Meta-Heuristics and Information Diffusion Based Community Detection Approaches using Hybrid Semantic Algorithm for Online Social Networks

SankaraNayaki K[1], SudheepElayidom M[2], R Rajesh[3]
[1]Research Scholar, Division of CS, SOE, CUSAT, Cochin, Kerala, India.
[2] Professor, Division of CS, SOE, CUSAT, Cochin, Kerala, India.
[3] Scientist III, NPOL, Cochin, Kerala, India.
[1] sankaranayaki@gmail.com,[2] sudheepelayidom@gmail.com,
[3] drrrthampi1971@gmail.com

Abstract

The semantic social network is a type of network that has enormous nodes and intricate semantic information; however, the conventional community detection algorithms were unable to give the expected coherent communities in its place. We present a clustering community detection algorithm that is based on the Hybrid Honey Badger Optimization Modified Latent Dirichlet Algorithm (HBOmLDA) model in order to solve the problem of detecting semantic social networks. We use the Particle Swam Optimization (PSO), which is able to make quantitative parameters map from semantic information to semantic space. Given that the semantic model is an LDA model, this is necessary. Then, in order to solve the problem of overlapping community detection, we present a HBO strategy that incorporates a semantic relation. In the end, we establish semantic modularity, in order to evaluate the semantic communities that were found. The experimental analysis demonstrates that the Hybrid HBOmLDA model's validity and feasibility, in addition to the semantic modularity, the search space is also optimized using Meta-Heuristics Algorithm..

Index TermsLDA, HBO, Community Detection, Meta-Heuristics Algorithm.

DOINumber:10.14704/nq.2022.20.8.NQ44720 **NeuroQuantology2022;20(8):6938-6951**

INTRODUCTION

The use of social media has, over the course of several generations, essentially replaced more conventional modes of social interaction and conventional ways of communicating with others by utilizing the technology that is now a part of our day-to-day lives [1]. The price was recently lowered, and some minor to moderate concessions were made in order to appease a large number of customers. Researchers believe that there will be 12.74 billion people using the internet in the year 2021, and they estimate that there will be 11.79

billion people using social media [6]. These days, more than fifty million businesses use Facebook as a channel through which they can communicate with their clientele. Such a specific interest as providing a spark to an event can occasionally lead to a viral enthusiasm. This is especially true among those individuals who are not users but are frequently interested in the item; buzzy media refers to these types of media [10].

Over the course of the past two decades, social networking websites have witnessed a meteoric rise in the total number of viewers as well as the quantity of content that has been shared on a global scale. Users are able to communicate with one another and with each other, discover and distribute content, and share information thanks to the services that we provide. The phrase ”as having social links” is used to describe a significant portion of these social media websites (e.g., Facebook and Twitter). Some of them, like YouTube and Flickr and Facebook, are helpful for exchanging ideas and developing social connections with fellow students and employees. Finding the actions of specific people on websites is one of the more significant challenges people face, and it is also one of the more challenging problems to solve.

Our topic of discussion is a scenario that involves social media platforms like Twitter and other similar websites. At the end of 2017, Twitter had over 500 million active users and a day in activity, which was only a slight increase from the figures seen in 2012 and 2016, when it had over 300 million and then 330 million active users respectively. Twitter was established in July of 2006 and had over 500 million active users and a day in activity at the end of 2017. I checked Twitter in late 2020, and there were 187 million daily users. In addition, the questions that can be asked through Twitter are limited to 280 characters each (2019). T there will be an influx of data, whether it be linked to followers or following other people online (for example, groups and sites), or what social media may suggest to the user via notifications (i.e., trying to follow or you), it is a continuous stream of information. T there will be an influx of data whether it be linked to followers or following other people online (for example, groups and sites) (notifications that let you have continuous streams of data). Getting suggestions on how to use the top stories is one of the relevant strategies that the user is suggested to implement for geographic location as a Facebook user. This is done as part of the Facebook user’s experience.

It takes time and knowledge to properly configure certain hidden parameters.In this article, we propose a novel community detection algorithm with the goal of grouping nodes into clusters. This algorithm’s name is clustering nodes.The fact that members of the same community share interests that are either identical or very similar is the defining feature of the communities that are uncovered by this algorithm. After taking into account the information on the topic and keywords in the text that comes from the words of individuals through the mLDA model, we quantize the semantic nodes and then map them into the semantic space. After utilizing the Honey Badger Optimization algorithm, we are able to create the perfect digital communities for social interaction.

LITERATURE SURVEY

Because we expand the knowledge and expertise to boost the LDA models by adding the entity relations generated by our research, we now review 4 classes of related to our Current Paper.

A. Latent Semantic Analysis:

The techniques of LSA, which are also known as latent semantic analysis, are the foundational work that goes into developing conceptual models. Our content consists of a very specific ”parasite matrix,” as well as the terminology associated with it. This terminology is divided into two parts: a document matrix that contains documents, and a matrix of topics and terms. In order to accomplish this, the first step is to create a document-term matrix. When we have a list of m documents and a list of n words, we are able to construct a matrix with the dimensions m by n, in which each row represents one document and each column represents one word. Each LSA entry is comprised of an unrounded word count regardless of the number of times the letter 'j' appears in the document to which it corresponds that has not been analyzed. This is the entry's most basic form. Counting the raw words does not work as well in practice, but it does provide a more accurate estimate because the influence that each word has on the meaning and importance of the whole is taken into account. When attempting to convey knowledge about the subject of a document, it is frequently preferable to use general terms such as "nuclear" rather than specific terms such as "fission", "test," or "analyzer," because the former enable us to investigate the full scope of conceivable concepts related to the latter [11]. This, As a consequence of this, LSA models are typically utilized in the process of substituting a tf-idf score for the raw counts contained within the documentterm matrix. A score is assigned to the term j in document I based on its frequency, which is calculated by multiplying the total number of times it appears in the document by the number of times it appears in the document for the first time.

