An Certain investigations of Emerging Big Data Technologies and its Applications

T.Mohana Priya[1] DrA.Saradha[2] R.Rajeshkanna[3]
[1]Research Scholar [2]Head [3]Assistant Professor
[1]Department of Information Technology [2]Department of Computer Science and Engineering
[3]Department of Computer Applications
[1]Dr. SNS. Rajalakshmi College of Arts and Science (Autonomous), Coimbatore, Tamilnadu [2]Institute of Road and Transport Technology, Erode, Tamilnadu, India [3]Dr.N.G.P Arts and Science College (Autonomous), Coimbatore, Tamilnadu, India

Abstract— In the digital era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. Furthermore, decision makers need to be able to gain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data. Such value can be provided using big data analytics, which is the application of advanced analytics techniques on big data. This paper aims to analyze some of the different analytics methods and tools which can be applied to big data, as well as the opportunities provided by the application of big data analytics in various decision domains.

Key words: Big data, Mining, Medical image processing, Analytics, Hadoop

I. INTRODUCTION

The 21st century is an era of big data involving all aspects of human life, including biology and medicine.With the fast increasing popularity of mobile smart devices, mobile crowd sensing has become a new paradigm of applications that enables the ubiquitous mobile devices with enhanced sensing capabilities, such as smartphones and wearable devices, to collect and to share local information towards a common goal. Most of the smart devices are equipped with a rich set of cheap and powerful sensors, including accelerometer, digital compass, GPS, microphone, and camera. These sensors can be utilized to monitor mobile users' surrounding environment and infer human activities and contexts. In recent years, a wide variety of applications have been developed to realize the potential of crowd sensing throughout everyday life, such as environmental, noise pollution assessment, road and traffic condition monitoring, road-side parking statistics, and indoor localization. The data acquired through mobile crowd sensing exhibits a number of important characteristics, such as being large in scale (volume), being fast generated (velocity), being different in forms (variety), and being uncertain in quality (veracity). The 4 Vs of crowd sensing data make it extremely interesting and challenging in designing participatory and opportunistic sensing technologies, human centric data management and analytics models, and novel visualization tools. Clustering is a widely used technique for big data analytics and mining. However, most of current algorithms are not effective to cluster heterogeneous data which is prevalent in big data. In this paper, we propose a high-order CFS algorithm (HOCFS) to cluster heterogeneous data by combining the CFS clustering algorithm and the dropout deep learning model, whose functionality rests on three pillars: (i) an adaptive dropout deep learning model to learn features from each type of data, (ii) a feature tensor model to capture the correlations of heterogeneous data, and (iii) a tensor distance-based high-order CFS algorithm to cluster heterogeneous data.

II. BIG DATA AND NETWORK BIOLOGY

Recently, biology has become a data intensive science because of huge datasets produced by high throughput molecular biological experiments in diverse areas including the fields of genomics, transcriptomics, proteomics, and metabolomics. In molecular biology, the list of components at the genome, transcriptome, proteome, and metabolome levels is gradually becoming complete and well-known to scientists. However, it is not holistically known how these components interact with each other to grow and maintain and reproduce life at different phases, in different environments, or with different challenging conditions.

Networks at the molecular level are constructed to understand and explain processes and subprocesses of the cell. New tools and algorithms are being continuously developed for the purpose of handling and mining big biological data and networks aiming to serve humanity by developing smart health care systems, new generation medical tests, drugs, foods, fuel, materials, sensors, and so on. Overall, this improves the understanding of the cell or in other words the life as a system. Therefore, the range of topics under big data and network biology is extensive and the present special issue is not a comprehensive representation of the subject. Nonetheless, the articles selected for this special issue represent versatile topics concerning the title that we have the pleasure of sharing with the readers.

The review paper "A Glimpse to Background and Characteristics of Major Molecular Biological Networks" focuses on biological background and topological properties of gene regulatory, transcriptional regulatory, protein-protein interaction, and metabolic and signaling networks. Versatile information contained in this article is helpful to facilitate a comprehensive understanding and to conceptualize the foundation of network biology.

