grzywaczewski, A., Iqbal, R., Shah, N., James, A., 2010. E-marketing strategy for businesses. In: e-Business Engineering (ICEBE). 2010 IEEE 7th International Conference. IEEE, pp. 428–434.

Gulden, T., Harrison, J.F., Crooks, A.T., 2011. In: Modeling Cities and Displacement through an Agent-based Spatial Interaction Model. Proceedings of the Second Annual Conference of the Computational Social Science Society of the Americas CSSSA-2011, Santa Fe, New Mexico.

Gupta, V., Kumar, R., Reddy, K.S., Panigrahi, B.K., 2017, July. In: Intelligent Traffic Light Control for Congestion Management for Smart City Development. IEEE Region 10 Symposium (TENSYMP), 2017. IEEE, pp. 1–5.

Habtie, A.B., Abraham, A., Midekso, D., 2016. A neural network model for road traffic flow estimation. In: Advances in Nature and Biologically Inspired Computing. Springer, Cham, pp. 305–314.

Harding, C., Ilves, P., Wilson, S., 2015. Digital mental health services in general practice. Br. J. Gen. Pract. 65 (631), 58–59.

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. Inf. Syst. 47, 98–115.

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. Int. J. Inf. Manag. 36 (5), 748–758.

Hawkins, J., 2004. On Intelligence. Times Books, New York. Hawkins, J., George, D., 2006. Hierarchical Temporal Memory: Concepts, Theory, and Terminology.

Haykin, S.S., Haykin, S.S., Haykin, S.S., Haykin, S.S., 2009. Neural Networks and Learning Machines. Vol. 3 Pearson Education, Upper Saddle River.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.

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 Process. Mag. IEEE 29 (6), 82–97.

Hirsch, D.D., 2013. Glass house effect: big data, the new oil, and the power of analogy. The. Me. L. Rev. 66, 373.

Iqbal, R., Shah, N., James, A., Duursma, J., 2011. ARREST: from work practices to redesign for usability. Expert Syst. Appl. 38 (2), 1182–1192.

Iqbal, R., Shah, N., James, A., Cichowicz, T., 2013. Integration, optimization and usability of enterprise applications. J. Netw. Comput. Appl. 36 (6), 1480–1488.

Iqbal, R., Grzywaczewski, A., Halloran, J., Doctor, F., Iqbal, K., 2015a. Design implications for task-specific search utilities for retrieval and reengineering of Code. In: Enterprise Information Systems. Taylor and Francis, pp. 1751–7575. http://dx.doi.org/10.1080/17517575.2015.1086494.

Iqbal, K., Odetayo, M., James, A., Iqbal, R., Kumar, N., Bama, S., 2015b. An efficient image retrieval scheme for colour enhancement of embedded and distributed surveillance images. J. Neurocomputing Vol. 174 (Part 2), 413–430.

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). IEEE, pp. 995–1004.

Karyotis, C., Doctor, F., Iqbal, R., James, A., 2015. In: An Intelligent Framework for Monitoring Students Affective Trajectories Using Adaptive Fuzzy Systems. IEEE International Conference on Fuzzy Systems, 2–5 August 2015, Istanbul, Turkey.

Karyotis, C., Doctor, F., Iqbal, R., James, A., Chang, V., 2018. A fuzzy computational model of emotion for cloud based sentiment analysis. In: Information Sciences. 433–434. Elsevier, pp. 448–463.

Kattan, A., Doctor, F., Arif, M., 2015. Two brains guided interactive evolution. In: Systems, Man, and Cybernetics (SMC). 2015 IEEE International Conference. IEEE, pp. 3203–3208.

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. Springer Berlin Heidelberg, pp. 169–176. http://www.landrover.com/experiences/news/self-driving-car.html As accessed online on 24 May 2016.

Larson, D., Chang, V., 2016. A review and future direction of agile, business intelligence, analytics and data science. Int. J. Inf. Manag. 36 (5), 700–710.

Le, Q.V., 2013. Building high-level features using large scale unsupervised learning. In: Acoustics, Speech and Signal Processing (ICASSP). 2013 IEEE International Conference. IEEE, pp. 8595–8598.

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. J. Supercomput. 1–12.

Liu, Y.X., Doctor, F., Fan, S.Z., Shieh, J.S., 2014. Performance analysis of extracted rulebase multivariable Type-2 self-organizing fuzzy logic controller applied to anesthesia. Biomed. Res. Int. 2014.

Lu, S., Li, R.M., Tjhi, W.C., Lee, K.K., Wang, L., Li, X., Ma, D., 2011. 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. IEEE, pp. 618–622.

Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.Y., 2015. Traffic flow prediction with big data: a deep learning approach. Intell. Trans. Syst. IEEE Trans. 16 (2), 865–873.

Lwowski, J., Kolar, P., Benavidez, P., Rad, P., Prevost, J.J., Jamshidi, M., 2017, June. In: Pedestrian Detection System for Smart Communities Using Deep Convolutional Neural Networks. System of Systems Engineering Conference (SoSE), 2017 12th. IEEE, pp. 1–6.

Mahmud, S., Iqbal, R., Doctor, F., 2016. Cloud enabled data analytics and visualization framework for health-shocks prediction. Futur. Gener. Comput. Syst. 65, 169–181.

Maniak, T., Iqbal, R., Doctor, F., Jayne, C., 2013. In: 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.

Maniak, T., Jayne, C., Iqbal, R., Doctor, F., 2015. Automated intelligent system for sound signalling device quality assurance. Inf. Sci. 294, 600–611.

