**Knowledge Discovery From Social Media Data**

Name: Dedeepya Yarra studentID:11594115 EmailID:DedeepyaYarra@my.unt.edu

**Purpose of study** social media networks generate enormous amounts of data that can be utilized to get insightful data on the dynamics, sentiment, and opinions of people. Data analytics models can help with the difficult work of extracting a bigger amount of data from social media platforms and analyzing it to determine public opinion, identify user group behavior in group events, and do sentiment analysis.

**Related literature** Techniques of opinion mining by Neha Gupta and Rashmi Agrawal, Authors discussed some key area related to feature extraction, deep learning and ontologies. [Trajectory data mining](https://digitalcommons.library.umaine.edu/cgi/viewcontent.cgi?article=1084&context=josis) by Jean Damascene Mazimpaka, Sabine Timpfmain. Authors explained integrated view relating to applications of mining trajectory data.

**Research design and strategy**

**Social Big Data Analysis Applications**

**1.Frequent Trajectory Mining from Social Media Data** It uses AUDSOME Methodology. It is method of extraction of most frequent user trajectory from Point of Interest (POI). In this method keywords are extracted from social media posts which are used to identify the Point of Interest. Key word extraction is as follows. **keyword discovery**-The region of interest was split into equal cells, and each cell was allocated a post based on its geolocation. Then, in each cell, keywords are discovered based on the frequency and description of the post. **keyword selection-** To discriminate between low frequency and high frequency terms, a discrete L curve elbow point is employed. The method will compute frequency by taking global and local frequency into account in each cell. keywords are selected based on their frequency. **keyword grouping-** Using Levenshtein's metric technique, comparable terms are clustered together. The algorithm generates a series of terms that are related, and each set represents a POI.

**ROI detection**: DBScan algorithms is used in Region of Interest (ROI) detection. The DBSCAN method requires two critical parameters: eps, which is the neighborhood radius with respect to point, and MinPts, which is the minimal number of points necessary to create a cluster. For each location, compute the k closest neighbors. Determine the threshold point and group points depending on the threshold value. The threshold value is determined by estimating the amount of noise, threshold point K-dist value is used as the first parameter eps, and the minpoint value is k+1.

**Trajectory mining:** To determine the most common set of ROI’s, a sequential pattern analysis approach is applied. For each frequent trajectory, we must determine the support and confidence. by determining a threshold value for these two variables. We can determine the user's most common path of travel.

**2. Opinion Mining from social Media Data**

**Methodology:** IOM-NN methodology was used for opinion mining. Process is as follows. **Keyword-identification**: The task of identifying postings that include a certain set of predefined keywords and fractions is known as keyword identification. **Classification of posts:** Feed forward, neural network, and multilayer perceptron techniques are utilized for categorization of postings. The procedure is iterative; in the first round, the postings are categorized in favor of groups, and in the subsequent rounds, neural networks find new rules and are trained with them. It will take use of the previous categorization; the cycle of iteration continues until either the threshold is met or there are no more posts to categorize. **Polarization of users:** Users are taken into consideration if they have at least a certain number of postings related political events and if at least two-thirds of their posts have been published to a faction. The categorized posts from the previous stage are grouped to users, and each user's posts are grouped together. In order to get the polarization percentage for each faction, which is result of political event, the faction score for each qualifying user is computed, and all fraction scores are ultimately calculated and normalized based on this faction score outcome of event is predicted

**3.Topic Discovery from social media Methodology**: The hashtag recommendation model HASHSET was created for the proposal of a suitable group of hashtags for a certain content. **Problem:** Although postings are hash tagged, there are no guidelines or limitations for hashtag selection, hence many posts lack a representative hashtag. IOM-NN performance is impacted by this, however it may be fixed by using the right hashtag recommendation model. **Process**: semantic mapping is done with teaching a multilayer perceptron to reduce cosine distances for subsequent latent vectors. For Input post latent vector and K closest hashtags in order of cosine similarity are recommended. Based on the recommended hashtag posts are classified into as assignable, ambiguous, and neutral and topic are discovered.

**Conclusion** Social networking is a daily activity that more than 70% of internet users engage in for at least an hour. we must conclude that social networks have developed into a type of reality in which people talk, engage, and obviously trust. By this we need to determine the significance of social media data analytics and the amount of effort we should continue to put into future research on social media data analytics.

**Contribution** Author explained social media data-analytics in 3 major viewpoints (i) Major tourist sites based on posts published by tourists. (ii) Opinion analysis of political event based on user posts regarding political factors (iii)hashtag recommendation and topic discovery from user’s posts

**Overall assessment** The author's detailed explanations made the material easy to read. In addition, the author did a great job of keeping a pleasant flow while offering three top-notch illustrations of social media data analytics. The paper's primary merit is its succinct treatment of the three commercial use cases, the underlying algebra, and outcome analysis. I must commend writers' writing abilities.

**Research methodology**

The study explained Socle framework for data normalization, data enrichment, and data trimming, and MapReduce framework for data collecting, storage and analytics, unsupervised data framework for filtering, extraction, and categorization, and cloud-based frameworks for managing substantial amounts of social media data platforms APIs.

**Model for Data Analytics: ParsoDA (Parllelsocial data analytics)**  ParSoda is employed to enable the cloud-based operation of these applications. The ParSoDA's runtime was designed from the ground up to handle massive volumes of data. It is based on the MapReduce framework. It was a library where a sizable amount of data gathered from social media sites was processed and analyzed. Complex algorithms can be parallelized with seven simple phases as follows obtaining data, filtering it, mapping it, dividing it up, reducing it, analyzing it, and visualizing it.

**Meta data Model** The major concern when dealing with social media data, is adopting to a standardized data model for data extracted from heterogenous sources. JSON document with 2 sections, first section called basic is for major data like source, location coordinates, user information etc. which is common across all social media sites, second section called extra is for fields specific to the source.

**Future Research**

Future social media data analytics study will use models built to increase customer satisfaction. The models have been improved for greater marketing effectiveness. The models are made to learn from customer’s, according to their interests. The scope of innovation is a crucial component of future research.

**New knowledge learned** From the article I learned socle, MapReduce, Unsupervised and cloud-based frameworks. And understand the implementation of opinion mining, trajectory mining and topic discovery and Advantages of social media data analytics.

**Questions to discuss** Some of the questions we need to consider while dealing with social media data are as follows. does social media analytics was compromising on individual’s privacy? How is data transformed from one platform to another platform? Every second large amount of data is collected from social media, where and how these data is stored and managed?

**References**

1. User Trajectories from Social Media Posts by Loris Belcastro; Fabrizio Marozzo; Emanuele Perrella. [[CrossRef](https://link.springer.com/article/10.1007/s00146-014-0549-4)]
2. Learning Political Polarization on social media Using Neural Networks by Loris Belcastro; Fabrizio Marozzo; Emanuele Perrella. [[CrossRef](https://linkinghub.elsevier.com/retrieve/pii/S0957417421011131)]
3. Learning Sentence-To-Hashtags Semantic Mapping for Hashtag Recommendation by Cantini, R.; Marozzo, F.; Bruno, G.; Trunfio P.