The paper titled "METSP: A Maximum-Entropy Classifier Based Text Mining Tool for Transporter-Substrate Identification with Semistructured Text" discusses a method for identifying transporter-substrate pairs by text mining and applied it to human transporter annotation sentences collected from UniProt database. The substrates of a transporter are not only useful for inferring function of the transporter, but also important in discovery of compound-compound interactions and reconstruction of metabolic pathways.

Volatile organic compounds (VOCs) play an important role in chemical ecology specifically in the biological interactions between organisms and ecosystems. The paper titled "Development and Mining of a Volatile Organic Compound Database" discusses creation of a new VOC database by collecting information scattered in scientific literature and analyzed the accumulated data to show relations between biological functions and chemical structures of VOCs. This work also shows that VOC based classification of microorganisms is consistent with their classification based on pathogenicity.

When inconsistent policies are applied to hospital computer systems, it can produce enormous problems, such as stolen information, frequent failures, and loss of the entire or part of the hospital data. The paper "EMRlog Method for Computer Security for Electronic Medical Records with Logic and Data Mining" presents a new method named EMRlog for computer security systems in hospitals based on two kinds of policies, that is, directive and implemented policies.

The paper "Cellular Metabolic Network Analysis: Discovering Important Reactions in Treponema pallidum" focuses on identifying critical reactions by analyzing the topological structure of the metabolic network. Treponema pallidum is the syphilis-causing pathogen and the critical reactions of its metabolism are important drug targets and such information can lead to invention of effective vaccine of syphilis.

The paper "Systematic Analysis of the Associations between Adverse Drug Reactions and Pathways" focuses on identifying relations between adverse drug reactions (ADRs) and biological pathways by integrating clinical phenotypic data, biological pathway data, and drug-target relations. This work suggests that drug perturbation in a certain pathway can cause changes in multiple organs, rather than in a specific organ.

The paper "Discovering Distinct Functional Modules of Specific Cancer Types Using Protein-Protein Interaction Networks" proposes a new graph theory based method to identify distinct functional modules associated with specific cancer types. The method was applied to nine different cancer PPI networks. The distinct modules identified by this work have a high correlation with those found in the experimental datasets related to specific cancer types.

The paper "Shaped 3D Singular Spectrum Analysis for Quantifying Gene Expression, with Application to the Early Zebrafish Embryo" presents a new approach for studying gene expression in nuclei located in a thick layer around a spherical surface by integrating several methods such as depth equalization on the sphere, flattening, interpolation to a regular grid, pattern extraction by shaped 3D singular spectrum analysis (SSA), and interpolation back to original nuclear positions.

The paper "An Effective Big Data Supervised Imbalanced Classification Approach for Ortholog Detection in Related Yeast Species" proposes a supervised Pairwise Ortholog Detection (POD) approach by combining a set of gene pair features based on similarity measures, such as alignment scores, sequence length, gene membership to conserved regions, and physicochemical profiles. The performance of the proposed method has been compared with several existing methods in the context of three pairs of yeast genomes[1-4].

III. Military Simulation Big Data

Military simulation (MS) is another application domain producing massive datasets created by high-resolution models and large-scale simulations. It is used to study complicated problems such as weapon systems acquisition, combat analysis, and military training. MS big data can be produced by numerous simulation replications containing high-resolution models or large-scale battle spaces or both. The data size increases rapidly with larger simulation and higher performance computer resources. This poses significant challenges in management and processing of MS big data. First, great efforts have been made to address the requirements from high performance simulation, while only a few is toward data processing. However, the processing of MS big data makes some differences between business data; for example, the computing resources are often used for data analysis together with simulation execution and this may lead to more complex resource scheduling. Second, new requirements will emerge with availability of high fidelity models and timeliness of large amounts of data. Traditional data analytic methods are limited by traditional database technologies with regard to efficiency and scalability. Now big data based new applications allow military analysts and decision-makers to develop insights or understand the landscape for complex military scenarios instead of being limited to only examining experimental results. An example is supporting the real-time planning of real combat by simulating all kinds of possibilities. This requires iterative simulation and analysis of result in very short time[5].

