

## Papers presented during the conference on 16<sup>th</sup> December, 2019.

Time: 4.15 PM to 6.00 PM

### Paper-1:

**Title:** Emotion-based Mining for Gender Prediction in Online Social Networks

**Authors:** Mudasir Ahmad Wani, Patrick Bours, Nancy Agarwal, Suraiya Jabin.

**Abstract:** In online digital platforms, it is easy for people to either hide or provide false personal information such as gender. Therefore, with the rapid growth of the Online Social Network (OSN) user population, the gender prediction problem has received extensive attention from researchers, as it has many applications in several domains, such as marketing, political campaigning, and security. For designing a gender prediction system earlier studies have majorly focused on stylistic and content-based features, mined from user-generated texts. However, research communities have also outlined that texts written by the users on OSN are highly driven by their emotions. Hence, this paper aims to investigate the potential of emotion-based features in the gender identification task, which has been unnoticed by researchers so far. The experimental study is carried out on the texts of two widely used OSNs, Facebook and Twitter.

In this paper, we attempt to answer three research questions: 1) whether there exists emotion-based gender difference in the text content generated on Facebook network, 2) would the emotion behavior change across OSN platforms (e.g., Facebook vs. Twitter), and 3) do the emotion-based features aid in designing gender prediction system. We formulated 12 emotion features based on Plutchik's basic emotions. Furthermore, we also used the Moodbook lexicon to mine the emotions from the user-generated text. The findings revealed that Facebook users exhibit markedly more emotional differences than Twitter users. With several machine learning techniques, the highest AUROC (Area under the Receiver Operating Characteristic) values reported by Facebook and Twitter gender classifiers are around 87% and 60% respectively.

**KEYWORDS:** *Online Social Networks, Facebook, Twitter, Emotion Analysis, Gender Identification, Machine Learning, Gender Prediction*

## Paper-2:

**Title:** Hybridization of Constriction Coefficient based Particle Swarm Optimization and Gravitational Search Algorithm for Constrained Engineering Optimization

**Authors:** Sajad Ahmad Rather, P. Shanthi Bala.

**Abstract:** The hybrid constriction coefficient-based particle swarm optimization and gravitational search algorithm (CPSOGSA) is a novel heuristic optimization algorithm. The PSO is inspired by the group behavior of bird flocking and fish schooling whereas GSA is a physics-based stochastic technique inspired by the law of gravitation. Moreover, the constriction coefficient is important for increasing the convergence speed of the standard PSO. The exploration and exploitation capabilities of PSO and GSA respectively are utilized to find the global minimum. Generally, stochastic algorithms perform well in many practical real-life problems. In this work, the CPSOGSA is applied to four constrained engineering design benchmarks including welded beam design, compression spring design, pressure vessel design, and speed reducer design. The experimental results of CPSOGSA have been compared with seven heuristic algorithms. Also, the Wilcoxon rank-sum test is performed to statistically validate the results. The simulation results indicate the efficient performance of CPSOGSA in terms of minimization of the cost function, design variable optimization, and successful constraint handling as compared to other participating algorithms.

**KEYWORDS:** *Gravitational Search Algorithm (GSA), Particle Swarm Optimization (PSO), Engineering Optimization, Hybridization.*

## Paper-3:

**Title:** Comparative Analysis of Machine Learning Algorithms for Fake News Detection

**Authors:** Nikhil Madaan, Siddharth Dharm, Aruna Malapati.

**Abstract:** With a surge in the number of electronic devices connected to the Internet, the propagation of news (both, legitimate and fake) is inevitable. The dataset comprises of 9,400,000 labeled news articles. In this paper, we propose to use readability and linguistic features extracted using NLP (Natural Language Processing) techniques, which help in substantially improving the performance of the classifiers. Our experimental study consists of comparing the performance of eight classification algorithms, thereby suggesting the best algorithm for detecting fake news.

**KEYWORDS:** *Natural Language Processing, Computational Linguistics, Artificial Intelligence, Deep Learning, Artificial Neural Networks, Recurrent Neural Networks.*

#### Paper-4:

**Title:** Efficient computation of top-k skyline objects in data set with uncertain preferences

**Authors:** Nitesh Sukhwani, Venkateswara Rao Kagita, Vikas Kumar.

**Abstract:** Skyline recommendation with uncertain preferences has drawn the attention of AI researchers in recent years due to its wide range of applications. The naive approach of skyline recommendation computes the skyline probability of all objects and ranks them accordingly. However, in many applications, we are interested in determining top-k objects rather than ranking all the objects. In this paper, we solve the problem of determining top-k skyline objects without having to compute the skyline probability of all the objects. The most efficient algorithm to determine the skyline probability of an object is based on the concepts of zero-contributing set and prefix-based k-level absorption. We show that the performance of this approach relies on the order of the objects. We also propose and analyze different ways of ordering these objects to maximize performance. We show the efficacy of our approaches through experimental analysis on real and synthetic data.

