IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

Intention Mining for Introspective Behavior Modelling in Business Intelligence

Deepali N. Pande[1*], Kaushik R. Roy[2], Satyajit S. Uparkar[3]

[1,2,3]Dept. of Computer Application, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

[*]*Corresponding Author: dnpande30@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i4.10921106 | Available online at: www.ijcseonline.org

Accepted: 19/Apr/2019, Published: 30/Apr/2019

Abstract— Mining user-intents has been a core platform in semantic web search and intelligent test mining. Prior art on these arena lacks infeasibility in materialization of theoretical foundations on factual viewpoints. Literatures and artifacts are needed to twitch conceptualization of formalism and methodologies on distinct domains under user-intent mining. This chapter provides a basis formulation of automata, theories and algorithm design approach for user-intent mining on social networks. The aim and the scope of this chapter is introspection of user's aspiration in online search mechanics symbolizing business intelligence.

A concrete approach to retrieve named entities from live social networks has been modelled in this chapter. The source for retrieval are public logs, blog, social channels and web-o-media has been constructed as an activity model which is modelled as transient in nature. Formulation of automata to recognize intent-keywords and the algorithm to reason the context of dialogue on live –talks have been described. This chapter describes principles and mathematical approach to design ontologies for intelligent mining for reasoning in live-talks overcoming the problem of out-of-vocabulary (OOV).

Keywords—opinion mining, social networks, ontology, live-talk reasoning, out-of-vocabulary (OOV)

I. INTRODUCTION

Information mining has always been a major research goal in data mining. The primitive data mining tasks for feature extraction only pay attention on data. They do not disclose information wrapping a problem [1-3]. In fact, extracting information along with data has been subjectively useful in ubiquitous-data intelligence mining. Moreover, the domain factors consisting of qualitative, quantitative and statistical aspects hide knowledge required for problem-solving [3-5]. This problem occurs due to instantiation of domain constraints and data constraints that thwart exploration of intelligence from featured data. To deal with these empirical research prospects, various techniques for knowledge mining were proposed and experimented [6-20]. But, ubiquitous hidden intelligence in real life constrained-based domains like human constraint, organizational constraint, evaluation constraint, deliverable constraints, deployment constraints and application constraints are still in exploration [21-30].

Intelligence mining in the form of opinions and intentions has been the basic block of predictive business intelligence [31-35]. Myriad researches in the field of opinion and intention mining comply with the mathematical models of ubiquitous hidden intelligence mining but extracting semantic intelligence from constrained-domains has always been a challenging aspect [36-45]. Accuracy in semantic intelligence mining in the form of customer opinions and intentions can be trustworthy and effective in predictive business intelligence. Intuitive Intelligence in this form can be exploited for proactive customer relationship management by using intelligence mining models. Hence, constraint-based mining is an unexplored research area to extract intelligence from real-time model space restricted by constraints [41-45]. Some constraints restrict knowledge exploration but some unnecessarily increase the frequent patterns. This makes intelligence discovery a tedious task because the search space exponentially increases. Contradictorily, a reverse situation occurs in restricted domains wherein the information is hidden in form of constraints. So, all except the most needful information, cannot be extracted. This situation decreases the search space exponentially. The research out-front addresses these challenges in intelligence mining for predictive business intelligence and propose methodologies to diminish the constrained-based mining problem in Social Network (SN) time-series data [41-51]. Owing to this, the following theoretical problems in mining ubiquitous intelligence in knowledge discovery have been identified,

i. Methods for extracting objective and subjective interestingness do not provide strong background and are purely generalized,
ii. Metasynthesis of ubiquitous intelligence in KDD provide techniques only for syntactic and semantic synthesis but lack linguistic connotation,
iii. Scaffold for mining useful and intelligent data from constraint based domains is only hypothesized,
iv. Gap resolution between statistical significance and business expectation is still challenging.

II. State-of-Art Literature Review

W. Ryu, J. Lee, and S. Lee in (1) proposed a methodology to identify the verbal intent in webpage using action perspective with topical intent. The method used a verbal advertisement ranking system. The results of precision improvement lack significant improvement in verbs identification.

R. Santos, P. Vieira, J. Barbosa, S. Carlos, E. Moura, and R. Moura in (2) proposed a novel algorithm using natural language processing and sentiment analysis methods to identify the most important user-feedback to be taken into consideration for decision making. The method proves satisfactory only in case of specific feedback already tagged using parts-of-speech model.

A. Prof and A. Z. Adamov in (3) proposed keyword based relevance approach using natural language processing and sentiment analysis to examine the attitude of the user regarding the context. Accuracy is left as a part of further improvement, though the authors have evaluated the method.
W. Songpan in (4) illustrated the use of naive baye's classifier to classify contrasting comments in context of opposite ratings. The model shows comparatively better results in case of decision tree algorithm.

K. Yang, Y. Cai, D. Huang, J. Li, Z. Zhou, and X. Lei in (5, 67) used a domain sentiment dictionary to classify documents according to their opinion using hybrid model adopted by combining different singleton models. But the method lays its application only to some restrictive sentiments.

E. Zahedi in (6) used a joint sentiment and topic modelling scheme designed using spark framework. The method shows significant speed up only in synthetic datasets. Its speed-up and accuracy is least guaranteed in real-time, timeframe based analysis.
S. M. Singh (9) used an aspect based opinion mining system using reviews generated on mobile phones. The method exhibit sentiment analysis task in determination of user reviews.

S. Mandal and S. Gupta in (10, 65) use a novel algorithm to classify and comprehend the emotional state of the user. The method uses lexicon based approach and calculates polarity levels.

K. Bhattacharjee and L. Petzold in (11) used a neural network model to build a vector space representation of user generated aspect based sentiments. The method provides solution to user generated data having variations for prediction of sentiments accurately.

V. S. Rajput in (12) presented a novel algorithm using firefly technique to optimize the results of structuring the reviews of the user has been successfully shown. The technique retrieves only correct and relevant reviews, restructures it and postulate sentiment analysis.

P. K. Mahajan, S. Ingale, K. Khairnar, K. Gaikwad, and G. Aher in (15) use syntax based feature mining to reduce parsing errors in user-attitude analysis. The authors used word alignment model in feature mining to reduce errors in customer reviews cum feedback.

R. Alfrjani, T. Osman, and G. Cosma in (16) use semantic modelling to link opinions from publically linked data sources. Their research perspective illustrates use of Dbpedia ontology in ground facts mining.

J. Desai and J. Majumdar in (17) use naive baye's, logistic regression and sentiword algorithm to classify contextual intentions from the user feedbacks. The method aimed to work over correct recognizes and recover emotions.

M. Y. Raut in (18) used a statistical classifier in training machine learning, to curb the polarity shifting problem using dual polarity shift technique. A dual classification algorithm is used for model complexity in reviews.

H. Li, Q. Peng, and X. Guan in (19) use subjective mining of user data using affinity propagation clustering algorithm. The authors used vector space model to model similarity in text domain.

G. Khodabandelou in (22), implemented method to identify process traces in identification of processes for current user's context. The method is helpful to provide clues to recognize the plausible set of operations which can be performed.

A. Di Sorbo, S. Panichella et.al in (23), implemented a construct to classify the contents of emails according to their purpose using semi-supervised natural language parsing. The method outperforms the traditional machine learning strategies.

K. Park and et.al in (24), proposed a method to improve the performance of searching using user's intention. They developed an intention map to cluster the selective features from the users given search query.

T. Lu and et.al in (25) proposed a novel method for intention extraction using Verb-BiTerm topic model. They used verb clustering using word to vector deep learning tool to collect short text collections. The results exhibit better verb clustering based on mining coherent topics.

