Knowledge-driven Analytics and Sensor Signal Processing in Human-centric Applications

Arijit Ukil
Tata Consultancy Services
Kolkata, India
arijit.ukil@tcs.com.com

Leandro Marin
University of Murcia
Spain
leandro@um.es

Antonio Jara
University of Applied Sciences Western Switzerland (HES-SO)
Switzerland
jara@ieee.org

John Farserotu
CSEM, Switzerland
john.farserotu@csem.ch

Abstract

Technology disruption through knowledge driven intelligent systems is increasingly controlling human life. Management of the present and future knowledge-driven artificial intelligence- based technologies is of highest importance to maximize its progressive influence to human life and human society. Life style diseases, social network affinity, impulsive financial decision, technology-abuse negatively affect our physical, emotional, social and mental health. Conversely, intelligent systems can bring positive impact on human life. This paper brings forward those positive applications and technologies as well as the path towards transformation of intelligent systems through some exemplary analysis that minimizes the negative impact. The push is to promote the development of human-centric intelligent technologies like precise and personalized medication and treatment plan, drug discovery of untreatable diseases, improved elderly care, minimizing private data theft, big data analytics for prediction of macro or micro economic condition, effective and fair trading practices, retail decision management, knowledge-driven energy and resource management, deep learning and artificial intelligence based applications for risk prediction and augmented human capability generation. The main focus of this paper is to demonstrate the knowledge-driven technologies, developments, applications for ensuring improvement of human quality of life. The impact would be micro-level, where human life is impacted in daily basis and at macro-level where human life would be impacted in long term that eventually influences the betterment to human society.

1. Introduction

This paper is intended to demonstrate the capability of knowledge-driven analytics for building human-centric applications. We envisage the knowledge-driven human beings, knowledge-driven societies and knowledge-driven technologies should co-operatively co-exist to create a better knowledge-driven world. Our focus is to minimize the risks, conflicts and hazards of adapting to intelligent systems. This paper illustrates exemplary impactful ideas and proposals to achieve the goal of knowledge-driven life.

Technology advancements of last few years have produced number of exquisite applications and penetrating influences in human life. The ubiquity of smartphones, large scale deployment of Internet of Things, high end computing, big data, impactful and gross engagement to social networks along with the advent and promise of powerful artificial intelligent tools like deep learning algorithms result in abundance of information generation, dissemination of knowledge and analytics- driven human decisions and choices. Such conglomeration of technologies, applications and the big data resources paves ways for knowledge-driven human life, society and economy.

We bring forward the applications and technologies that through knowledge-driven analytics bring positive outcomes to the human life and to the world at large. For example, knowledge-managed learning techniques have the capability of providing robust prediction of medical condition, automated summarization, report generation, minimization of diagnosis error, enabling remote disease screening. It can predict the suicidal trend or state of depression from analyzing Facebook posts, tweets or recent posted images. Prediction of psychiatric disorders like schizophrenia, which physicians find difficult to anticipate would have immense impact on millions of human life. Traditional coarse evidence driven medical treatment needs to be more precise and personalized. Big data and availability of vast information invite severe data privacy attacks which can potentially ruin one's life and reputation. One of the challenging applications is the controlled release of private data without compromising the beneficial influence, prediction and subsequent prevention of cyber-attacks and privacy breach incidents. Knowledge-driven analytics will restrict an individual to venture into risky investments, traps of false social requests.

The goal of this paper is to inculcate the realization of long term co-existence of human-life with big data, artificial intelligence and deep analytics. Powerful tools, applications and ever increasing knowledge sources will drive human life, its micro and macro conditions for augmenting the human capabilities, minimizing the nuisances of infiltratory technologies and overall betterment of human experiences.

2 Knowledge Management for Human Quality of Life

We are at the crucial juncture of welcoming the knowledge-driven management of our life with the apparent arrival of inflection point of big data analytics based industry solutions and research outcomes. Knowledge-driven technologies and applications for improving human quality of life will potentially enable long term human-centric convergence of futuristic applications.