Algorithm	Description	Size	Reference	Data
Dictionary-based analysis	Development of a statistical learning model	118,000	Quinn 2010	Speeches (Text)
LDA	Identify literature themes	3346	Jockers 2013	Books
LDA	Identify topics of German politicians	2581	Baum 2012	Speeches (Video)
LDA	Identify tweets related to obesity	2,581,283	Ghosh 2013	Tweets
Dictionary-based analysis and LDA	Compare dictionary-based analysis vs. LDA	77,000,000	Guo 2019	Tweets
LDA	Investigate the validity and reliability of LDA	344,456	Maier 2018	Web pages
LDA	Investigate how climate change is discussed in blogs	1,300,000	Elgesem 2014	Blogs
LDA	Investigate the political agenda of Russians in LiveJournal	> 100,000	Koltsova 2014	Web forum posts
LDA	Investigate how Twitter has been used in academic conferences	109,076	Parra 2016	Tweets

Figure 1. Literature Survey of LDA based Algorithm

A truncated singular value decomposition can be used to accomplish this kind of mathematical reduction. The singular value decomposition [8] (SVD) is accomplished by first multiplying the given matrix by a diagonal matrix, and then multiplying the result by the singular values of the initial matrix to factorization (or other weights), which results in the production of a new ones-based submatrix that is denoted by the letter U. only the t of the largest values and the singular vectors, V, with a signal to firsts that has a reduced entropy, are entered into the singular value decomposition. Because we have a specific preference for the number of topics that we would like to search for, t can be defined to give us control over that particular aspect of the search. An intuitive way of thinking about this situation would be to imagine that only the dimensions that are the most important to our space-expanded form are still there. In relation to this scenario, if we select U to serve as our document matrix and V to serve as our term matrix, then U can be interpreted as a function, whereas V cannot. There is at least one entry in either U or V that is pertinent to the one or more topics that we are discussing. In the case of topic U, a topic can be stated as an assortment of documents that are pertinent to the topic, and in the case of topic V, a topic can be stated as a collection of documents that are pertinent to the topic.

B. Latent Dritchlet Analysis:

Dirichlet Allocation is referred to as an expanded model. LDA is a statistical, probabilistic form of pLSA. A primary advantage of Dirichlet word is that lends itself to better generalisation in particular is that it uses Dirichlet-random priors for the document-topic and word-topic distributions. Dirichlet is basically means that something has a more wide distribution than anything that falls into any other class or group or type. Essentially, to this question, it provides the response: ”in light of this distribution, what are some common probability distributions that I am likely to see?” Looking at the example of comparisons of PD of topics may help to expand on your understanding of these concepts Let’s consider the body of data that we are analysing to be made up of records from three different topic areas. If we wish to have a distribution heavily weighting on a specific topic, we would want to use weights to come into play in only a small percentage of its calculations. Since we’re dealing with 3 topics, the most likely results would be to occur include the following:

Container A: 80% T A1, 15% T B1, 5% T C1
Container B: 15% T A1, 80% T B1, 5% T C1
Container C: 15% T A1, 5% T B1, 80% T C1

If we randomly drew probabilities from this sample from this Dirichlet distribution, the parameters of which have equal topic weight, would produce a distribution that is likely to resemble either a container distribution. We’d be very doubtful to collect if we sampled the triangle shaped, cube shaped or to sample a distribution that is 30% TA1, 30% TB1, and 40%TC1.

That is basically what a Drichlet solution provides: a random variable in which the type of random process is specifically expanded and Studied in various papers as shown in Figure 1.

C. Grouping Product Features in Opinion Mining:

Zhongwu Zhai, Bing LiuHua, XuPeifa Jia proposed a paper on the review opinion mining of products,which is frequently looking for features. Nonetheless, the same feature can be described in different ways. To be an effective synopsis, all such words or phrases must be associated with the same concept. Topic modelling is appropriate for the task However, we believe that topic modelling can be performed better by using pre-established constraints instead of free exploration. First, we extend a popular topic modelling approach, called Latent Dirichlet Allocation (LDA), to handle large-size limits. Finally, two new constraint extraction methods are proposed. Finally, the resulting circumscribed features are used on the products. Experiments demonstrate that LDA wins by a large margin

D. Semi-supervised learning using a constrained labeling LDA model:

In the case of problems where labelled data is present, semi-supervised learning (mSSL) methods are powerful but semi-supervised techniques with this approach can additionally make use both labelled and unlabeled data. In this paper, we describe an alternate generative semi-supervised discriminant semi-expansion algorithm, the semi-supervised generative discriminant (LSDA) is utilised. Empirical results with synthetic data to find out how the method behaves, and compare that to actual world data; then we then we perform experiments on several real and representative data sets to find the actual strengths and weaknesses of the method.