C. Chung in (26) exploited the intention verification using functional assertions. The method used a property mining technique to extract the properties in the design view based on constraints.

M. Hamroun and et.al in (27) used semantic modelling techniques to find user intentions and demonstrated that ontologies outperform in semantic pattern mining.

T. Mei and et.al, used capture intention based on video sequences, to cluster singular value decomposition based intention segmentation and learning based classification. The method is applicable in online visual social media mining.

J. Liu and Y. Lin in (29) used random walk model to cluster the relevant query log. The aim of the method is to consider the initial query into consideration in design of a set of path-edges to reach the destination goal node as the final intention.

N. Ideas and et.al in (30) worked over the documentation goal of intention sequencing. They formalized two levels of abstraction on intention mining perspective, one using source code and the other using statistical machine translation perspective.

III. System Architecture

The conceptual diagram of illustration of facts and findings for the design of Intention-keyword mining and ontology design has been canvassed and pictorized below in figure-1. The User Intention mining engine is a factual diagram of intention keyword extraction from SN web crawls. The process flow diagram shows the steps involved in intention word fetching. The system architecture shows how the query system understands intelligent keywords which can reason a user's wishes regarding search. The knowledge base (KB) stores intelligent phrases perusal for the user's intent of search.

The KB helps in reducing the transient time in putting up the information alongwith the search query, while the user is confused about what exactly he/she wants to search or find.
The KB helps like a auto-complete feature in taking recommendations about any social network activity.

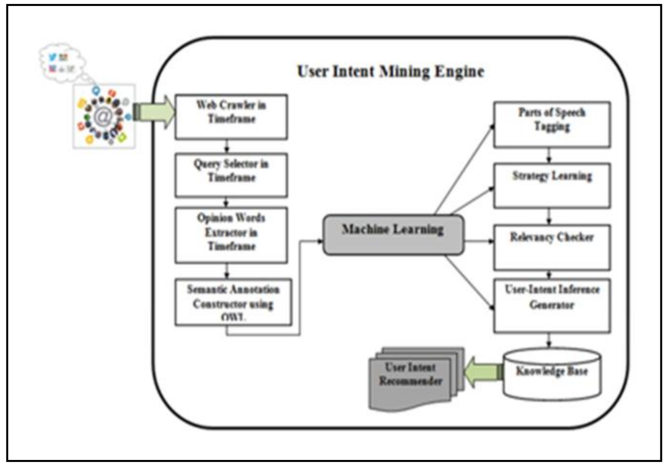

Figure 1. User-Intent Mining Engine

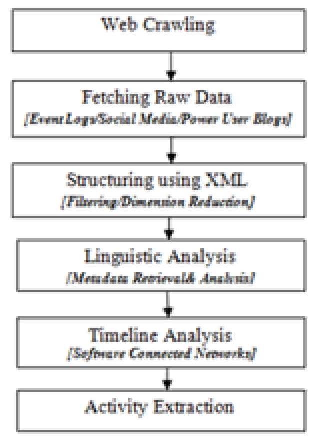

Figure 2. Process Flow Diagram for Fetching User-Intent Vocabulary

IV. Brief Description of the Drawings

The disclosed methods and systems are directed to solve one or more problems set forth above and proposed as a set of objectives in the introduction section.

FIGURE. 1 illustrates a prototypical environment integrating certain embodiments of the present theoretical formulation. It illustrate the facts and origin of social media vocabulary origin and its usage in assembling bag-of-words for social network dictionary design as an auxiliary support to conquer out-of-vocabulary (OOV) problem.

FIGURE. 2 illustrates a definitive computing system consistent with the disclosed epithet of the proposed theoretical formulation. It describes the flow of processes in reasoning activity-words and live-talk activity dialogues from communications on social networks.

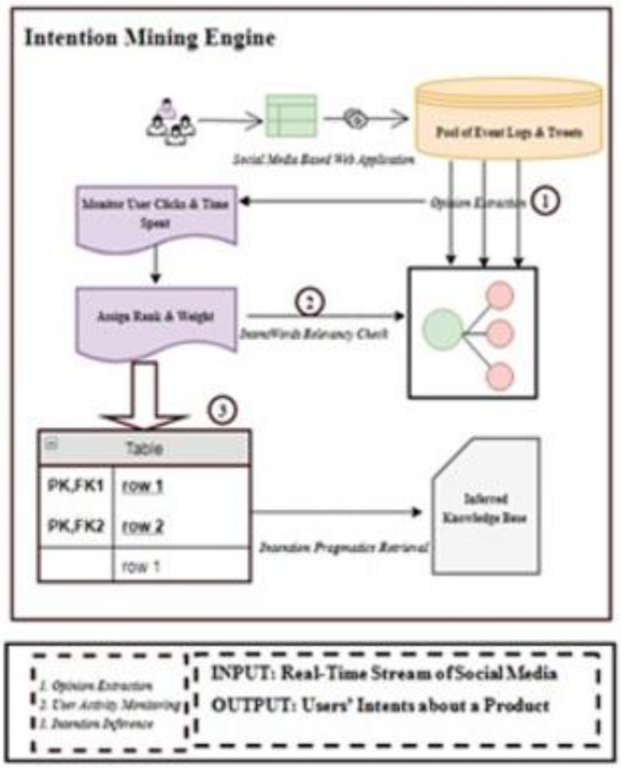

Figure 3. Process Flow Diagram for Fetching User-Intent Vocabulary

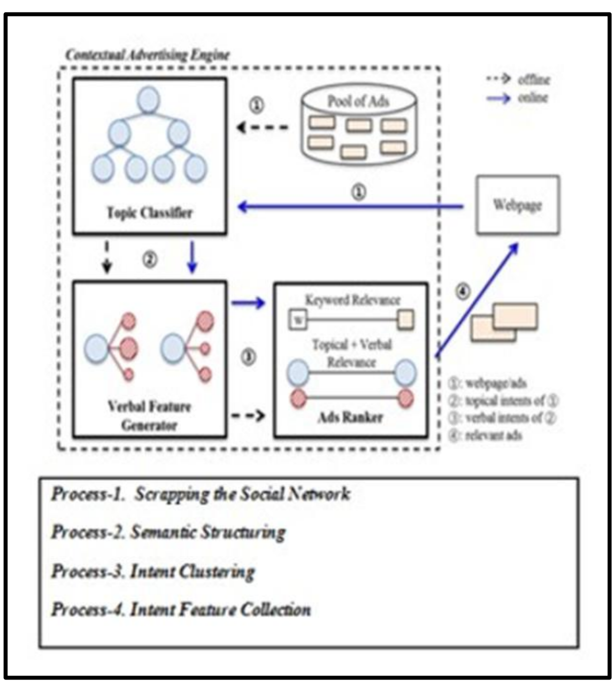

Figure 4. Process Flow Diagram for Out-of-Vocabulary (OOV) Bag-of-Words

FIGURE. 3 illustrates an exemplary query process and context identifying system consistent with the disclosed epithet of the proposed theoretical formulation. It describes the flow of processes in retrieving and sematic framing of user-activity-words and live-talk activity dialogues from communications on social networks.

FIGURE. 4. It illustrate the vocabulary used in social networks is on-grammatical, lacks sematic, so reasoning of user-dialogues over a particular context is major issue. This illustration shows fetching of activities and clustering of out-of-vocabulary words (OOV).