Martin, J.M., Ortigosa, A., Carro, R.M., 2012. SentBuk: Sentiment analysis for e-learning environments. In: Computers in Education (SIIE). 2012 International Symposium. IEEE, pp. 1–6.

Melo, F.S., Silva, J.L.M., Macedo, H.T., 2016. Flood monitoring in smart cities based on fuzzy logic about urban open data. In: Telematics and Information Systems (EATIS). 2016 8th Euro American Conference. IEEE, pp. 1–5.

Mital, M., Chang, V., Choudhary, P., Papa, A., Pani, A.K., 2017. Adoption of internet of things in India: a test of competing models using a structured equation modeling approach. Technol. Forecast. Soc. Chang. http://dx.doi.org/10.1016/j.techfore.2017. 03.001. (in press).

Mitchell, T., 1997. Machine Learning, 1st ed. McGraw-Hill Education, New York.

Monostori, L., 2003. AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing. Eng. Appl. Artif. Intell. 16 (4), 277–291.

Murdoch, T.B., Detsky, A.S., 2013. The inevitable application of big data to health care. JAMA 309 (13), 1351–1352.

Nagaraj, K., Sridhar, A., 2015. A Predictive System for detection of Bankruptcy Using Machine Learning Techniques. (arXiv preprint arXiv:1502.03601).

Ngo, D.C.L., Teoh, A.B.J., Hu, J. (Eds.), 2015. Biometric Security. Cambridge Scholars Publishing, Eds.

Ortigosa, A., Martín, J.M., Carro, R.M., 2014. Sentiment analysis in Facebook and its application to e-learning. Comput. Hum. Behav. 31, 527–541. Pang, B., Lee, L., 2008. Opinion mining and sentiment analysis. Found. Trends Inf. Retr. 2 (1–2), 1–135.

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. Enterp. Inf. Syst. 9 (1), 86–116.

Parpinelli, R.S., Lopes, H.S., 2011. New inspirations in swarm intelligence: a survey. Int. J. Bio-Inspired Comput. 3 (1), 1–16.

Picard, R., 1995. Affective Computing Technical Report no. 321. MIT Media Laboratory, Cambridge.

Picard, R., 1997. Affective Computing. MIT Press, Cambridge, MA.

Poli, R., Langdon, W.B., McPhee, N.F., Koza, J.R., 2008. A Field Guide to Genetic Programming.

Qureshi, Faisal Fayyaz, Iqbal, Rahat, Qasim, Mohammad, Doctor, Faiyaz, Chang, Victor, 2017. Integration of OMNI channels and machine learning with smart technologies. J. Ambient. Intell. Humaniz. Comput. 1–17.

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. Springer International Publishing, pp. 119–142.

Rehman, Muhammad Habib ur, Chang, Victor, Batool, Aisha, Teh, Ying Wah, 2016. Big Data Reduction Framework for Value Creation in Sustainable Enterprises. In: International Journal of Information Management.

Sagiroglu, S., Sinanc, D., 2013, May. Big data: a review. In: Collaboration Technologies and Systems (CTS). 2013 International Conference. IEEE, pp. 42–47.

Sokolowski, J.A., Banks, C.M., Hayes, R.L., 2014. In: Modeling Population Displacement in the Syrian City of Aleppo. Proceedings of the 2014 Winter Simulation Conference. IEEE Press, pp. 252–263.

Sultan, N., 2014. Making use of cloud computing for healthcare provision: opportunities and challenges. Int. J. Inf. Manag. 34 (2), 177–184.

Suthaharan, S., 2014. Big data classification: problems and challenges in network intrusion prediction with machine learning. ACM Sigmetrics Perform. Eval. Rev. 41 (4), 70–73.

Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., Li, P., 2011. In: User-Level Sentiment Analysis Incorporating Social Networks. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 1397–1405.

Tirachini, A., 2013. Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services. Transp. Res. C Emerg. Technol. 30, 239–256.

Tolk, A., 2015, July. In: The Next Generation of Modeling & Simulation: Integrating Big Data and Deep Learning. Proceedings of the Conference on Summer Computer Simulation. Society for Computer Simulation International, pp. 1–8.

UK Smart Cities Index, 2016, May 17. Assessment of Strategy and Execution of the UK's Leading Smart Cities. Commissioned by Huawei, Navigant Consulting, Inc.

Vegni, A.M., Biagi, M., Cusani, R., 2013. Smart Vehicles, Technologies and Main Applications in Vehicular Ad Hoc Networks. INTECH Open Access Publisher.

Villars, R.L., Olofson, C.W., Eastwood, M., 2011. Big Data: What it Is and why you should Care. White Paper, IDC.

Vlahogianni, E.I., Kepaptsoglou, K., Tsetsos, V., Karlaftis, M.G., 2016. A real-time parking prediction system for smart cities. J. Intell. Transp. Syst. 20 (2), 192–204.

Wahid, A., Shah, M.A., Qureshi, F.F., Maryam, H., Iqbal, R., Chang, V., 2017. Big data analytics for mitigating broadcast storm in vehicular content centric networks. Futur. Gener. Comput. Syst (in press).

Whitley, D., 2001. An overview of evolutionary algorithms: practical issues and common pitfalls. Inf. Softw. Technol. 43 (14), 817–831.

Wu, D., 2010. Using physiological signals: virtual reality stroop task. IEEE Trans. Affect. Comput. 1 (2), 109–118.

Wu, D., 2012. 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.

Zadeh, L.A., 1965. Fuzzy sets. Inf. Control. 8 (3), 338–353.

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).