It is assumed that knowledge-driven analytics, information management will attempt to ensure positive influence for society and quality of life. Broadly, the areas would be: managing and analysis of knowledge for human mental and physical health condition improvement, maximizing the benefits of social network interactions while minimizing the ill-effects, assisting human decision making in financial domain, social network foot-printing, behavioral understanding and subsequent necessary action recommendation, ensuring personal data privacy preservation, as well as attempting to address few pertinent questions: Can machines understand how are we feeling and act accordingly? How will I be alerted before a devastating financial decision? How can a doctor be given augmented knowledge on diagnosis? All of us are different. Why are we not given personalized treatment instead of average case treatment plan? How can we use big data and knowledge mining for developing sustainable societies by optimizing energy, waste and perishable resource management? And many others.

The pertinent areas of human quality of life improvement through intelligent knowledge management would be:

- Macro-action analytics to identify cognitive dissonance.
- Computational method of automated disease detection.
- Social network usage analytics to identify suicidal tendency and psychiatric abnormality.
- Finding efficacy of prescription drugs in the presence of concept drift.
- Identifying wrong or ineffective economic decisions based on spent and requirement analysis.
- Recommendation of personalized retail and financial decisions and plans.
- Big data management by proactive control of data misuse and incorporating proactive data privacy.
- Value alignment to highly automated intelligence systems to restrict greedy outcomes.
- Algorithmic fair trading.
- Deeper personalization by understanding the retail behavior, prognosis trend, sentiment analysis, drug abuse, online surfing habits and other related personal studies.
- Patient-specific tailored medication and treatment plan.
- Virtual assistant for elderly and infant care.
- Knowledge-driven energy, waste, perishable resource management.
- Artificial intelligence for changing the responsibilities of human workers, where mundane, repetitive, stressful jobs would be by robots or other humanoids.
- Game theoretic investigation for conflict resolution of actions in knowledge-driven intelligent system.
- Long term prediction on knowledge driven human life and society.

- Crowd sourcing for knowledge aggregation and exploiting wisdom of the crowd.

In this paper, we illustrate two important case studies:

1. Analytics for unobtrusive cardiac condition identification and inference: ways to minimize loss of human life due to cardiac diseases.
2. Privacy preserving sensor signal mining: ways to minimize human value loss due to intended and unintended privacy breaching attempts.

3 Analytics For Unobtrusive Cardiac Condition Identification And Inference: Ways To Minimize Loss Of Human Life Due To Cardiac Diseases

It is estimated that more than 25% of worldwide deaths are due to cardiac ailments. Fortunately, cardiac diseases are preventable when early signs of cardiac health abnormality systems are captured.

With the advent of sophisticated body sensors, smartphones and Internet-of-things (IoT), we can affordably capture various fundamental physiological signals, which are definite markers of cardiac health [Fras14]. For example, photoplethysmogram (PPG) can be reliably captured by smartphones, electrocardiogram (ECG) can be reliably captured by external sensors like Alivecor [Alive]. AliveCor has developed Kardia heart monitor that has prediction capability of fatal cardiac condition like Atrial Fibrillations [Heart]. In their investigation by concerned team, total 1001 persons in vulnerable age group of cardiac diseases (65 years and more) are studied and disease detection prediction of Kardia outperforms the doctor's capability [Jul]. It is well-known that prognosis is significantly better when Atrial Fibrillations is detected early and treated with appropriate anticoagulation. Such proactive diagnosis will have high probability of decreasing stroke morbidity and mortality. We observe that the entire study and analysis were performed ion smartphones, which encourage the ubiquity of deployment and building a penetrative eco-systems of cardiac disease monitoring.

In a further study, researchers attempted to predict the presence of Atrial Fibrillations and other cardiac diseases including arrhythmia, coronary artery diseases using single lead Alivecor ECG sensor attached with a smartphone [Ukil17A].