A. *Big Data Analytics in Healthcare*

The rapidly expanding field of big data analytics has started to play a pivotal role in the evolution of healthcare practices and research. It has provided tools to accumulate, manage, analyze, and assimilate large volumes of disparate, structured, and unstructured data produced by current healthcare systems. Big data analytics has been recently applied towards aiding the process of care delivery and disease exploration. However, the adoption rate and research development in this space is still hindered by some fundamental problems inherent within the big data paradigm[6].

Currently healthcare systems use numerous disparate and continuous monitoring devices that utilize singular physiological waveform data or discredited vital information to provide alert mechanisms in case of overt events. However, such uncompounded approaches towards development and implementation of alarm systems tend to be unreliable and their sheer numbers could cause "alarm fatigue" for both care givers and patients.

B. Medical Image Processing from Big Data Point of View

Medical imaging provides important information on anatomy and organ function in addition to detecting diseases states. Moreover, it is utilized for organ delineation, identifying tumors in lungs, spinal deformity diagnosis, artery stenosis detection, aneurysm detection, and so forth. In these applications, image processing techniques such as enhancement, segmentation, and denoising in addition to machine learning methods are employed. As the size and dimensionality of data increase, understanding the dependencies among the data and designing efficient, accurate, and computationally effective methods demand new computer-aided techniques and platforms. The rapid growth in the number of healthcare organizations as well as the number of patients has resulted in the greater use of computer-aided medical diagnostics and decision support systems in clinical settings[7]. Many areas in health care such as diagnosis, prognosis, and screening can be improved by utilizing computational intelligence. The integration of computer analysis with appropriate care has potential to help clinicians improve diagnostic accuracy. The integration of medical images with other types of electronic health record (EHR) data and genomic data can also improve the accuracy and reduce the time taken for a diagnosis.

C. Medical Signal Analytics

Telemetry and physiological signal monitoring devices are ubiquitous. However, continuous data generated from these monitors have not been typically stored for more than a brief period of time, thereby neglecting extensive investigation into generated data. However, in the recent past, there has been an increase in the attempts towards utilizing telemetry and continuous physiological time series monitoring to improve patient care and management.