V. Activity Goal Mining

The goal of mining for a given set of activities are to find Explicit Activity pages and Implicit activity pages and extract object-usage information fromt them. One way of accomplishing this would be to search for these pages by a search engine, download the pages and obtain information of objects from them using natural language processing algorithm, hypothesized by Van der Alast in [20], [30]. However, it will not be feasible for us since downloading th e-pages could take hours or even days for a single activity.Therefore, an algorithm that extracts the desired information in a real-time without downloading the web pages is needed. A set of web search engines (e.g., Google and Bing) already have downloaded the pages and stored all the information on their server. The mining will be very fast if we can dig out the desired data from their server. Fortunately, almost all the search engines (SEs) provide special mechanisms for querying the required information. For example, Table 1 provides three query modifiers and operators that can be used along with queries. Our objective is to use these to get the desired information.

An activity is a task, which once completed leads to a full or partial accomplishment of the goal it is related to. The activity is the building block of a process. A process consists of a collection of activities represented by several activity complex relationships like sequence, parallelism and mutual exclusion. The process instances are different executions of the same process. A process model is the representation of a process. Process mining is the discipline aiming at automatically discovering process models from process traces. In real life web streaming, the process traces are logs extracted from the information systems where processes are executed. A log recording the execution of activities in form of business process provide logs such a ERP, WFM, CRM, SCM and B2B recording transactions in a systematic way. The event logs are of the form {estart, A, eend}.

Definition: *An intention is a goal that can be achieved by the performance of an activity e.g. automated, semi-automated or manual. A strategy is an approach and a manner to achieve the intention. [1-5].*

VI. Formal Assumptions

Activity Mining consists of a suite of techniques and tools which has been effectively used to discover and visually represent complex process models from logged traces of information system user and web users. [1], [2], [3]. An activity model can be formalized by a quadruple, A = {S_{Start}, $S_{stop,}$ P, T} where each element is

i. P = $\{a\}_{n=1}^{N}$ is the finite set of activities,
ii. $S_{Start} \neq \varphi$ the set of initial activity states,
iii. $S_{stop} \neq \varphi$ the set of final activity states,
iv. T ⊆ ((S_{Start} X P) ∪ (P X P) ∪ (A X S_{stop})) the finite set of transitions between activities or states.

A process execution E = Sstart < .....< ai <...Sstop represents a sequence in the process from one start state to one end state with at least one activity in between. In most of the software application now-a-days, either web based or desktop based events are generated and logged during their usage. Process mining exploits these events for revealing the process models in a bottom-up approach. Let 'e' be such an event captured during the user iteraction with an application. The minimum information about an event required by a process mining technique is: eid the event's unique identifier, et the event's timestamp, ea ϵ A the event's associated activity. The activity is usually inferred and attached to the event in a pre-processing phase, preceding the mining phase. This happens by reasoning on the existing event's meta-data.
Example of these activities include User Registration, Restaurant Search. A specific process execution E by a User u ϵ U is logged as a trace T = {e1, e2,.....,er}. Let L = {tv }Vv=1 be a log containing a finite set of traces. The process mining technique is a function ϒ that maps the log 'L ' on a process model 'P'. The most challenging part to which a process discovery should deal with, is structuring spaghetti processes. These type of processes are highly unstructured and unreadable. To ensure structured process mining, the function ϒ is sequenced using,

Sequence: t ϵ T: t = $a_i < a_j$ or t = $S_{start} < a_k$ or t = $a_l < S_{stop}$ with $a_i < a_j$.

Step-1) Web Log File Mining

One of the most promising sources is the World Wide Web. Numerous web pages exist on the web that explain how to do activities of daily livings. Each of these pages provide details about activities such as what objects to use, how to use them, and in what sequence. Researchers have been working to mine this data to train activity models. Perkowitz et.al introduced the notion of mining generic activity models from the web. They showed that is possible to convert web data into activity models that can be used in conjunction with RFID tags to recognize activity. Wyatt et.al improved the system by introducing a model that includes idiosyncracies of the environment in which it will be deployed. Hu and Yang showed how to use web knowledge as a bridge to help link different activity label spaces in transfer learning for activity recognition. The hidden markov model are most suitable for web usage mining in which the output of a state depends on the sequence of object usage at a given time. Since the model exploits the sequential pattern of object usage, it is not suitable to train with the real-world data in a conventional way, as the complexity to determine the object usage sequence probabilities will be very high. However, this will not be the same for web activity data since we propose that the model be trained while the system is online and recognizing activity.

Step-2) Ensembling the Activity Logs

Consider a user u that visits the social media site and navigates to another activity of his interest through a click event. This click event is e^c is generated to capture the recommendation by the user. During this process, the recommended impression event e^i is generated and shown in the browser with meta-data <U_{id} ; art_{id} ; art_{cat} = $\{c_m\}^M{}_{m=1}$; timestamp θ> where in the following representations hold true,

- User-id :Uid
- article-id :artid
- categories associated with the article : artcat = {cm}Mm=1
- timestamp θ

In order to generate the event logs necessary for process mining, the following mappings are needful,

i. The Session id is associated with the unique identifier of the process execution traces,
ii. For each impression event, the categories of the read article are sorted, concatenated by commas, and associated with the activity of one trace event,
iii. For each impression event, its timestamp is associated with the timestamp of one trace event,
iv. For each impression event, a unique identifier is generated during the pre-processing and associated with the id of one trace event.

The process activities are mapped with the event log files and categorized in the form of processed log events.Process mining also describes temporal aspects of the process flows such as: how much time an activity lasts on an average, what is the average minimum and average maximum time between two activities. The primary task is extraction of temporal metrics.

VII. Formulation of Intentional Process Model

Ghazaleh et. al.defines Intention Mining as extracting sequences of actor's activities from sets of event logs to infer intentions of related actor. A set of activities corresponding to the achievement of an intention, is termed as Intentional Process modelling. Intention mining tackles the four challenges of discovery, conformance, enhancement and recommendation.

VIII.Prior-Art-Algorithms for Intention-Process Mining

Ghazaleh et.al proposed algorithms based on probabilistic and statistical techniques to formulate intentional processes. The probabilistic models provide the information about the nature of data. The conversant data fitting into probabilitic models is harvested taking into account the temporal aspect. Forney G.D. proposed the Viterbi Algorithm which finds the most likely sequence of hidden states termed as Viterbi paths., for a given observed sequence using trellis, a type of finite state machine with states and transitions. The HMMs enable to find he hiddden structure by estimating the parameters of observaion sequences. The Verbi uses evaluation metrics to find which path is the most likely one. The resultant hidden states are the intentions of the actors that generate obsevations. Baum L.E et al, proposed the Baum-Welch algorithm which allows estimating the optimal solution accurately. It is an optimization over known observation sequences so that the frequent data can be directly counted from the priori information.

IX. Algorithm to Mine Intention-Strategy Keywords

To collect a clear denomination of intentional processes, an intention map has to be designed. A strategy is defined by the triplet {s, A, e}, where s is the starting activity, A is the set of activities in between and e is the ending activity. The following rules realize an activity into strategy,

i) A bottleneck activity a occurs with a probability of 1, with atleast two consequent activities preceding it.

ii) A sequence of activity such that the probability of having this sequence is higher than a threshold β. This rule ensures that the activities composing a strategy have to be frequently used sequentially.

Formal Assumptions

For each strategy {s, A, e}

If max P(A, e|s) < β then

Split Strategy

X. Intention Inference from Strategies

To hold a significant factor of confidence and support measure, a human reasoning and inference factor is needed to link the strategy paths from the group strategies beseiged in an algorithmic way. An ontology containing strategy paths has to be chained to automate the strategy inference system. Owing to this fact in intention mining, we describe the methodology and formalization of algorithms to automate strategies, define intentions and name them using ontologies in this article.