E. Labeled LDA - A supervised topic modelling:

A large percentage of the world's written text is searched and reviewed by people on social relinking websites. Innumerable tags are an alarming concern because almost all pages have them, but they're not evenly applied throughout the entire document credit attribution solution is required in order to be solved. To get that done, the first step is attributing each word with the relevant tags and then the process follows. This paper defines Labeled Dirichlet Allocation (LDA), a model that permits Dirichlet allocation to be both constrained to have one-to-one interaction between topics and tag values. If you enable the word feature 'Expand', Labeled LDA will learn word-equivalences directly.We emphasize the benefits of De-elaborate labelling a small lexicon of the URLs in the Delicious library by mapping them to delicious. bookmarks. Labeled LDA outperforms SVMs by over 3 to 1 in its ability to find tag-specific text, and is about 1.4x faster at random snippet extraction. While the bigoted baselines for multiplelabel classification models have not yet been studied, we compare competitively to existing models.

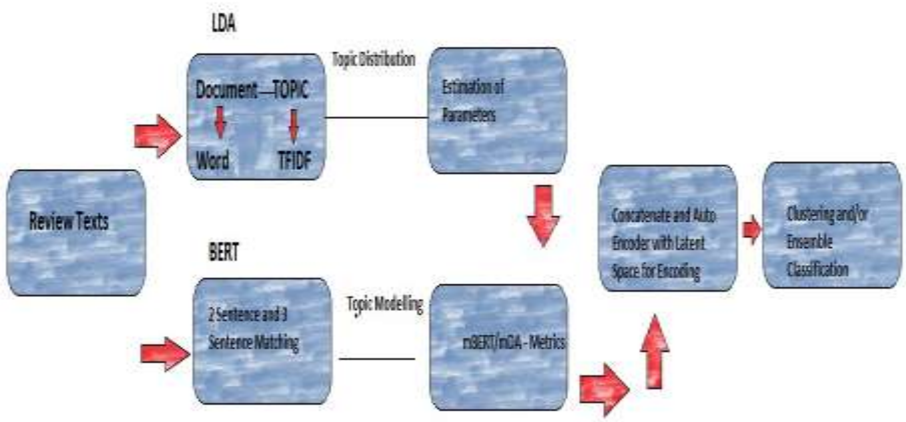

Figure 2. Overall Process.

6942

PROPOSED SYSTEM

This design platform enables the searcher to identify the market's best devices. Predicted on the phone with user comments Association/association prediction is improved with Constrained LDA The information is fed into BERT Restricted LDA is a generative model that establishes topics in a document collection a bag-of-word representation of documents is given to the system And every document will have a set of latent topics, as well as a set of topics assigned. Access levels will be determined by system architecture, system functionality, security level, knowledge standard, and levels of privilege. They aim to use the solution , Regardless of their knowledge or abilities.The Data is given as input to The Transformer method of applying bidirectional training which is introduced via BERT. This differs from previous studies which viewed the text from left to right and then also right-to-left.The study's findings show that bidirectional language models can store more context and content than single-direction models They describe a novel method named Masked (ML) which had previously been unenviable.Let us now discuss on the Working and Concepts. BERT

[20] uses Transformer, a mechanism that focuses on contextual relationships in a text Transformer has two separate mechanisms — the input reader and the decoder. Only the encoder mechanism is needed in order to generate a language model in BERT.In contrast to models, which read text in a sequence (leftto-to-right or right-to-to-left), the Transformer.

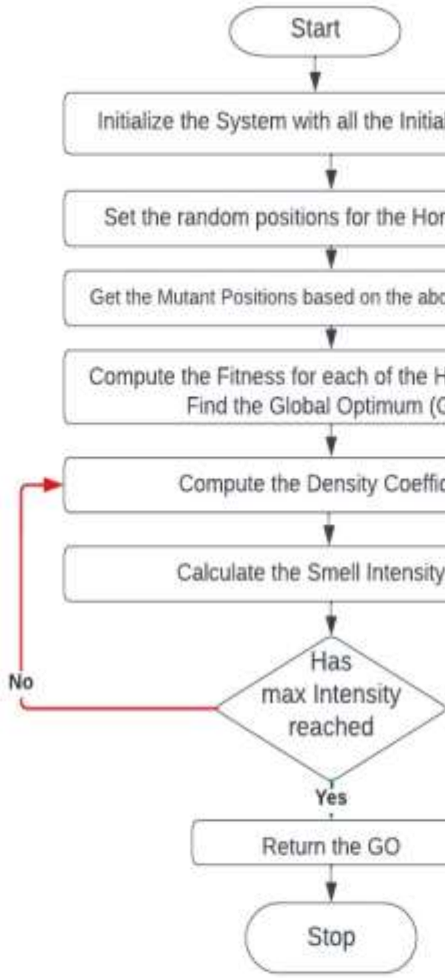

Figure 3. Flow Chart of HBO Algorithm

Reads the entire sentences.The TF-IDF of our data is depicted in Figure 2. Thus, it is bifurcated: Non-directional is a better word to describe its status. This dictionary characteristic enables the model to discover the surrounding context of any word (left and right of the word).The Figure 2 also exploits the components of the Transformer encoder. The activation function is a series of vectors, which will then be mapped to numbers. the sequence has H-dimensional input tokens, each of which corresponds to an output token with the same index. The most difficult problem faced in learning language models is defining a forecast goal. There are many models that state the next words will be spoken (e.g. "The child returned home from ") as sequential, which can be expected to restrain how much the child learns about the context. To surmount this obstacle, two training approaches are employed:

1) PSO-LDA: : PSO is a smart optimization algorithm. Kennedy and Eberhart [35] first proposed it.PSO has simplified, quick convergence speed and fewer controlling parameters, etc.Compared with Genetic Algorithm (GA), Ant Colony Optimization (ACO), and Simulate Anneal (SA), PSO has two attractive features: first, PSO optimizes the solution from the local optimumfirst and runs fast, making the algorithm more adaptable to the evolution of networks; second, particles in PSO can be mapped to nodes in semantic network; the process of finding the optimal solution in PSO is consistent with the PSO uses iterative search to find optimal solutions from random startup solutions [37]. PSO solutions are called ”particles.” Each particle is fit. We design a PSO-based heuristic to detect communities. Each particle shares social information to find the best solution. PSO-LDA incorporates LDA semantics. We map semantic social network nodes to PSO ”particles” and use each node’s semantic information vector to particle velocity. Information similarity replaces fitness value. In PSO, we normalize that semantic social network nodes simulate a ”bird flock,” where individuals gain from the discoveries and previous experience of other nodes during the search for food [38]. Each node, called a particle, in a semantic social network called a swarm ”flies” over the search area looking for promising regions.

2) Honey Badger Optimization (HBO): Honey badgers are extremely important to HBA because of the qualities they possess [19]. In order to locate the source of its food, this one either employs the help of a honey assist bird or relies on its keen sense of smell in conjunction with digging. As a direct consequence of this, the activities that are connected to the honey badger’s search for food can be separated into two distinct groups: the honey mode and the digging mode. When this same badger is in honey mode, it will pay attention to the directions that are provided to it by the honey guide bird. In the latter mode, the badger will encircle the food by using its acute sense of smell, and then it will dig it up to eat it. In either case, the badger will ultimately be successful in eating the prey it pursues. Figure 2 Depicts the Flow Chart of HBO Algorithm.

In this paper, Twitter is used as a mean of collecting viral marketing topics. Also, correlating the amount of tweets that are related to each topic to generate an accurate depiction of the level of interest on a given topic. The Process and Functionality of the Proposed System is Depicted in Figure 3 and Figure 4. The emphasis on the basic task that is essential for building the Model on top of it.

	0	1	2	3	4	5	6	7	8
-PRON-	-0.910577	0.825126	-0.917275	-1.122652	-0.483990	-0.594812	-1.040798	-0.507671	1.489593
the	0.453245	0.366712	1.438942	-0.417440	-0.315771	2.167688	-0.553316	0.779073	-0.102743
of	2.001768	0.781753	2.231088	0.574766	-0.307000	2.653598	0.163683	0.371369	0.799808
and	0.390101	0.233469	1.193282	-0.857497	-0.088257	0.787290	0.072328	-0.521108	1.300832
be	-0.292456	-0.626095	0.588986	0.243185	0.313843	0.092111	0.266125	0.468745	1.051247
to	-0.846508	0.643082	0.806062	-1.439596	-1.229732	0.672711	-1.373244	-0.440372	2.239923
a	0.418980	-0.261030	0.860957	-0.547008	0.076805	2.237270	-1.078582	-0.322116	-0.113774
in	0.925783	0.103659	1.915761	0.256893	-1.104098	1.970993	-0.280158	-0.580547	1.756419
have	-0.885816	-0.035996	-0.115731	-0.625770	-0.231968	-0.426998	-0.253416	-0.883949	0.610476
that	-0.870101	0.546461	1.019510	-1.130501	-0.829928	0.064485	0.226518	-0.079798	0.679610
with	1.054381	0.648423	1.624364	-0.738685	-0.224442	1.700122	-0.534436	-1.594193	2.349406

Figure 4. TF-IDF

A. Topic modeling:

Conventional topic modeling schemes, for example Latent Dirichlet Allocation, are known to perform insufficiently when connected to tweets, because of the sparsity of short documents. Figure 4 Depicts the Topic Modelling being Done in our Procedure. To mitigate these disadvantages, we apply few pooling techniques, aggregating comparative tweets into individual documents, and explicitly study the aggregation of tweets sharing creators or hash tags. The results show that aggregating comparative tweets into unique documents significantly increases topic coherence.