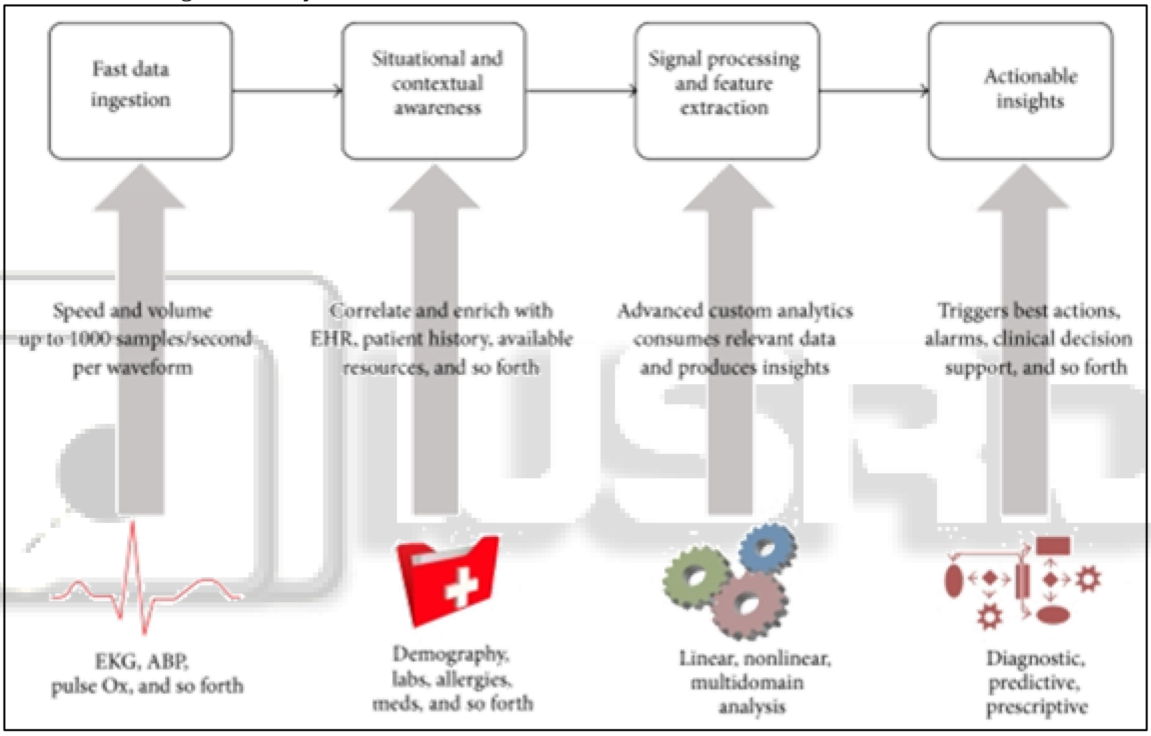

Fig. 1: Generalized analytic workflow using streaming healthcare data.

Streaming data analytics in healthcare can be defined as a systematic use of continuous waveform (signal varying against time) and related medical record information developed through applied analytical disciplines (e.g., statistical, quantitative, contextual, cognitive, and predictive) to drive decision making for patient care. The analytics workflow of real-time streaming waveforms in clinical settings can be broadly described using Figure 1. Firstly, a platform for streaming data acquisition and ingestion is required which has the bandwidth to handle multiple waveforms at different fidelities. Integrating these dynamic waveform data with static data from the EHR is a key component to provide situational and contextual awareness for the analytics engine. Enriching the data consumed by analytics not only makes the system more robust, but also helps balance the sensitivity and specificity of the predictive analytics. The specifics of the signal processing will largely depend on the type of disease cohort under investigation. A variety of signal processing mechanisms can be utilized to extract a multitude of target features which are then consumed by a pretrained machine learning model to produce an actionable insight. These actionable insights could either be diagnostic, predictive, or prescriptive. These insights could further be designed to trigger other mechanisms such as alarms and notification to physicians[8-16].

D. Signal Analytics Using Big Data

Research in signal processing for developing big data based clinical decision support systems (CDSSs) is getting more prevalent. In fact organizations such as the Institution of Medicine have long advocated use of health information technology including CDSS to improve care quality. CDSSs provide medical practitioners with knowledge and patient-specific information, intelligently filtered and presented at appropriate times, to improve the delivery of care.

As complex physiological monitoring devices are getting smaller, cheaper, and more portable, personal monitoring devices are being used outside of clinical environments by both patients and enthusiasts alike.

However, similar to clinical applications, combining information simultaneously collected from multiple portable devices can become challenging. Pantelopoulos and Bourbakis discussed the research and development of wearable biosensor systems and identified the advantages and shortcomings in this area of study [125]. Similarly, portable and connected electrocardiogram, blood pressure and body weight devices are used to set up a network based study of telemedicine.

IV. TECHNOLOGIES

A. Column-Oriented Databases

Traditional, row-oriented databases are excellent for online transaction processing with high update speeds, but they fall short on query performance as the data volumes grow and as data becomes more unstructured. Column-oriented databases store data with a focus on columns, instead of rows, allowing for huge data compression and very fast query times. The downside to these databases is that they will generally only allow batch updates, having a much slower update time than traditional models.

B. Schema-Less Databases, Or Nosql Databases

There are several database types that fit into this category, such as key-value stores and document stores, which focus on the storage and retrieval of large volumes of unstructured, semi-structured, or even structured data. They achieve performance gains by doing away with some (or all) of the restrictions traditionally associated with conventional databases, such as read-write consistency, in exchange for scalability and distributed processing.

C. Map Reduce

This is a programming paradigm that allows for massive job execution scalability against thousands of servers or clusters of servers. Any MapReduce implementation consists of two tasks:

1) The "Map" task, where an input dataset is converted into a different set of key/value pairs, or tuples;
2) The "Reduce" task, where several of the outputs of the "Map" task are combined to form a reduced set of tuples (hence the name).