XI. Web Activity Data Mining Algorithm

Their model consists of a sequence of states and is based on a particle filter implementation of Bayesian reasoning. The model extractor works as follows,

(i) Select the set of website which describes activities and retrieve its HTML structure.

(ii) Search each of the page for the activity direction and extract the direction.

(iii) Set the label of an activity as the title of the direction.

(iv) Extract the object phrases from the direction.

(v) Remove the phrases without noun and directed verbs.

(vi) Determine the object-usage probability as a Google Conditional Probability (GCP)

$$GCP(O_i) = \frac{hitcount(object\ activity)}{hitcount(activity)}$$

Where hit count(x, y) is the number of pages Google returns if we search with x and y.

(vii) Finally select only the objects embedded with RFID tags from the phrases.

Wyatt et al proposed an unsupervised activity recognition system using web activity data mining. They developed a document genre classifier that identifies the pages describing an activity and an object identifier that extracts all the objects from a page and calculates the weight of an object.

(i) It queries Google with the activity name along with "how to" as the discriminating phrase. Google returns the number of pages it has indexed in its server.

(ii) It then retrieves P pages as the top z pages within the total pages returned by Google. They did not define the optimal value of z. The efficiency of mining clearly depended on z, with a larger value of z meaning more efficient mining.

(iii) The algorithm uses the genre classifier to determine Ṕ, a subset of P, as the activity pages.

(iv) Using the object identification technique, for each page p in Ṕ, it extracts all the objects mentioned in the page and calculates their weights Ẃ.
(v) Finally, the algorithm calculates the objects usage probabilities as,
P (objects | activity) = $\frac{1}{|Ṕ|} = \sum_p w_{object,p}$

From the mined information a Hidden Markov Model can be best used to assemble the three traditional parameters namely, prior probabilities for each state $\prod$ which were set to be uniformly distributed, the transition matrix T of constant probabilities and observation matrix B where $B_{ji} = p(object_i \,|activity_j)$.
Let A = {$a_{1,}a_{2,}a_{3,}$ ......, a_m} be the set of activities to monitor, let O = {$o_{1,}$ $o_{2,}$ $o_{3...}$ o_n} be the set of objects in the environment, where m and l are the total number of activities and the objects respectively. Let θ = {$\theta_{1,}$ $\theta_{2,....,}\theta_n$} ϵ O be the set of object-usage sequences at any given time, where n is the total number of object usage. The goal is to map the observation sequence θ into the predefined activity labels. For each pair of object usages, the activity recognizer checks if it already knows the probability which it determines from the web with the help of mining engine if it does not already know. It uses the probabilities to recognize the current activity. The activity model is based on Hidden Markov Model. Each of the states is an activity and the observation probabilities are the sequence of object usages.

The observed variable at time t depends on the state at time t. The goal is to find the joint probability distribution,
$P\ (a,\ \theta) = \prod_{t=1}^{T} P\ (a^t|a^{t-1})P(\theta^t|a^t)$ *where* $P\ (a^{t-1}|a^t)$ is the transition probability from state
a^{t-1} to a^t. If we train an AR system using real-world activity data, we count the number of occurrences of transitions, observations and states to find the probabilities that maximize the joint probability. However, as we consider the web activity data to train the system, there is no way we can count the transitions because the transitions between activities are highly subject dependent. Therefore, the transition probability matrix T is set to constant probabilities. For T, the duration of any activity is set to ϒ and all the self-transition probabilities, T_{jj} are set to 1-1/ϒ. The remaining probability mass is uniformly distributed over all transitions to other activities.
During the inference, the Viterbi algorithm is used to find the most likely labels for the new observation sequences. This algorithm has been successfully applied with HMM to solve many activity recogniton problems.

XII. Web Activity Data Mining

As described above, to train the system we need to know two types of probabilities: the probability of using an object given an activity that is P($\theta_1^t \,|\, a^t$) and the probability of using an object given another object and an activity that is P($\theta_k^t|\theta_{k-1}^t, a^t$). The purpose of web activity data mining is to determine these probabilities. According to this requirement, the web activity pages are categorized as two types, namely explicit activity page and the implicit activity page.

Definition 1. (Explicit Activity Page): A web page is called an explicit activity page if it provides detailed instructions about performing an activity. It has a title, which in most cases contains the activity name. It has a text section that provides details of an activity such as what the objects to use and their sequence.

Definition 2. (Implicit Activity Page): A web page is called an implicit activity page if it does not provide explicit description of the activity but instead provide information that is implicitly related to an activity.
Digging out activities and object usage sequences requires the following steps to be accomplished in mining the day to day data. For each activity a_i in A, the following tasks are done. The local database is checked for the number of web activity pages, WAP_i, indexed by the search engine that either explicitly or implicitly describes the activity a_i. It determines this from the web with the query intitle "a_i" if it is not locally available. It then determines the number of pages that mentioned the first object θ_0. We denote this as $POS_{i1.}$ Finally, for each sequence of object-usage pairs θ_{ik} and $\theta_{i\ (k+1)}$ in θ, it searches and stores the number of pages indexed by the search engine, POS_{i1}, using the query intitle: $a_i = \theta_k * \theta_{k+1.}$

XIII. Number of Queries Required for Mining

The Algorithmic approach to do this, is as follows, the number of queries needed 'r' depends on two factors,
(i) the set of activites the system is recognizing 'm',

(ii) the total number of objects used |θ| for an activity. For each of the activities it requires |θ| number of queries. Therefore, we can write,

$$r = m * |\theta|.$$

This formulation of 'r' ensures to counter total number of queries in account.

Input: Activity List A, Object List θ

Output: Web Activity Page Matrix m X1;
Pages with Object-usage Sequences
(POS): m X n matrix

(i) Fetch Web Activity Pages using in title: a_i in the search engine.

(ii) Web Activity Pages will behold the number of pages indexed by the search engine for the given query.
(iii) The Pages with Object-usage Sequences (POS) can be fetched using the intitle: a_i: $\theta_k * \theta_{k+1.}$
(iv) Defining the observed vaiables into the log file to fetch IP address, date and time of access, URL address, etc.
(v) Data Matrix is created from the information in log file & web contents.

Event Summary	User Id	Temporal Features	IP Address	Event Type	Event ID	Task ID
Asset Porting	101	13:00:00 01-01-2018.	192.X.X.X.	Asset Registration	Asset Selection	111111
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Figure 4. Activity Scenario Mining from Live-talks

An event monitor to crawl through social networks (SN) has to be set up to record the key features like Event Summary, User-ID, Temporal features like time and date, IP Address, Event Type and Event ID. In accordance to development perspective, enterprise java beans are used to capture session activities in real time streaming. Multiple aspects of an event using temporal patterns from event log sequence can be generated using event-narration maps.

XIV. IDENTIFYING TYPE OF INTENTION

Intention can be grouped into two categories based on the way of expression of a user's context. When the user directly expresses his requirement mentioning the subjective information and his need, then the intention is tagged as explicit user intention. But when the user just mentions the problem without expressing his need, then the intention is tagged as implicit user intention. The implicit intentions usually lack the need of the user and therefore intention inference algorithms are strongly needful at this stage of intention mining. This is described using Case-1 and Case-2. The example in case-1 shows that the user requires "a book for learning development of machine learning softwares". The need of the user is expressed directly in this post. Hence, this context represents explicit intention, whereas the example in case-2, describes that the user suffers from cold and flu so, the intention

Inference system should recommend medicines to heal cold and posts of companion users to protect flu spread.
Case -1) Explicit User Intention: I am searching a book for "How to develop algorithms based on machine learning".
Case-2) Implicit User Intention: I am suffering from cold and fever.