Formally the analytics problem to solve the disease prediction can be formulated as:

Let instance space be $\mathcal{X}$, label space be $\mathcal{S} = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \ldots, \mathcal{D}_N\}$, where $\mathcal{D}_{n=1,2,3,\ldots N}$ are the different diseases (for e.g. $\mathcal{D}_1$ be Atrial Fibrillation, $\mathcal{D}_2$ be Coronary Artery Disease, $\mathcal{D}_3$ be the normal sinus rhythm) and prediction space be $\hat{\mathcal{S}}$ and our model be $\mathcal{M}$, such that:

$$\mathcal{M}(\mathcal{X}, \mathcal{S}) \rightarrow \text{minimize}\left(\mathcal{L}\left(\hat{\mathcal{S}}\right)\right) \quad (1)$$

Where, $\mathcal{L}$ is certain loss function.

Another vital aspect that needs considerable attention is to identify distortion and noise in the sensor captured physiological signals. For example, Alivecor captured single lead ECG contains significant noise particularly due to motion artifacts. In order to ensure mobility to the sensing applications and smartphone being the integral part in the ecosystem, noise identification and removal play important role for getting acceptably accurate clinical inference. In [Silv], the authors show that physiological signals captured even at controlled setup like in the ICU (Intensive care Unit) requires signal quality estimation and noise cleaning action. We have to note that presence of noise would invariably impact the prediction outcomes negatively and consequently false alarm rate would increase [Ukil17B]. Heartmate scheme described in [Ukil17A] proposed a robust denoising algorithm that identifies and eliminates corruption in physiological signals like PPG. In [Ukil16], an integrated analysis of unobtrusive cardiac health management and remote monitoring system CardioFit is proposed. Authors in [Ukil16], emphasize the aspect of clinical utility enhancement by physiological signal cleaning and removing distortion and noise. The complete learning pipeline in data-driven clinical analytics pipeline consists of:

- Pre-processing and noise cleaning
- Feature listing and feature selection
- Model building

Apart from pre-processing and noise cleaning, model building; feature listing and feature selection play a major role for the construction reliable aptly fitted learning model with the objective of avoiding overfitting on the training datasets.

Heart sound or phonocardiogram (PCG) is another vital marker of cardiac health which can conveniently captured using digital stethoscope or smartphone acoustic sensors. PCG signal is characterized by different markers like S1, S2 which are predominant, whereas murmurs, S3, S4 indicate the presence of cardiac anomalies. Authors in [Ukil17A] have demonstrated that smart analysis of PCG signal would reveal cardiac health condition and prediction of cardiac abnormality can be performed by studying PCG signals. Further, in [Ukil17B], noise reduction of PCG signal is presented. It has been shown that disease prediction model preceded by appropriate noise cancellation and removal block results in better clinical utility and higher accuracy of detection. One of the significant decision model of clinical analytics is that sensitivity of the model should be ensured very high and $\rightarrow 1$, which means that presence of cardiac anomaly will be captured with negligible failure rate, while specificity is maintained at decent rate, say > 0.8. We re-formulate equation (1) for practical model development purpose as:

$$\mathcal{M}(\mathcal{X}, \mathcal{S}) \rightarrow \text{maximize}\left(\mathcal{Specificity}\left(\hat{\mathcal{S}}\right)\right) \qquad (2)$$

Such that:

$$\mathcal{Sensitivity}\left(\hat{\mathcal{S}}\right) \geq \delta$$

Where, δ is typically > 0.9.

However, the proposed predictive analytics for remote cardiac health management would be useful for the care givers and partially adds value to the patients. The main outcome of predictive analytics like the presence of cardiac abnormality in a patient or the probability of cardiac damage recurrence are meaningful to the doctors, who can immediately provide diagnostic actions. We envision that smartphones with body sensors in the form of smart bands, smart patches will extract the physiological signals like ECG, PPG, PCG and analytics would be either performed locally at the smartphone or at the cloud. The prediction outcome when found important (i.e. cardiac anomaly is detected) is shared to the concerned stakeholders like doctors, hospitals or emergency service providers. Such ecosystem solves the problem of building cardiac health management partially. In such predictive analytics model, patients provide the data, which based on the action by the doctor, results in remote cardiac care. The main crux of this system is the complete dependency on the actions rendered by the human-in-loop. Circumstances may arise when timely action could not be taken. We envisage that prescriptive analytics, where the actions need to be taken is also part of the analytics system as illustrated in Figure 1.