D. Hadoop

Hadoop is by far the most popular implementation of MapReduce, being an entirely open source platform for handling Big Data. It is flexible enough to be able to work with multiple data sources, either aggregating multiple sources of data in order to do large scale processing, or even reading data from a database in order to run processor-intensive machine learning jobs. It has several different applications, but one of the top use cases is for large volumes of constantly changing data, such as location-based data from weather or traffic sensors, web-based or social media data, or machine-to-machine transactional data.

E. Hive

Hive is a "SQL-like" bridge that allows conventional BI applications to run queries against a Hadoop cluster. It was developed originally by Facebook, but has been made open source for some time now, and it's a higher-level abstraction of the Hadoop framework that allows anyone to make queries against data stored in a Hadoop cluster just as if they were manipulating a conventional data store. It amplifies the reach of Hadoop, making it more familiar for BI users.

F. PIG

PIG is another bridge that tries to bring Hadoop closer to the realities of developers and business users, similar to Hive. Unlike Hive, however, PIG consists of a "Perl-like" language that allows for query execution over data stored on a Hadoop cluster, instead of a "SQL-like" language. PIG was developed by Yahoo!, and, just like Hive, has also been made fully open source.

G. Wibi Data

WibiData is a combination of web analytics with Hadoop, being built on top of HBase, which is itself a database layer on top of Hadoop. It allows web sites to better explore and work with their user data, enabling real-time responses to user behavior, such as serving personalized content, recommendations and decisions.

H. PLATFORA

Perhaps the greatest limitation of Hadoop is that it is a very low-level implementation of MapReduce, requiring extensive developer knowledge to operate. Between preparing, testing and running jobs, a full cycle can take hours, eliminating the interactivity that users enjoyed with conventional databases. PLATFORA is a platform that turns user's queries into Hadoop jobs automatically, thus creating an abstraction layer that anyone can exploit to simplify and organize datasets stored in Hadoop.

I. Storage Technologies

As the data volumes grow, so does the need for efficient and effective storage techniques. The main evolutions in this space are related to data compression and storage virtualization.

J. Sky Tree

SkyTree is a high-performance machine learning and data analytics platform focused specifically on handling Big Data. Machine learning, in turn, is an essential part of Big Data, since the massive data volumes make manual exploration, or even conventional automated exploration methods unfeasible or too expensive.

V. APPLICATIONS

A. Applications Enabled by Smartphones

Today's smartphones serve not only as important communication devices, but also as computing and sensing devices with rich sets of embedded sensors, such as accelerometers, digital compasses, gyroscopes, GPS, microphones, and cameras. Generally, combining growing computing abilities, these sensors are enabling new applications across a wide variety of domains, such as human health care, social networks, safety, environmental or climate monitoring, and transportation. They lead to a new research area called mobile phone sensing. As the number of smartphone users increases rapidly across the whole world, large amount of data is generated, transferred, aggregated, and analyzed. The ubiquity of mobile phones and the increasing size of the data generated by sensors and applications lead to a new research domain across computing and social science. Big data, as a data science to process high volume information, is consequently involved in this field. Researchers have begun to address big data issues by using large-scale mobile data as an input to characterize and understand real-life phenomena, including individual traits, human mobility, communication, and interaction patterns.

B. Smartphones for Internet of Things

Semantic-oriented vision, as one of the broader visions of Internet of Things (IoT), emphasizes on data integration and management from vast number of smart devices, such as smartphones, pads, sensor nodes, and other devices with the ability to send out information. As one of the most important constituent parts of IoT, smartphones can not only provide more information than other devices, but also act as information collecting and distributing terminals. How to integrate diverse information is a big challenge of utilizing smartphones for IoT. Mobile data collected from wireless sensor networks are strongly spatial correlated; however, traditional methods are usually in static setting and the so-called optimal data collection trees are fixed and their performance suffers from link problems when mobile users change virtual sinks. The model proposed in this paper initializes an optimized tree and updates it according to users' accessing virtual sinks by locally modifying the previously constructed data collection tree. Their model is easy to implement, has low cost, and provides real-time data acquirement even when updating the tree structure.