(vi) Activity map is generated after identifying sessions/agents/path.

(vii) Behaviour patterns of users will be analyzed from the web log file.

(viii) Activity map for articulation of behaviour rules and measures of quality is designed, subjectively in every session.

XV. INTENTION WORD EXTRACTION

Step-1) A sentence 'S' is extracted from User activity monitoring log from social network tagged with micro blogs and hash tags. The Sentence 'S' denotes contextual expressions of a user based on target social media post.
Step-2) The micro blogs and hash tags are pre-grouped based on user-interest genre.
Step-3) Each sentence is tagged with a significant probability weight.
Step-4) A training sequence for all words in the sentence S = $\{w_{1,} w_2, w_{3...}w_n\}$ is constructed.
Step-5) A supervised learning procedure is chosen to classify intention words from the given training sequence 'S' and its classification probability is obtained.
Step-6) A dictionary $D_{User-Intention}$ is the union of all intention words tagged by user 'U' are ensemble.
Step-7) For each word in the Sentence S, a domain score is computed based on grouping all genres in the micro-blogging and hash-tagging social network.
Step-8) Each word is ranked according to the domain score.
Step-9) The word with the highest domain score is ensemble into the dictionary $D_{User-Intention}$ indexed based on user-ids.

XVI. IDENTIFYING THE INTENTION PHRASES

An intention phrase shows inclination towards some object or an associated activity. An intention comment has the syntax,
<Intention Subject + Intention Verb + Intention Object>
In a certain post, the following comment can be intention comment,
"Hello, we ***require software*** ***to clean*** hyperactive internet advertisement and fake webpage crawlers. Can you tell us, which software can protect our computers from these web instincts? We ***have tried some free malware-removal softwares like*** v-pro, zetero and key-cleaner ***but the problem still exist***."
The above post is related with malware removal softwares and the users are requesting suggestions for it. The attributes of the intention in the above post sent by the user can be "require software", "to clean", "can protect", "free malware-removal softwares", "problem still exist".
The fetched comments from the posts are pre-processed by passing each sentence in the comment through the Natural

Language Processing pipeline, in the forms sentence-> token ->Parts-of Speech tagging <noun; verb; adverb>. The corpus generated in this way, is labelled and matched against the specified intention phrase in the real time stream posts along with their frequencies of occurrence.

XVII. INTENTION ATTRIBUTE EXTRACTORS

Once the option statements are extracted from the posts, the next step is to extract all relevant information attributes to ensemble intention attributes. A list of information extractors is generated to be used as a look up table holding labels indexed based on genre of posts. A rule set can be designed to extract intention word in form of tokens. A regular expression in the context of intention words have to be set by framing the conditional part in the surrounding context of intention phrases.

Out of a huge volume of java classes, Scala is one library of java which provides 'Regex' class to set regular expressions. In the context of intention phrases extraction, the Regex class can be set in the following manner,

***Class Regex* {**

1. def findAllIn (text: String): Iterator = iterator over the sequence of all matches in
   User's Post

}

For example, the following regular expression contains three categories of genres related to users post in social media,

2. val expPattern = """**(purchase)**\s* **(sell)**\s***(suggest)**""".r

Given a user post that matches the pattern, the substrings that match the genres can be extracted and be declared as intention words.

3. def eval (exp: String) = {
   val expPattern (str) = exp
4. op match {
   case "purchase" => str.Hastag_ExplictIntention#01
   case "sell" => str.Hastag_ExplictIntention#02
   case "suggest" => str.Hastag_ExplictIntention#03
   }
   }

One of the most important things to be worried about intention mining from social media posts is that the micro-blog posts often lack proper grammatical construction. Also, some of the contexts relate sarcasm and over exaggeration which creates a reasoning problem, also in case of human reasoning.

XVIII. EXTRACTING USER INTENTION

In this study, a key-word based representation is described for a cluster of user's search logs data which includes the user intention. The extracted words reflect user's intention. To extract the key-words of a word, there are two methods. One is to extract key-words of a cluster by considering weights on candidate which contains the complaint data, while the other is to extract key-words showing significant difference among clusters. But, considering the aspects of practical implementation, the algorithm should be designed which considers both methods for performance improvement. The term-frequency-inverse document frequency is the most prominent method for information retrieval and text mining. But, it can be used effectively on domains like intention mining from social media, if the weight factor is defined as the as per the frequency of appearance of the intention word in the social media corpus as well as the threshold offset. The outcomes of learning the user's strategy and intention phrase mining vary if the statistical relationship between weighting and ranking the phrases is unabridged. It requires a designing a suitable weighting function and a ranking function to beget the output of as per the relevance of the search query fired.

User intention phrase can be extracted using TF-IDF as follows,

Step-1) Store the query fired by the user,
Step-2) Find the genre of the need of the user from the query,
Step-3) Cluster the users based on the genre need,
Step-4) Compute the statistics of occurrence of an intention phrase,
Step-5) Calculate the Score Indicator for each phrase in the corpus by multiplying the three metrics, namely, the 'f' frequency of occurrence of the intention words in user corpus, TF(t) is the frequency of the user cluster, and IDF(t) is the inverse of the intention occurrence cluster.
TF (t) = (Number of times a term t appears in the user cluster)/ (Total number of terms in the cluster)
IDF (t) = Log (Total number of clusters/Number of clusters with term't' in it)
Step-6) Assign Weight to each intention phrase,
Step-7) Allocate Rank to each intention word in the bag-of-intention words.
Step-8) Finally, the user intention can be classified using the tf-idf weight allocation to intention feature of each query.

XIX. INTENTION MAP REPRESENTATION

The search intention extracted by the user relating to the query is represented in the form of intention map. A schema of the user intentions retrieved is a set of interrelations between concepts and their features.

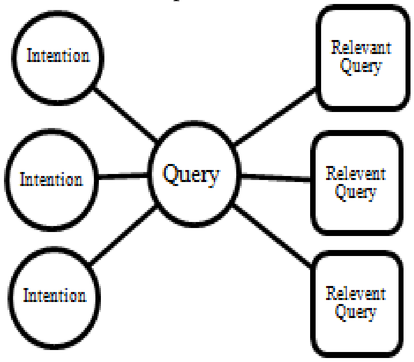

Figure 5. Intention Map Representation [Gazaleh et.al]

In Figure-3 a relevant key-word is added to the intention map in order to extend relevance relations with a search query by using the following rules,

i. Collaborative Filtering: Analyze a relationship between search log and a query and extract a query with high relevance to the search log.
ii. Query Sequence: Extract a query with high input frequency before and after query input.
iii. Query Clustering: Extract a query with high input frequency among similar queries.
iv.

XX. Validation Process

The effectiveness of the intention map is important for its usefulness and applicability. In the validation and evaluation process of an intention map, some of the tests have to be done on the following factors,

i. Evaluate the expressiveness of the intention map,
ii. Verify the effectiveness of the intention map.

The Likert scale is a behaviour measurement technique designed by R. Likert. It works by summing the rating scales and combining the total scores. It computes an internal sequence scale about the similar content and finds the consistency scale using item analysis. Tweets2011 is a collection which includes approximately 16 million tweets between January 23rd and February 8th 2011. It is a real-world data collection and has standard short text collection. Since, it is benchmark dataset; it can be used as a pair with intention corpus under test.