Prescriptive analytics includes predictive analytics and descriptive analytics on prediction to instruct the patient to take actions. The outcome of the prescriptive analytics engine directly provides the patient with advices. Prescriptive analytics systems that reliably deliver instructions in healthcare applications are yet to be in deployable shape. The development process of prescriptive analytics involve enormous involvement of domain experts (in remote cardiac health management, cardiologists are the domain experts) such that the knowledge is sufficiently captured and a resilient rule engine is generated. Natural language processing based techniques can also be employed to build the knowledge representation. However, definite methods and systems to construct predictive analytics engine particularly for cardiac health data analysis and patient care instruction-based knowledge building researches would usher the development of complete cardiac health management with prescriptive and predictive analytics engines.

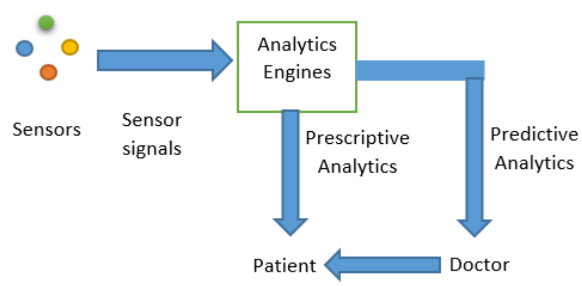

Figure 1. Architecture sketch of the automated unobtrusive cardiac care ecosystem catering both predictive and prescriptive analytics

4 Privacy Preserving Sensor Signal Mining: Ways To Minimize Human Value Loss Due To Intended And Unintended Privacy Breaching Attempts

Human quality of life improvement through knowledge management and analytics largely depend on sensor signals and data captured through sensing human activities. Such data often contains sensitive information. For example, energy consumption forecasting for optimal energy generation and carbon footprint minimization require smart energy meter data. Smart energy meter data contains granular information of inside home human activity, which are private and sensitive. Privacy breaching attacks on gaining access through Non-Intrusive Load Monitoring (NILM) needs to be minimized by detecting the sensitivity content of the shared information [Ukil15]. In [Ukil14A], 'Dynamic Privacy Analyzer' is proposed that controls involuntary leakage of smart meter data. The salient aspects of the proposed solution is that: It is completely unsupervised and attempts to find the optimal privacy-utility trade off while obfuscating the private smart meter data to third parties.

Traditionally, privacy-preserving data mining is implemented using *k-anonymity* [Swee02], *l-diversity* [Mach07] or other sensitive data anonymization techniques [Gentry]. However, we need to consider few of the specific aspects of security and privacy of the sensor data that capture human activity signatures. For example,

- Sensor devices, particularly body sensors are constraint with energy resources. Data transmission energy cost needs to be minimized to maximize the life span of such devices. Data transmission security with minimum energy consumption needs to be achieved using Constrained Application Protocol (CoAP) [Ukil14B].
- Sensitivity information requires secure storage and execution at the analytics engine at the analytics platform [Ukil10]. With the help of trusted computing (e.g. Trustzone), sensor data and computation are to be made secure resistant to data stealing attacks [Ukil11].

We depict the architectural sketch of the security and privacy methods of sensor data analytics management in Figure 2. Firstly, sensor data captured by the sensing device is to be securely transmitted with lightweight security implementation to the analytics platform, which may be at the cloud or locally available (smartphone). The captured sensor data is securely stored and executed by trusted computing setup. Further, the sensor data is privacy protected by required obfuscation and anonymity. The privacy preserved data is securely transmitted to the users. In fact, there are mainly three aspects of sensor data security-privacy framework:

- Data at transit:
  - Lightweight secure transmission from the sensing devices to the analytics platform [Ukil14B].
  - Secure transmission from analytics platform to the clients (users) [Ukil10].
- Data at storage:
  - Storage security for secure storing and execution of sensitive sensor data [Ukil11].
  - Privacy preservation of sensitive data before sharing with the clients (users) [Ukil10].

