Guest Editorial Preface

Special Issue on Artificial Intelligence in Data Science (AIDS)

Neetu Sardana, Department of Computer Science and Engineering, Jaypee Institute of Information Technology, Noida, India
Vikas Saxena, Information Technology, Jaypee Institute of Information Technology, Noida, India

In this special issue of the *Journal of Cases on Information Technology*, IGI Global on the theme of "Artificial Intelligence in Data Science" is envisioned to find the recent advancements, research, insights and viewpoints from scholars, researchers and practitioners on the use of Artificial Intelligence which is being used for data analytics and predictions.

Nowadays Artificial Intelligence presence can be felt in almost every area due to evolution in technology and development in complex algorithms. The algorithms that can handle the huge data easily. Artificial Intelligence has become a vital part of our society. Innovations in artificial intelligence and growing trend in data have driven new research opportunities in a variety of areas, such as social networks, education, healthcare, manufacturing business, mobile apps development, Internet of Things (IoT) etc. In this special issue we seek for articles that have applied artificial intelligence in the various applications where continuous data growth has been observed. For instance, World Wide Web (WWW), which is the fastest-growing segment of the Internet, growing at rate of 3,000 per cent every year. The WWW provides the ocean of data which is exponentially growing. The researchers have shown great interest in harvesting the knowledge from this platform as we have received ample papers that have explored and analyzed the data on the web with different perspective.

The special issue selected 11 high quality papers among those eight papers had an application towards information retrieval on WWW with different domain of study. These eight papers highlighted aspects of sentiment analysis on twitter dataset, developed location recommender system based on GPS trajectories, applied collaborative filtering on wiki pages, recommended user with the summary based on their reading pattern, personalized user with news articles and tweets, enhanced user interest features vector based on their profiles, examined the relevancy of each answer to why question posted on question and answering websites and finally visualized context-based sentiment language using signed social network. The remaining three papers worked on Android apps, Textile Production Line Monitoring System and Road traffic management systems. Each of these revised and extended papers has undergone full double blind peer review, prior to being selected for this special issue.

The special issue first paper "Aspect-Based Sentiment Analysis of Unlabeled Reviews Using Linguistic Rule-Based LDA" uses the Aspect based analysis for information processing. This paper describes the primary differentiation in sentiment analysis and aspect-specific sentiment analysis, such that the former only detect the sentiment of an overall text while the later investigates each text sentence to find out various aspects and then determine the emotion associated with each of them. The aspect-based analysis looks more closely at the information behind a text, resulted in an accurate and detailed results as compared with traditional methodology.

A location-based recommender system has been proposed in "EMD-Based Semantic User Similarity Using Past Travel Histories." The paper presents a novel earth mover distance (EMD) based semantic similarity measure using the past travel histories of users in form of GPS logs. It considers the semantic features while computing similarity along with the geographical features. The stay sequences for the users are computed from their GPS trajectories and then semantic tags are associated with every point in the stay sequences. Finally, the similarity score of the semantic trajectories were computed using the EMD. The paper demonstrates experimentally that the semantic stay sequence-based approach is computationally effective in comparison to semantic trajectory-based approach.

An interesting work on Wikipedia has been presented in "PerSummRe Gaze-Based Personalized Summary Recommendation Tool for Wikipedia." The work proposes a personalized summary recommendation technique which learns user's reading pattern in the form of their past summaries and uses it to find users with similar interests. The method works in collaboration with an eye gaze-based summarization system. The authors describe how one could utilize a user's eye gaze to capture their ROI in a Wikipedia article. The system filters the similar readers present in the system and appropriately integrates their summaries to generate a recommendation for the user.

"Quantifying the Impact of Biopics on Wikipedia Articles" presents the different domain of study that uses the collaborative filtering on wiki pages. The study presents the insights about the quality statistics of Wikipedia links and pages. The presented work relates on the impact of external factors over the quality of wiki pages, whose impact was even seen on the search engine retrieved results like Google search, once wiki pages were either modified, updated, or even in cases of broken and deleted page links. The study impacts on the use of words, wiki links, and references on page to measure the quality of the page statistically.

The more updated work of similar aspect has been presented in "Semantic Term-Term Coupling-Based Feature Enhancement of User Profiles in Recommendation Systems," where the Non-IIDness learning concept is discussed with semantic information retrieval. The work is presented on the retrieval of the user profile using term-term coupling which acts as a content in profiling. The presented approach refers to understanding, modeling, analyzing, and representing not independent and identically distributed data. The work also put emphasis on the coupling and heterogeneity, which acts as an important aspect of Non-IIDness, in information retrieval domain. The study fetches the attention towards the terms of the feature vectors which are considered to be independent but there exist couplings between the terms. The empirical study of the work proved that with Non-IIDness learning, the sematic information can be preserved and used in making better user profiling in recommendation systems.

In the paper "**Towards Intelligent Road Traffic Management Over Weighted Large Graphs Hybrid Metaheuristic-Based Approach,**" a new hybrid genetic algorithm named IOGA (Improved Optimization Genetic Algorithm) that integrates a genetic algorithm with Dijkstra's algorithm for computing the k-optimal shortest paths on a road network to minimize distance and time from the origin node to the destination node. The results demonstrate the method to be efficient in terms of runtime and quality.

The work in "Analyzing Linguistic Features for Answer Re-Ranking of Why-Questions" re-ranks answer candidates of why-questions using feature selection methods. Various features covering lexical-syntactic, semantic, and contextual similarities have been employed to find the relevancy of each answer candidate to a question.

A work in different dimension has been shown in "Malware Detection in Android Apps Using Static Analysis." The paper proposed a permission induced risk interface namely 'MalApp' classifier to identify privacy violations that rise from granting permissions during app installation. In their approach concept of reverse engineering using static analysis has been applied to construct an app permission matrix. Permission ranking was applied to identify the riskiest permission across a category. Heuristic approach has been proposed which integrated machine learning and ensemble learning techniques to classify malicious Android apps.

Another recommender system has been proposed in "PNTRS-Personalized News and Tweet Recommendation System." The work personalize user with precise, trending and specific news articles of their preference and also exhibits tweets having discussions, viewpoints, and opinions of different people on those recommended articles. The works utilizes the fuzzy inference system to recommend personalized tweets. The proposed method has obtained encouraging results.

In "Self-Aware Contextual Behavior Analysis for Service Quality Assurance Over Social Networks," a self-aware computational model has been built for context-based sentimental language analysis, visualized through a signed social network. The overall sentiment analysis provides knowledge about an abstract sentiment, while context-based analysis helps us to find the major contexts from a list of topics and the mood associated with each such context. The analysis has been performed on real time user comments and feedbacks from Twitter portals on Zomato and Swiggy food delivery services.

Finally, "Textile Production Line Monitoring System Using Wavelet-Regression Neural Network" presents a novel approach towards the wavelet transform using a regression neural network to monitor the textile production power line system. The work shows how to access and utilize the center data for generator stability. Furthermore, learn the strategies to reduce the risk of instability by making efficient and realistic plans and designs.

We are extremely happy to bring out this special issue and sincerely thank Editor JCIT, IGI Global and all reviewers for valuable comments and suggestions, which helped us to improve the quality of the article. We dedicate this issue to all those who have made their best efforts to contribute to this publication. May these contributions pave the way for the future research in the area of Artificial Intelligence in Data science.

Neetu Sardana
Vikas Saxena
Guest Editors
JCIT