In a corpus of 'n' intention phrases, to test its expressiveness, and draw conclusions about its impact, a paired't-test' is used. Suppose x = test score before the module, y = test score after the module, a null hypothesis proving that the true mean difference is zero, the 't-test' can be performed as follows,

1. A difference Di = yi – xi between two intention phrase observation record is computed. It is important to find the positive and negative differences before testing an observation record.
2. A mean difference Ď is to compute to estimate the target deviation.
3. A standard deviation of the difference S_d is used to calculate the standard error of the mean difference SE (Ď) = S_d / √n.
4. A t-statistics factor computed by T = Ď/SE (Ď) is computed to test the null hypothesis by finding the the t-distribution with n-1 degrees of freedom.
5. Finally, a look-up table for t-distribution is used to compare the value under test for T to the tn-1 distribution which gives a p-value for the paired t-test.

XXI. Intention Mining Algorithm

Self-Organising maps, Bi-Term Topic Modelling, Latent Dirichlet Analysis, and Probabilistic Latent Similarity Analysis.

Self-Organizing maps operate in a two-step way, namely training phase and a mapping phase. The training phase constructs a map using input samples and the mapping phase classifies a new input vector as per the subjective samples in the training phase. The training phase works on the basis of vector quantization. The self-organizing maps aim to make different parts of the network to respond in a similar way to a certain input pattern. It prepares the training sample by vector quantization method described below,

1. A sample vector has to be randomly chosen,
2. Use simulated annealing procedure to converge the sample vector towards the centroid through quantization vector as the source iterator.
3. Loop until all samples are covered.

XXII. Parameters Necessary to Tune the Dimensionality Reduction Algorithm

Phrasal word clustering is in fact akin to topic clustering. The traditional algorithms like Latent semantic analysis and Probabilistic latent semantic analysis use dimension reduction techniques which consume time and memory during the process. Also, the latent semantic analysis method, which is a three step method namely, token parsing, word frequency matrix construction, tf-df computation, outperform if the parameters used in dimension reduction equation are tuned with the information vectors as per the estimated threshold of the information retrieval phase. The results of clustering can be fruitful in use of LSA if the parameters like the weighting scheme, selection of clustering method, the function criterion and optimal number of dimensions are tuned during the feature mining phase.

The Probabilistic Latent Similarity Analysis algorithm can also be used in cases where the expectation maximization algorithm is used with iterations in range 150-300 and a voluminous feature vector as input.

XXIII. Discussion and Conclusion

In this literature review and proposed system design of intention mining, a discussion on how a large set of opinions are extracted from social media and intention keywords are derived is visualized. A system architecture for intention keyword ontology mining has been proposed and its theoretical summary is covered. In the first part of the review, the challenges in identifying intention keyword from opinion mining context have been discussed. In the second part, a detail method to mine intention keyword and design its ontology has been discussed. The contextual framework focuses on customer feedback reviews collection, and mining of user reactions to large scale events on microblogs and online blog conversations.

References

[1]. Agrawal R.D, Gunopulos and F. Leymann (1998), "Mining Process Models from Workflow Logs", Sixth International Conference on Extending Database Technology, pp. 469-483.

[2]. H. Yin et al, "A Temporal Context-Aware Model for User Behavior Modelling in Social Media Systems," Proc.2014 ACM SIGMOD Int'l Conf.Management of Data(SIGMOD),2014,pp. 1543-1554.

[3]. A.Ferraz Cotsa et al., "RSC: Mining and Modeling Temporal Activity in Social Media," Proc.21th ACM SICKDD Int'l Conf. knowledge Discovery and Data Mining(SIGKDD), 2015,pp.269-278.

[4]. R.Hinami and S.Satoh, "Audience Behavior Mining by Integratig TV Rating with Multimedia Contents", IEEE Int'l Symp.Multimedia(ISM), 2016, pp.44-51.

[5]. J.Machajdik and A.Hanbury, "Affective image Classification Using Features Inspired by psychology and Art Theory", Proc.18th ACM Int'l Cont.Multimedia(MM),2010,pp,83-92

[6]. A.Krizhevsky,I,Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Network Systems(NIPS)", 2012,2012,PP.1097-1105.

[7]. X.Wu and S.Satoh, "Ultrahigh-Speed TV Commercial Detection, Extraction, and Matching", IEE J.of forecasting, Vol.27, no. 4, 2011, pp.

[8]. N.Katayama et al., "Mining large-Scale Brodcast Video Archives Towards Inter-Video Structuring", Advances in Multimedia Information Processing-PCM 2004,LNSC 3332, Springer,2005,pp489-496.

[9]. C.Rolland, N.Prakash, and A.Benjamen, "A Multi-Model View of Process Modelling," Req.Eng.,Vol.4,N.4.Springer-Verlag,1999.

[10]. W.M.P.Van der Aalst, "Process Mining: Discovery, Conformance and Enhancement of Business Process", Ist Ed.,2011,pp.184-352.

[11]. C.Hug.R.Deneckere, and C. Salinensi, "Map-TBS: Map process enactment traces and analysis", RCIS,2012,pp.204-209.

[12]. L.Maruster, "Machine Learning Approach to Understand Business Process", PhD Thesis, Tech.Univ.Eindhoven,The Nrtherlands,2006.

[13]. C.Rolland, "A Comprensive View of Process Engineerng", CAiSE98, LNCS 1413, Springer, Pisa, Itly, June 1998.

[14]. H.Schonberg, B.Weber and B.V.,Dongen, "Suporting Flexile Process through Recommendation baised on History", Springer, BPM, vol 5240, pp.51-66. 2008

[15]. A.Rozinat, M.Veloso, and W.M.P. van der Aalst, "Evaluating the quality of discovered Process Models", Antewrp, Belgium, 2008, pp 45-52.

[16]. R.Agrawwal, D.Gunopulos, and F.Leymann, "Minng Process Models from Workflow Longs", EDBT,LNCS 1377, pp. 469-483, Springer. 1998.

[17]. J. Herbist and D. Karaginnis, "Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models", Workshop and database and Expert System Applications, 1998.

[18]. G.Khodabandelou, C. Hug, R. Deneckere, and C. Salinei, "Supervised International Process Models Discovery Using Hidden Markov Models", 7th IEEE RCIS conference, Parice, France, 2013.

[19]. J. Cook, and A. wolf, "Automating Process Discovery through Event- Data Analysis," Int. Conf. on soft. Eng, ACM process 1995, pp. 73-82

[20]. F. van der Aalst, A. Weijters, "Generic Process Mining." Application and Theory of Petri Nets, LNCS 3536, 2005.

[21]. W. Van Der Aalst, A. Weijters and L. Maruster, "WorkFlow Mining: Discovering Process Model from Event Logs," IEEE Trans, on Knowledge and Data Eng Vol. 16 N.6, 1996, pp. 47-60.

[22]. J. Laflaquiere, "A Web Service Composition System Based on Fuzzy Dominance for Preference Query," PhD thesis, University de technologies de Troyes, Campus France, 2009

[23]. T.K. Moon, "The Expectation Maximization Algorithm", Signal Process Magazine, IEEE, Vol 12, N.6, 1996, pp. 47-60.

[24]. G. Greco, A. Guzzo, and L. Pontieri, "Mining Hierachies of Models: From abstract Views to Concrete Specifications," BPM, 2005, pp. 32-47.

[25]. I.vanderfesten, H. Reijers, W. van der Aalst, "Product-based workflow support Information Systems," Vol.36, No.2, pp. 517-

[26]. G. Greco, A. Guzzo, and L. Pontieri, "Mining Hierarchies of Models: From Abreact Views to Concrete Specifications," BPM, 2005, pp.32-47.