C. Smartphones for Crowd Sensing

Static sensing is traditional and mature but has node coverage, maintenance, and scalability issues. Mobile crowd sensing is more flexible, manageable, and scalable, especially when vast numbers of smartphones are used as sensing nodes in cities or towns. The fast increasing number of smartphone users, various inherent mobile applications, and exponential increasing capacity of 3G/4G networks lead to this new mobile sensing paradigm. Currently, smartphones are used as sensors for localization, personal/surrounding context recognition, traffic monitoring, and other daily life related applications. But, in the near future, other applications, such as environmental pollution detection, health care monitoring, and social life analysis, will generate large amount of sensing data. Unlike conventional sensor networks, mobile crowd sensing is more human related; therefore privacy and security should be carefully considered. Otherwise, smartphone users will be unwilling to share their devices and subsequent data with others. To the best of our knowledge, there is no mature platform for mobile crowd sensing and researchers are working in that direction. For example, researchers proposed Medusa, which can provide high-level abstractions for stages in completing crowd sensing tasks and a distributed system which can coordinate the execution of these tasks between smartphones and the cloud.

D. Smartphone's for Environment Monitoring

Weather and environment monitoring are usually the responsibility of governments and some specific institutions. But if billions of mobile phones can be utilized for such jobs, more diversified and abundant information can be used to improve human's living conditions. Currently, combined with a cloud of supporting web services, large amount of smart mobile devices make such a distributed data collection infrastructure possible, though not immediately usable. An appropriate platform can be used in this field for further applications.The Personal Environmental Impact Report (PEIR), a system that combines web and personal mobile techniques to inform users of environmental impact and exposure, which can help people make more informed and responsible decisions. PEIR is built on location tracing and GPS records that are sampled. Based on the GPS information, users' trips are predicted and environmental impact or exposure measurements are aggregated from each trip. This platform can be used for a number of applications, such as traffic condition measurement, environmental pollution monitoring, and vehicle emission estimating. Though only four applications were proposed by the authors, new models can be developed based on this platform and scalability, stability, performance, and usability are the foreseeable promising directions for this kind of platforms.

E. Common Issue of Smartphone Related Applications

In previous sections, we introduced different applications enabled by smartphones. One common research issue among the wide variety of applications that use smartphones as sensing data sources is power consumption. With the development of smartphones, more and more embedded devices and powerful processors are attached. Therefore, smartphones consume significantly more energy than the previous generation of cellular phones. A smartphone which never stops using its GPS, not to mention those applications which might combine GPS with other components, may run out of energy within several hours. So, for every newly developed application, power consumption is an unavoidable problem[16-20].

VI. CONCLUSION

Big data, just as its name implies, is a data science which cannot be easily processed using existing infrastructure or data processing methods. Currently, researchers are working in two directions to solve this problem. One is modifying and improving current infrastructures, for instance, strengthening processing abilities or optimizing computing structures, to handle data more efficiently. Another direction is developing new data management methods. Various techniques are applied in each direction and it is hard to categorize them precisely. One of the significant features of sensing in future is "gigantism." Concepts like smart cities and IoT require vast number of sensors to work together under certain control policies. Conventional topologies, policies, architectures, and methods are no longer suitable. Platforms which can deal at city level, country level, or even world level with sensor data are in need. In future, everything goes beyond big data technologies.

Big data analytics which leverages legions of disparate, structured, and unstructured data sources is going to play a vital role in how healthcare is practiced in the future. One can already see a spectrum of analytics being utilized, aiding in the decision making and performance of healthcare personnel and patients. Here we focused on three areas of interest: medical image analysis, physiological signal processing, and genomic data processing. The exponential growth of the volume of medical images forces computational scientists to come up with innovative solutions to process this large volume of data in tractable timescales. The trend of adoption of computational systems for physiological signal processing from both research and practicing medical professionals is growing steadily with the development of some very imaginative and incredible systems that help save lives.