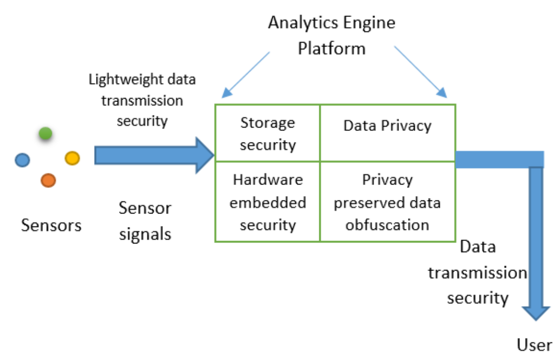

Figure 2. Architecture sketch of the secure and privacy preserved sensor data analytics

Another significant sensor data privacy protection policy would be privacy-preserving computation, where the analytics function is computed over encrypted data, without data being decrypted. Let, x_1, x_2 be the data from sensors s_1 and s_2. The analytics function $\mathcal{F}$ computes mean of x_1, x_2. The analytics engine receives encrypted data $x_1' = \mathcal{E}(x_1)$, $x_2' = \mathcal{E}(x_2)$, where $\mathcal{E}$ is the encryption function. The analytics engine can compute mean(x_1, x_2) from x_1', x_2' using homomorphic encryption technique [Ukil10]. In practice, useful fundamental analytics functions like summation can be computed in real-time through simplistic computational set up [Gentry].

5 Conclusion

Knowledge-driven technologies and applications for improving human quality of life will potentially enable long term human-centric convergence of futuristic applications. We have demonstrated exemplary cases of analytics for unobtrusive cardiac health management and privacy-preserving data mining of sensitive sensor signals. We observe that human-centric applications work closely with human activities and capture human behavior or other related sensitive information. Owing to the sensitive nature of such applications, security-privacy framework should be considered at the initial design time, as an integral part of the entire application eco-system. Another crucial aspect is to incorporate larger network of analytics to fathom the human actions and cognitions. For instance, social networking posts, retail consumption pattern, frequency of visit to physicians may be combined to derive the plan for personalized medication or cognition therapy. We envision that knowledge management, sensor signal processing and intelligent analytics system would immensely impact human life and the thrust of human-centric application would significantly improve the human quality of life.

Acknowledgments

Leandro Marin is partially supported by Research Project TIN2017-86885-R from the Spanish Ministery of Economy, Industry and Competitivity and Feder (European Union).

References

[Alive] https://www.alivecor.com/

[Heart] https://spectrum.ieee.org/the-human-os/biomedical/diagnostics/heart-monitor-for-your-phone-beats-doctors-at-diagnosing-atrial-fibrillation

[Jul]Julian P.J. Halcox, Kathie Wareham, Antonia Cardew, Mark Gilmore, James P. Barry, Ceri Phillips, Michael B. Gravenor. *Assessment of Remote Heart Rhythm Sampling Using the AliveCor Heart Monitor to Screen for Atrial Fibrillation: The REHEARSE-AF Study*. Circulation, 136 no. 19 (2017): 1784-1794

[Silv]Silva, Ikaro, Joon Lee, and Roger G. Mark. *Signal quality estimation with multichannel adaptive filtering in intensive care settings*. IEEE Transactions on Biomedical Engineering 59, no. 9 (2012): 2476-2485.

[Ukil17A]Arijit Ukil, Soma Bandyopadhyay, Chetanya Puri, Rituraj Singh, Arpan Pal, Ayan Mukherjee. *Heartmate: automated integrated anomaly analysis for effective remote cardiac health management.* IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2017): 6578-6579.

[Ukil16] Arijit Ukil, Soma Bandyopadhyay, Chetanya Puri, Rituraj Singh, Arpan Pal, KM Mandana. *CardioFit: Affordable Cardiac Healthcare Analytics for Clinical Utility Enhancement*. eHealth 360° (2016): 390 - 396.

[Puri17] Chetanya Puri, Rituraj Singh, Soma Bandyopadhyay, Arijit Ukil, Ayan Mukherjee. *Analysis of phonocardiogram signals through proactive denoising using novel self-discriminant learner*. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), (2017): 2753-2756.