[27]. Vanderfeesten, H. Reijers, W. van der Aalst, "Product- baised Workflow Support Information Systems." Vol. 36, No.2, pp.517-535, 2011.

[28]. G. Grosz, C.Rolland, S.Schwer, C. Souuveryet, V. Plihon, S. Si-Said, C. Ben Ashour and C. Gnho, "Modelling the Requirements Engineering Process: An Overview of the Nature Approach," RE'97, London, Vol.2,1997.

[29]. J. Cook and A. Wolf, "Discovering Models of Software Process from Event-Based data," ACM Trans. on Soft. Eng. and Methodology, Vol.7, No. 3, 1998, pp.215-249.

[30]. S. Minscok, C.W. Gunther, and W.M.P. Van der Aalst, "Trace Clustering in Process Mining Business Process Management Workshops," LNBIP 2009, Vol.17. pp 109-120.

[31]. F.Fibre and I.H. Witten, "Data Mining: Practical Machine Learning Tools and Techniques", 3rd Ed., Morgan Kaufmann Publishers Inc. San Francisco, USA, 2005.

[32]. Dardenne, A. Van Lamsweerde and S. Fickasu, “Goal-Directed Requirements Acquisition”, Sci. Comput. Program, Vol.77, No.2, 1989, pp.257-286.
[33]. Van Dingen, and W, Van Der Aalst, “Multi- Phase Process Mining: Building Instance Graphs,” Int. Conf. on Conceptual Modelling, LNCS 3288, Springer, 2004, pp. 362-376.
[34]. L.E. Baum, T. Petrie, G. Soules, and N. Wess, “A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains,” AT&T Technical Journal, Vol.64, No.2, 1985, pp.391-408.
[35]. G. Khodabandeleou, C. Hug. R. Deneckere, C. Salinesi, M. Bajec, E. Kornyshova and M. Jankovie, “Cots Products to Trace Method Enactment: Review and Selection”, 21st Ecis, Utrecht, Netherland, 2013.
[36]. E. Yu, “Modelling Strategic Relationship for Process Re-Engineering”, PhD thesis, University of Toronto, Dept of Computer Science, 1995.
[37]. T. Joachims, “Text Categorization with Support Vector Machines”, Technical Report, LS VIII No. 23, University of Dortmund, 1997.
[38]. Event Log Analyser IEEE Conference viewed October 2012.
[39]. J.Aranda, And G. Venolia, “The Secret Life of Bugs: Going Apart the Errors and Omissions in Software Repositories”, in Proceedings of the 31st International Conference on Software Engineering (ICSE), 2009, pp. 298-308.
[40]. G.Antoniol, K. Ayari, M. Di Penta, F. Khomh, Y. Guhneuc, “Is It a Bug or An Enhancement? A Text-Based Approach to Classify Change Requests”. CASCON. 2008:23.
[41]. J. Anvik, L. Hiew, and G.C. Murohy, “Who Should Fix This Bug?” In Proceeding of the 28th International Conference on Software Engineering (ICSE), 2006, pp. 361-370.
[42]. Bacchelli, T. dal sasso, M.D' Ambros, and M. Lanza. “Contest Classification of Development Emails”, In Proceedings of the 34th International Conference on Software Engineering (ICSE), 2012, pp. 375-385.
[43]. C.Vasallo, S. Panichella, M. Di Penta, And G. Canfora, CODES: “Mining Source Code Description From Developers Discussions”, In Proceedings of The 22th IEEE International Conference on Program Comprehension (ICPC), 2014, Pp. 106-109.
[44]. R. A. Baeza- yates, B. Ribiro-Neto, “Modern Information Retrieval”, Addison- Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1999.
[45]. V.R.Basili, L.C. Briand, and W.L. Mlo, “A Validation Of Object Oriented Design Matrices As Quality Indicators”, In IEEE Trans. Software Eng., Vol.22, No. 10, 1996, pp. 751-761.
[46]. Begel and N. Nagappan, Global Software Development. “Who Does It?” In Proceedings of the 2008 IEEE International Conference on Global Software Engineering (ICGSE), 2008, pp. 195-199.
[47]. D. M. Blei, A.Y. Ng, and M.I. Jorden, “Latent Dirichlet Allocation. In Journal of Machine Learning Research (JMLR)”, Vol. 3, 2003, pp. 993-1022.
[48]. G.Confora, M. Di Penta, R. Oliveto, and S. Panchella, “Who Is Going To Mentor Newcomers in Open Source Projects?” In Proceedings of ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering (FSE), 2012:44.
[49]. D. Cer, M.C. de Marneffe, D. Jurafsky, and C.D. Manning, “Parsing to Stanfold Dependencies: Trade-Offs between Speed and Accuracy”, in Proceeding of the 7th International Conference Language Resources and Evaluation (LREC), 2010.
[50]. L. Cerulo, M. Ceccarelli, M. Di Penta, and G. Canfora, “A Hidden Markov Model to Detect Coded Information Islands in Free Text” in Proceeding of 13th IEEE International Working Conference on Source Code Analysis and Manipulation (SCAM), 2014, pp. 157-166.
[51]. E.Ceylan, F. Kutlubay, and A. Bener, “Software Defect Identification Using Machine Learning Techniques”, in Software Engineering and Advanced Applications, 2006. SEAA 06, 32nd EUROMICRO Conference on, 2006, pp. 240-247.
[52]. A. M. Jehad Sarkar. "Hidden Markov Mined Activity Model for Human Activity Recognition" , International Journal of Distributed Sensor Networks, 2014
[53]. hal-paris1.archives-ouvertes.fr
[54]. Kinam Park, Taemin Lee, Soonyoung Jung, Sangyep Nam, Heuiseok Lim. "Extracting Search Intentions from Web Search Logs" , 2010 2nd International Conference on Information Technology Convergence and Services, 2010
[55]. Khodabandelou, Ghazaleh, Charlotte Hug, Rebecca Deneckère, and Camille Salinesi. "Process Mining Versus Intention Mining" , Lecture Notes in Business Information Processing, 2013.
[56]. mlsc.lboro.ac.uk
[57]. Kithmi Ashangani, K. U. Wickramasinghe, D. W. N. De Silva, V. M. Gamwara, Anupiya Nugaliyadde, Yashas Mallawarachchi. “Semantic video search by automatic video annotation using TensorFlow” , 2016 Manufacturing & Industrial Engineering Symposium (MIES), 2016.
[58]. Tingting Lu, Shifeng Hou, Zhenxiang Chen, Lizhen Cui, Lei Zhang. "An Intention-Topic Model Based on Verbs Clustering and Short Texts Topic Mining" , 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015
[59]. Lecture Notes in Computer Science, 2004.
[60]. Submitted to The University of Manchester
[61]. G. Murray, E. Hoque, G. Carenini. “Opinion Summarization and Visualization”, Elsevier BV, 2017.
[62]. Submitted to Universiti Tunku Abdul Rahman
[63]. S. Bansal, “Data Security by Steganography: A Review”, IJSRNC, Vol 7, Issue-1 Mar-2019
[64]. Submitted to University of Johannsburg
[65]. Santanu Mandal, Sumit Gupta, “A novel dictionary-based classification algorithm for opinion mining”, 2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2016
[66]. “Enterprise, Business-Process and Information Systems Modeling", Springer Nature, 2013
[67]. Sirisha Aguru, Batteri Madhavrao, “Data Security in Cloud Computing Using RC6 Encryption and Steganography Algorithms”, IJSRCSE, Vol 7, Issue 1 Feb 2019.
[68]. Kinam Park. "Automatic extraction of user’s search intention from web search logs",
Multimedia Tools and Applications, 01/13/2011.
[69]. export.arxiv.org
[70]. Process Mining, 2011