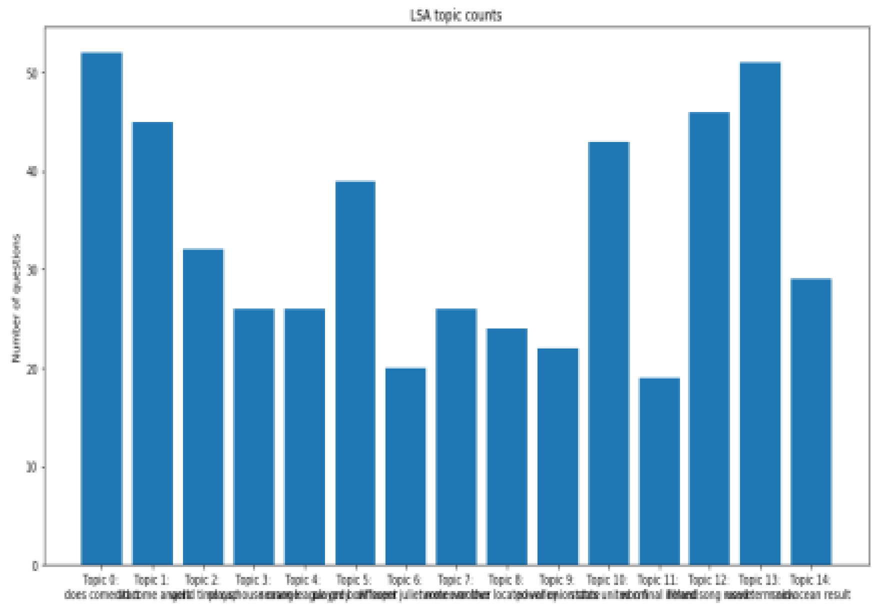

Figure 5. Topic Modelling

B. Sentiment analysis:

Using natural language processing (NLP), case studies, and computational semantics are used to capture and represent and retrieve the sentiments from the data the researcher's interest in finding a document's attitude toward a specific subject. Here, we'll take the opposite approach and search for only the positive and negative words[17]. If it's true, we will discover it amid these words. It is obvious on Twitter when someone's database/ community tweets include a hashtag, word, phrase, or both. A removal of sentiment is referred to as "opinion mining" in the assessment process.

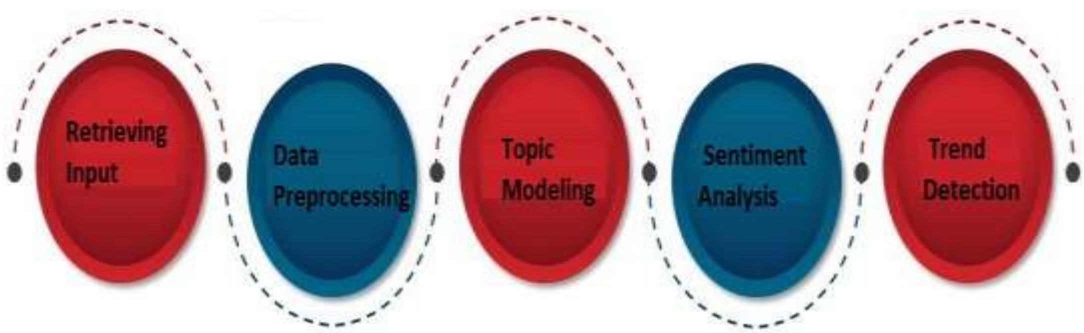

Figure 6. Process Diagram

In the most basic sense, text can be either positive or negative evaluation. It can be done in single words, phrases, and passages, too. The particular opinion must only refer to a single element within the level of analysis (word, phrase, etc.). sentiment is considered a classifier because it tells a content'[19] apart from lexicon-based approaches and language techniques, the machine learning is a means of classifying sentiments Computational analytical techniques determine how people feel about a topic or a text Tweets were evaluated in terms of ecological-based social momentary understanding:

A variety of reports use multiple scores, both positive and negative, throughout. If the basic method is used, a "sentiment score" could remove the bias from this document. thus, argumentation across a single document becomes increasingly difficult the exploration of explicit and implied meaning We should start by learning how to walk before we run this methodology. The Scores of various topics are recorded in Figure 5 and Figure 6.

C. Trend detection:

The ability to describe general situations as well as specific details, as well as the ability to identify patterns newly discovered, cause it makes total sense that we really want to know everything that is about this recent and emerging [upwardly[12]. The information providers should know what is growing in demand so they can take proper steps to meet the needs of that people who might want to use it.

METHODOLOGY

The High level Process Diagram is depicted in Figure 5 and the Functionality is depicted in Figure 6. At this stage, the programme will receive three sources of data: the keywords, the analysis duration of the session, and Tweets[13]. The app must use Twitter for it to provide us with the Twitter information, therefore, needs to validate with Twitter. This should be done as well as the develop a twitter account. For programmers, which will allow them to be included in the developer package, Twitter has four elements: they will have four parameters. The four parameters are the four points to be considered are: the user's key, access token to the customer, and passwords for access. Additionally, more information can be obtained by verifying the twitter page via context. Tweets are obtained from the search function in the site (or the Twitter API, if used in the code) using the term searchTwitter.

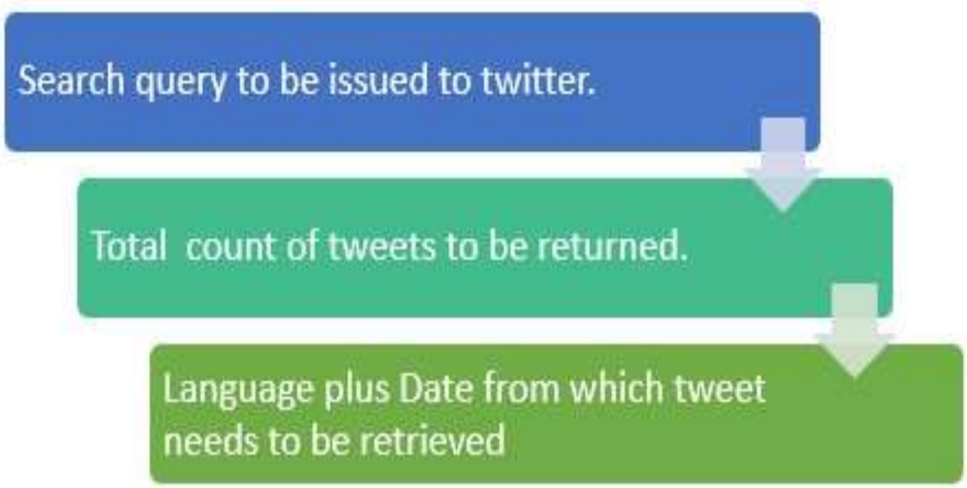

Figure 7. Functionality.