REFERENCES

[1] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. J. Patil, and D. Barton, "Big data: the management revolution," Harvard Business Review, vol. 90, no. 10, pp. 60–68, 2012. View at Google Scholar
[2] C. Lynch, "Big data: how do your data grow?" Nature, vol. 455, no. 7209, pp. 28–29, 2008. View at Publisher • View at Google Scholar • View at Scopus
[3] A. Jacobs, "The pathologies of big data," Communications of the ACM, vol. 52, no. 8, pp. 36–44, 2009. View at Publisher • View at Google Scholar • View at Scopus
[4] P. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw-Hill Osborne Media, 2011.
[5] J. Manyika, M. Chui, B. Brown et al., Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, 2011.
[6] J. J. Borckardt, M. R. Nash, M. D. Murphy, M. Moore, D. Shaw, and P. O'Neil, "Clinical practice as natural laboratory for psychotherapy research: a guide to case-based time-series analysis," The American Psychologist, vol. 63, no. 2, pp. 77–95, 2008. View at Publisher • View at Google Scholar • View at Scopus
[7] L. A. Celi, R. G. Mark, D. J. Stone, and R. A. Montgomery, "'Big data' in the intensive care unit: closing the data loop," American Journal of Respiratory and Critical Care Medicine, vol. 187, no. 11, pp. 1157–1160, 2013. View at Publisher • View at Google Scholar • View at Scopus
[8] F. Ritter, T. Boskamp, A. Homeyer et al., "Medical image analysis," IEEE Pulse, vol. 2, no. 6, pp. 60–70, 2011. View at Publisher • View at Google Scholar
[9] J. A. Seibert, "Modalities and data acquisition," in Practical Imaging Informatics, pp. 49–66, Springer, New York, NY, USA, 2010.
[10] V. Marx, "Biology: the big challenges of big data," Nature, vol. 498, no. 7453, pp. 255–260, 2013. View at Publisher • View at Google Scholar • View at Scopus
[11] M. L. Metzker, "Sequencing technologies—the next generation," Nature Reviews Genetics, vol. 11, no. 1, pp. 31–46, 2010. View at Publisher • View at Google Scholar • View at Scopus
[12] D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson, and E. R. Mardis, "The next-generation sequencing revolution and its impact on genomics," Cell, vol. 155, no. 1, pp. 27–38, 2013. View at Publisher • View at Google Scholar • View at Scopus
[13] M. Watson, "Illuminating the future of DNA sequencing," Genome Biology, vol. 15, article 108, 2014. View at Publisher • View at Google Scholar
[14] R. Chen, G. I. Mias, J. Li-Pook-Than et al., "Personal omics profiling reveals dynamic molecular and medical phenotypes," Cell, vol. 148, no. 6, pp. 1293–1307, 2012. View at Publisher • View at Google Scholar • View at Scopus C. Lynch, "Big data: how do your data grow?" Nature, vol. 455, no. 7209, pp. 28–29, 2008.
[15] Toscos T., Chen M. Y., Activity sensing in the wild: a field trial of ubifit gardenProceedings of the SIGCHI Conference on Human Factors in Computing Systems April 200817971806doi:10.1145/1357054.13573352-s2.0-57649188943
[16] Antonic A., Zarko I. P., Jakobovic D.Inferring presence status on smartphones: the big data perspective Proceedings of the 18th IEEE Symposium on Computers and Communications (ISCC ′13)July 2013600605doi:10.1109/iscc.2013.67550132-s2.0-84897460888
[17] Liu B., Jiang Y., Sha F., Govindan R.Cloud-enabled privacy-preserving collaborative learning for mobile sensingProceedings of the 10th ACM Conference on Embedded Networked Sensor Systems (SenSys ′12) November 20125770doi:10.1145/2426656.24266632-s2.0-84873449020
[18] Weiss G. M., Lockhart J. W.Identifying user traits by mining smart phone accelerometer dataProceedings of the 5th International Workshop on Knowledge Discovery from Sensor Data (SensorKDD ′11)August 2011ACM6169doi:10.1145/2003653.20036602-s2.0-80051680184
[19] D. W. Hosmer Jr. and S. Lemeshow, Applied Logistic Regression, John Wiley & Sons, 2004.
[20] R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, vol. 3, Wiley, New York, NY, USA, 1973.