**KEYWORDS:** *Skyline query, Skyline computation, Uncertain preferences.*

#### Paper-5:

**Title:** Synthesis/Programming of Hopfield Associative Memory

**Authors:** Garimella Rama Murthy, Lankala Krishna Vamshi, Devaki Nimmagadda, Sree Divya Bondalapati.

**Abstract:** In this research paper, synthesis of Hopfield Associative memory with desired orthogonal stable states is discussed when the threshold vector is non-zero (and the dimension,  $N$  is even). Synthesis of one desired stable state in Hopfield Associative Memory (HAM) when  $N$  is odd is discussed. Based on relationship between Hopfield Associative Memory and associated graph-theoretic code, noise immunity issues are discussed.

**KEYWORDS:** *Eigenvalues, Eigenvectors, Synthesis, Orthogonality, Hopfield network, Hadamard Matrix.*

## Paper-6:

**Title:** End to End Audio Transformation across Arbitrary Targets

**Authors:** Shashwat Aggarwal, Shashwat Uttam, Sameer Garg, Shubham Garg, Swati Aggarwal.

**Abstract:** We present a fully differentiable end-to-end audio transformation network to transform the style of one audio to another. Our method has three key features to offer: (a) It does not require parallel utterances, transcriptions, or time alignment procedures, (b) It employs a global conditioning mechanism and is training vocabulary agnostic and thus can transform the style from one audio to another irrespective of target identity and, (c) It achieves one-shot audio transforms without using any intermediate phonetic representations, thereby eliminating the need for phonetic alignments and speaker-independent ASR networks. We evaluate our method with the existing approaches on two audio transformation tasks, namely, voice conversion, and musical style transfer. Subjective evaluation confirms the superiority of our approach.

**KEYWORDS:** *Voice Conversion, Musical Style Transfer, Audio Transformations, End-to-End Audio Pipeline.*

## Paper-7:

**Title:** Relative Effectiveness of Machine Learning Techniques in Imputing Missing Data

**Authors:** V. Anand, Varsha Mamidi.

**Abstract:** Business and management disciplines have been dealing with the problem of missing values in data for decades using statistical methods. Machine learning is being used for data analysis but rarely used for imputing missing values during the data preprocessing phase. Also, there is very little evidence of systematic evaluation of machine learning-based methods vis-à-vis statistical methods in business and management. This study provides a comparative evaluation to advance machine learning as a prudent way of imputing missing values. We evaluate mean imputation, multiple imputation, k-nearest neighbors imputation, sequential regression tree imputation and sequential random forest imputation on four real-world datasets. Our results indicate that multiple imputation and sequential random forest imputation outperform other methods in specific scenarios. The generalizability of these results requires further research in this direction.

**KEYWORDS:** *Mean Imputation, MICE, KNN, Regression tree, Random forest.*

## Paper-8:

**Title:** TRACS: Transformer for Video Captioning and Summarisation

**Authors:** Aman, Anand Zutshi, Apaar Gupta, Swati Aggarwal.

**Abstract:** Video summarization is the process of creating a meaningful and self-explanatory summary of a given video by automatically selecting keyframes and thus creating a short and concise video summary of the original video clip. Video captioning refers to the task of automatically generating natural language description of a given video to provide some additional or interpretive information. Video summarization and video captioning are often considered as two different tasks, each with a diverse application field. Thus, we propose an approach to jointly combine these two tasks and present a model which generates a short video summary along with a relevant caption. We make use of self-attention-based transformer network combined with Multi-Layer Perceptron (MLP) with multiple hidden layers for video captioning and video summarization. We aim to demonstrate that the joint model can attain better performance than many of the previous approaches in both of the individual tasks. Our proposed model is based on the fact that both of the above-mentioned tasks of captioning and summarization will improve the performance of the other one as both the tasks are complementary to each other.

**KEYWORDS:** *msr-vtt, msvd, tvsum50, deep learning, natural language processing, video summarization, transformers, attention.*

## Paper-9:

**Title:** The Information Content of Analysts' Blogs: A Machine Learning Approach

**Authors:** Varsha Mamidi, Chandu S R A Kunisetty, Vijaya B Marisetty.

**Abstract:** Analysts' blogs regarding earnings of a company will carry significant amount of information regarding stock price movement before the earnings announcement day. However, each article will have a different view on the movement of stock price. Unlike existing studies in finance and accounting, we use a machine learning approach to arrive at analysts' consensus and create a trading algorithm. By using N-Grams based approach for features extraction from the blogs and Support Vector Machines to predict movement from the extracted features, an accuracy of 58.65% is obtained. Word embeddings approach for features extractions and Long Short-Term Memory based Recurrent Neural Network for prediction resulted in an accuracy of 59.62%. Instead of word embedding, when document embedding was used, the accuracy improved to 63.46%. Our trading strategy, when document embedding was used, around earnings announcements of India's top 30 companies, gave an average cumulative abnormal return of 1.17% for 10 days holding period. Our results suggest that machine learning based trading strategies have

**KEYWORDS:** *Analysts recommendations, SVM, N-Grams, Word2Vec, Doc2Vec.*