Pre-Processing is something that is done before work on the text data before any analysis is done Without question. Collecting data from a wide variety of sources to begin with is a necessary task to achieve success. You will find the main portion of the information you can glean from the text to be composed of detailed and noisy information. Use it and clean data to obtain higher levels of understanding, or, or design, formulate, and develop, algorithms from it. To cite an example, social media is low in structure; it's an informal in nature; Stop words are common; and post on-specific (users might use URLs and tags); and a clear majority of content is unstructured, consisting of GIFs, audio, and other short videos or pictures, with a wide variety of faces, types such as emojis, photos, and dood hand gestures[15].

The Improvement in the Algorithm is Possible if the Topic Modelling has been done Properly. Figure 10 DShows the Figure 11. Heat Map of Mobile Categorization Heat Map of the Topics and Helps to reduce the Topics that are Repeating. This would help in removing the Redundant Information. Stage 1 also includes going through the steps of cleansing the content, discovering components, and downsizing so that only essential information is emphasised. Other examples[16] of Tweets being ”less structured”signs and language combined with, as well as having a ”unstructured”ty in ”syntax and ” and being used in this way tend to support the belief that they contain more off-the-the-cuff (i. as a non-systematic) content than news and web pages.” the effect of using general-purpose software like content-mining tools will be to further expand the incorrect characters or special symbols that can be removed.

To perform an SDA, you must use the LDA. As clearly as can be seen, LDA analyses all terms and organises them into Themes. LDA can be understood here: mLDA is demonstrated here.The Figure 7 Shows the Comparison study of all the models discussed. mLDA performs better than the other model.

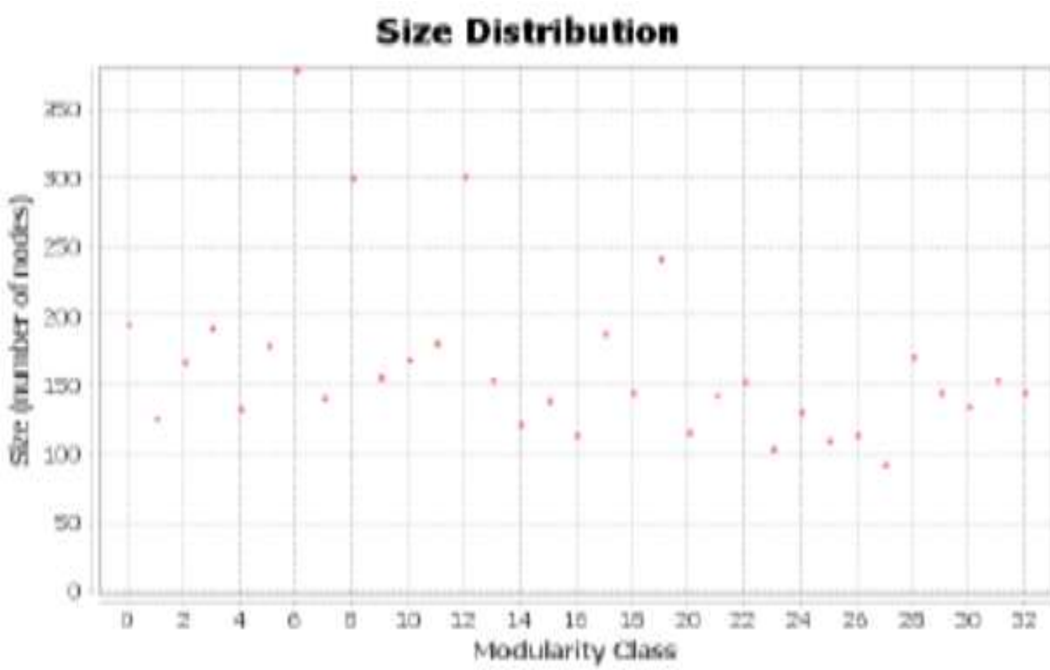

Figure 8. Comparison of Algorithm

Algorithm 1: Sentiment Analysis : HBOmLDA

Input: Twitter data
Result: Ensemble Classification output
Step 1 : Use the twitterR package to get the information - from Twitter
Step 2 : Fetching the information from the storage device to the R environment
Step 3 : Data cleaning eliminates retweets, special characters, re-generated content, emo(c)tions and relative clauses (pronouns)
Step 4 : Use of the 'tm' package to build a Term Matrix document database
Step 5 : Determining the TF-IDF for all the words of the previous Step
Step 6 : Stripping out the words which have less than a 0.1 weight in tf-idf
Step 7 : find a K-log Likelihood score solution to the TDM
Step 8 : Use the 'topicmodels' package to find the relevant topics
Step 9 : Ascertaining the model's implication.
Step 10 : Give the Input to the HBO Algorithm
Step 11 : Form the Clusters based on the Optimized Search Space