[Ukil17B]Arijit Ukil, Uttam Kumar Roy. *Smart cardiac health management in IoT through heart sound signal analytics and robust noise filtering*. IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), (2017).

[Fras14] Fraser, Graham D., Adrian DC Chan, James R. Green, and Dawn T. MacIsaac. *Automated biosignal quality analysis for electromyography using a one-class support vector machine*. IEEE Transactions on Instrumentation and Measurement 63, no. 12 (2014): 2919-2930.

[Puri16] Chetanya Puri, Arijit Ukil, Soma Bandyopadhyay, Rituraj Singh, Arpan Pal, Kayapanda Mandana. *iCarMa: Inexpensive Cardiac Arrhythmia Management--An IoT Healthcare Analytics Solution*. First Workshop on IoT-enabled Healthcare and Wellness Technologies and Systems (2016): 3-8.

[Gim18] Jangwon Gim, Sukhoon Lee, and Wonkyun Joo. *A Study of Prescriptive Analysis Framework for Human Care Services Based On CKAN Cloud*. Journal of Sensors, (2018).

[Puri16] Puri Chetanya, Arijit Ukil, Soma Bandyopadhyay, Rituraj Singh, Arpan Pal, Ayan Mukherjee, and Debayan Mukherjee. *Classification of Normal and Abnormal Heart Sound Recordings through Robust Feature Selection*. IEEE Computing in Cardiology, Vol. 43 (2016).

[Thor13] Thornton, Chris, Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. *Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms.* In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, (2013): 847-855.

[Ukil14A] Arijit Ukil, Soma Bandyopadhyay, Arpan Pal. *Sensitivity inspector: Detecting privacy in smart energy applications*. IEEE Symposium on Computers and Communication (ISCC), (2014): 1- 6.

[Moli10] A. Molina-Markham, P, Shenoy, K. Fu, E. Cecchet. and D. Irwin. *Private memoirs of a smart meter*. ACM BuildSys (2010): 61-66.

[Ukil15] Arijit Ukil, Soma Bandyopadhyay, Arpan Pal. *Privacy for IoT: Involuntary privacy enablement for smart energy systems*. IEEE International Conference on Communications (ICC) (2015): 536-541.

[Swee02] L. Sweeney. *Achieving k-anonymity Privacy Protection Using Generalization and Suppression*. Int. J. of Unc. Fuzz. Know. Syst, (2002): 571 – 588.

[Mach07] A. Machanavajjhala, D. Kifer, J, Gehrke, and M. Venkitasubramanian. *l-diversity:Privacy beyond k-anonymity*. ACM Trans. Knowl. Disc. Data, vol. 1, issue. 1 (2007).

[Ukil12] A. Ukil, J. Sen, and S. Ghosh. *An Efficient Distribution Sensitive Privacy for Real-time Applications.* Computer Science and Convergence, LNEE, vol. 114, (2012): 81-91.

[Ukil14B] Arijit Ukil, Soma Bandyopadhyay, Abhijan Bhattacharyya, Arpan Pal, Tulika Bose. *Lightweight security scheme for IoT applications using CoAP*. International Journal of Pervasive Computing and Communications, Volume 10, Issue 4 (2014): 372-392.

[Ukil11] Arijit Ukil, Jaydip Sen, Sripad Koilakonda. *Embedded security for Internet of Things*. IEEE National Conference on Emerging Trends and Applications in Computer Science (NCETACS), (2011): 1- 6.

[Gentry09] Craig Gentry. Fully Homomorphic Encryption Using Ideal Lattices. ACM Symposium on Theory of Computing (STOC), (2009): 169- 178.

[Ukil10] Arijit Ukil, Jaydip Sen. *Secure multiparty privacy preserving data aggregation by modular arithmetic.* IEEE International Conference on Parallel Distributed and Grid Computing (PDGC), (2010): 344-349.

[Sen11] J. Sen, S. Koilakonda, A. Ukil. *A mechanism for detection of cooperative black hole attack in mobile ad hoc networks*. IEEE International Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 338-343, 2011.