RESULTS DISCUSSION

Figure 7 shows the optmized Size distribution of the Nodes of Study based on the Titter data. These nodes have been preprocessed and then analysed for Community. Figure 8 Depicts the Sentiment against the mobile phones captured from the Twitter Data. The proposed Algorithm identifies the Semantics and the Nodes which are capable for effective Marketing. Figure 9 shows the Trend Analytics. Figure 10 shows the Heat map of the data being displayed. The longterm goal is to build a model which will be able to predict mobile phone market shifts. And we predicted phone trends correctly, the system came up with precise results and was able to identify the Seed nodes.

Figure 8 Depicts the Sentiment of the Mobile Phones. The purpose of the role is to serve as an interface between end users and system functionality, receiving inputs from both the end users and the system and integrating both of them[14]. Figure 9 Depicts the Trends of all the Mobile Phone in Twitter

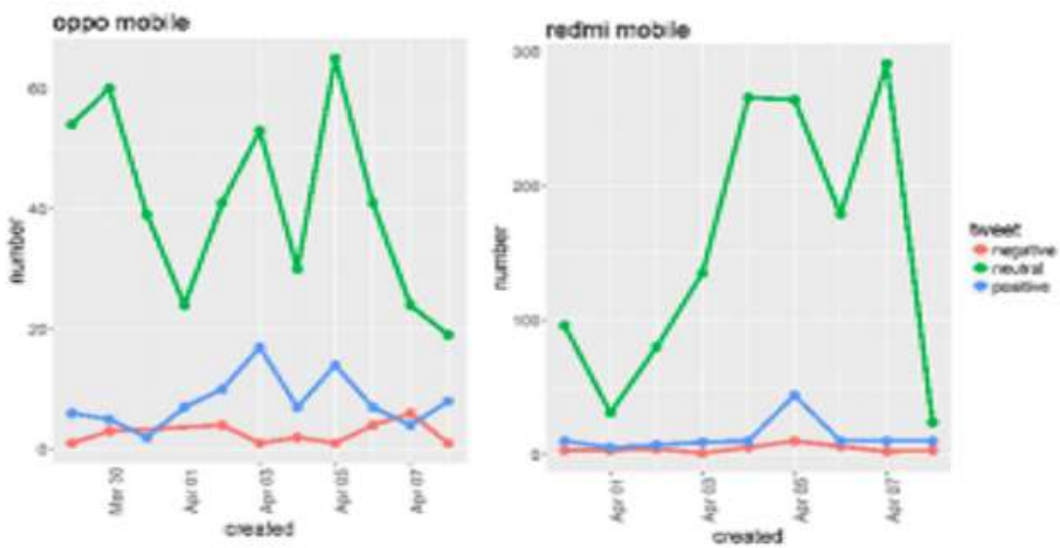

Figure 9. Sentiment analysis of oppo and redmi

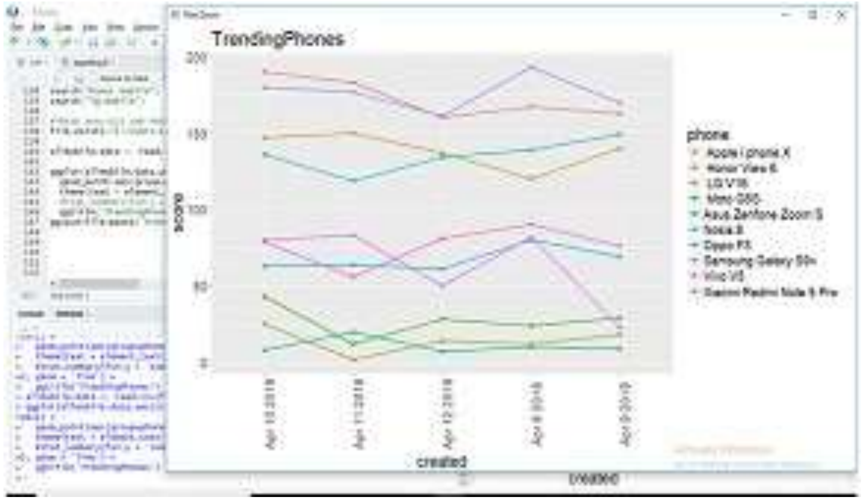

Figure 10. Trend detection

CONCLUSION

This paper is aided by the popular topic modelling approach which incorporates previously recognised concepts as well as all concepts that have been shown to be essential and any concepts that have been shown to be useless. The method applied here is linear discriminant analysis (LDA and mLDA). While there is a lot of interest in this proposed model, it is very unclear if it will support numerous areas of application, and, and if so, what the restrictions and the impacts will be in comparison to standard must/may-linking will be. The experiment findings demonstrate that the values that were chosen for the experimental attribute seem to work as expected. We ran two sets of synonym analyses on the product to see if we could identify functions the same as or variations on the same to identify synonyms. More constraints outperformed current approaches by a large margins, which means that greater previous training can be done with no supervision, allowing unsupervised background modelling to succeed. The paper also presented two new approaches for getting the two requirements that both discovered, as well as recommendations, were intended to simplify the extraction of them.

Figure 11. Heat Map of Mobile Categorization

Experimental results proved that they were of excellent quality We found social media analytics using a newly developed trend-finding algorithm on the concept of "twitter trending". To give it an initial reading, it does a text preprocessing followed by topic categorization and sentiment analysis. It found and reduced the size of the Twitter text files to normal data before going through the various steps in the text processing module. Then the topic modelling module then converts the documents into a set of topics and has analysed their sentiments and have assigned them to assigned to mobile phones according to their sentiment trends and potential neutral and generated the number of positive, negative, and detected the trend but using the Models Like LDA, mLDA and HBOmLDA. The Community Detection accuracy is improved by the Meta - Heuristcs Optimizer.

References

[1] Chakraborty A, Kar A K, "A Review Of Bio-inspired Computing Methods And Potential Applications", Proceedings of the international conference on signal, networks, computing, and systems. Springer India2016, pp 155–161.

[2] Aswani R, Ghrera SP, Chandra S, " A Novel Approach To Outlier Detection Using Modified Grey Wolf Optimization And K-nearest Neighbours Algorithm", Indian J SciTechnol 9(44):1–8 -2016.

[3] Ghanem TF, Elkilani WS, Abdul-Kader HM, " A Hybrid Approach For Efficient Anomaly Detection Using Metaheuristic Methods", J Adv Res 6(4):609–619 -2015.

[4] Aggarwal N, Ahluwalia A, Khurana P, Arora A, "Brand Analysis Famework For Online Marketing: Ranking Web pages And Analyzing Popularity Of Brands On Social Media", SocNetw Anal Min 7(1):21 -2017.

[5] Akcora CG, Carminati B, Ferrari E, Kantarcioglu M," Detecting Anomalies In Social Network Data Consumption", SocNetw Anal Min 4(1):231 -2014.

[6] F. Chen and D. Neill," Non-parametric Scan Statistics For Event Detection And Forecasting In Heterogeneous Social Media Graphs", In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, 1166-1175.

[7] S. Levine, "How Fast The News Spreads Through Social Media", http://blog.sysomos.com/2011/05/02/how-fast-the-news-spreads-through social-media/, 2012.

[8] J. Benhardus and J. Kalita, "Streaming Trend Detection In Twitter", International Journal of Web Based Communities 9, no. 1, 2013, 122139.

[9] J. Lau, N. Collier and T. Baldwin, "On-line Trend Analysis with Topic Models: Twitter Trends Detection Topic Model Online", In COLING, 2012, 1519-1534.

[10] R. Parikh and K. Karlapalem, Et: "Events From Tweets", In Proceedings of the 22nd International Conference on World Wide Web companion, 2013, 613-620.

[11] Guille and C. Favre," Mention-anomaly-based Event Detection And Tracking In Twitter", Advances in Social Networks Analysis and Mining(ASONAM), 2014, 375-382.

[12] Fisher T," ROI In Social Media: A Look At The Arguments", J Database Market Cust Strategy Manag 16(3):189–195 -2009.

[13] Gendreau M, Potvin JY, ”Metaheuristics In Combinatorial Optimization”, Ann Oper Res 140(1):189–213 - 2005.

[14] Goh KY, Heng CS, Lin Z, ”Social Media Brand Community And Consumer Behavior:Quantifying The Relative Impact Of User And Marketer-Generated Content”, InfSystRes .24(1):88–107 -2013.

[15] Song, D., Gao, J., Pang, J., Liao, L., Qin, L. (2020). Knowledge Base Enhanced Topic Modeling. 2020 IEEE International Conference on Knowledge Graph (ICKG).

[16] Lele Yu†,Ce Zhang, Yingxia Shao, Bin Cui,LDA*: A Robust and Largescale Topic Modeling System, Proceedings of the VLDB Endowment, Vol. 10, No. 11, 2017.

[17] Dandan Song; Jingwen Gao; Jinhui Pang; Lejian Liao; Lifei Qin, Knowledge Base Enhanced Topic Modeling, 2020 IEEE International Conference on Knowledge Graph (ICKG), 11 September 2020.

[18] M S Simi, K Sankara Nayaki and M. Sudheep Elayidom, An Extensive Study on Data Anonymization Algorithms Based on K-Anonymity, 2017 IOP Conf. Ser., Mater. Sci. Eng. 225 012279

[19] Annibale Panichella, A Systematic Comparison of search-Based approaches for LDA hyperparameter tuning, Information and Software Technology, Volume 130, February 2021, 106-411

[20] Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton and Toutanova, Kristina , BERT - Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Conference of the North

[21] American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, 2019

[22] S. Kianian, M. R. Khayyambashi, and N. Movahhedinia, “Semantic community detection using label propagation algorithm,” Journal of Information Science, vol. 42, no. 2, pp. 166–178,2016.

[23] H. Abadlia, N. Smairi, and K. Ghedira, “Particle swarm optimization based on island models,” in Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 49-50, 2017.

[24] 31. Sankara Nayaki, Sudeep Elayidom, Rajesh R, A Novel Compression based Community Detection Approach using Hybrid Honey Badger African Vulture optimizer for Online Social Networks, Concurrency Computation: Practice and Experience, 2022 (Accepted)