[71]. Song, B. L., Liao, and J.Lee -. “Intelligent Collaborative Quality Assurance System for Wind Turbine Supply Chain Management”, International Journal of Advanced Computer Science and Applications, 2013.
[72]. file.allitebooks.com
[73]. Ghazaleh Khodabandelou, Charlotte Hug, Rebecca Deneckère, Camille Salinesi. "Chapter- 33 Process Mining versus Intention Mining”, Springer Nature, 2013.
[74]. Tao Mei, Xian-Sheng Hua, He-Qin Zhou, Shipeng Li, “Modeling and Mining of Users Capture Intention for Home Videos”, IEEE Transactions on Multimedia, 2007.
[75]. Andrea Di Sorbo, Sebastiano Panichella, Corrado A. Visaggio, Massimiliano Di Penta, Gerardo Canfora, Harald Gall. “DECA”, Proceedings of the 38th International Conference on Software Engineering Companion - ICSE '16, 2016
[76]. xin-xia.github.io
[77]. “Conceptual Modeling – ER 2005”, Springer Nature, 2005
[78]. www.engineering.wright.edu

AUTHORS’ PROFILE

Deepali Narendra Pande is working as an Assistant Professor in the Department of Computer Application at Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India. She has qualified B.E. in C.T, M. Tech in CSE. Her area of research includes Network Security, Query Optimization in Heterogeneous, Distributed Computing and Parallel Computing.

Kaushik Rampriya Roy is working as an Assistant Professor in Department of Computer Application, at Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India. He has qualified BE in CSE, M.Tech in IT and M.Tech in CSE. His research endeavours include underwater communication and Internet of Things.

Satyajit Uparkar is working as an Assistant Professor in the Department of Computer Application at Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India. He is pursuing his PhD from Rashtrasant Tukodoji Maharaj Nagpur University. He has delivered several guest lectures on eminent areas of computer science and engineering. His research endeavours includes Distributed Computing, Big Data Optimization, Statistics and Operation Research.

APPENDIX
[OUTCOMES: ACTIVITY AND EVENT LOGS]

I. Examples: Event logs

The scheduler service logs information into the application event log and provides an event identification (event ID) number for each event in the log. This topic shows examples of events that are logged to the application event log.

II. Scheduler service

Event 4097 (informational message)

Example 1:

```
Event Type: Information
Event Source: AdsmClientService
Event Category: None
Event ID: 4097
Date: 10/31/2018
Time: 8:29:57 AM
User: DILE\Administrator
Computer: MIKEDILE
Description:
TSM 515 Scheduler halted.
```

III. Copy

Example 2:

```
Event Type: Information
Event Source: AdsmClientService
Event Category: None
Event ID: 4097
Date: 10/31/2018
Time: 8:29:57 AM
User: DILE\Administrator
Computer: MIKEDILE
Description:
Scheduler Terminated, service
ending.
```

IV. Copy

Example 3:

```
Event Type: Information
Event Source: AdsmClientService
Event Category: None
Event ID: 4097
Date: 10/31/2018
Time: 8:29:56 AM
User: DILE\Administrator
Computer: MIKEDILE
Description:
TSM Client Scheduler 'TSM 515
Scheduler'
Started.
```

V. Copy

Example 4:

```
Event Type: Information
Event Source: AdsmClientService
Event Category: None
Event ID: 4097
Date: 10/31/2018
Time: 8:29:56 AM
User: DILE\Administrator
Computer: MIKEDILE
Description:
Starting Scheduler.
```

VI. Copy

Example 5:

```
Event Type: Information
Event Source: AdsmClientService
Event Category: None
Event ID: 4097
Date: 10/30/2018
Time: 8:06:09 PM
User: DILE\Administrator
Computer: MIKEDILE
Description:
Incremental backup of volume
'\\MIKEDILE\C$'
```

VII. Copy

Event 4098 (warning message)

Example 1:

```
Event Type: Warning
Event Source: AdsmClientService
Event Category: None
Event ID: 4098
Date: 10/31/2018
Time: 8:29:56 AM
User: DILE\Administrator
Computer: MIKEDILE
Description:
Error Initializing TSM Api, unable
to verify
Registry Password, see
dsierror.log.
```

VIII. Copy

Example 2:

```
Event Type: Warning
Event Source: AdsmClientService
Event Category: None
Event ID: 4098
Date: 9/20/2018
Time: 6:20:10 PM
User: DILE\Administrator
Computer: MIKEDILE
Description:
ANS1802E Incremental backup of
'\\mikedile\
c$' finished with 3 failure
```

IX. Copy

Event 4099 (error message)

Example 1:

```
Event Type: Error
Event Source: AdsmClientService
Event Category: None
Event ID: 4099
Date: 9/17/2018
Time: 6:53:13 PM
User: DILE\Administrator
Computer: MIKEDILE
Description:
Scheduler exited with a result code
of 4.
```

X. Copy

Example 2:

```
Event Type: Error
Event Source: AdsmClientService
Event Category: None
Event ID: 4099
Date: 9/17/2018
Time: 6:27:19 PM
User: DILE\Administrator
Computer: MIKEDILE
Description:
ANS4987E Error processing
'\\mikedile\e$\
tsm520c\client\winnt\mak
\dsmwin32.ncb':
the object is in use by another
process
```

XI. Copy

Event 4100 (scheduler command message)

```
Event Type: Information
Event Source: AdsmClientService
Event Category: None
Event ID: 4100
Date: 10/31/2018
Time: 8:29:56 AM
User: DILE\Administrator
Computer: MIKEDILE
Description:
Next Scheduled Event Obtained from
Server
SNJEDS1 (MVS):
-----------------------------------
------
Schedule Name: NIGHTLY BACKUP
Action: Incremental
Objects: (none)
Options: (none)
Server Window Start: 19:00:00 on
10/31/2018
```

XII. Copy Event 4101 (backup or archive statistics) Displays backup and archive statistics, which might be useful in determining the success or failure of a command.

```
Event Type: Information
Event Source: AdsmClientService
Event Category: None
Event ID: 4101
Date: 10/30/2018
Time: 8:29:21 PM
User: DILE\Administrator
Computer: MIKEDILE
Description:
Backup/Archive Statistics for
Schedule Backup
NIGHTLY BACKUP :
-----------------------------------
----------
Total number of objects inspected:
158,688
Total number of objects backed up:
2,486
Total number of objects updated: 0
Total number of objects rebound: 0
Total number of objects deleted: 0
Total number of objects expired: 12
Total number of objects failed: 0
Total number of bytes transferred:
1.15 GB
```



```
Data transfer time: 104.35 sec
Network data transfer rate:
11,564.84 KB/sec
Aggregate data transfer rate:
866.99 KB/sec
Objects compressed by: 100%
Elapsed processing time: 00:23:11
```

XIII. Copy Event 4103 (backup-archive client service startup parameters)

```
Event Type: Information
Event Source: AdsmClientService
Event Category: None
Event ID: 4103
Date: 10/31/2018
Time: 8:29:56 AM
User: DILE\Administrator
Computer: MIKEDILE
Description:
Backup/Archive Client Service
Startup
Parameters:
-----------------------------------
---
Service Name : TSM 515 Scheduler
Last Update : Oct 14 2018
Client PTF Level : 5.1.5.2
Service Directory : D:\Program
Files\
Tivoli\TSM515\baclient
Client Options File :
E:\users\mikedile\
logfiles\dsm.opt
Client Node : MIKEDILE
Comm Method : (default or obtained
from
client options file)
Server : (default or obtained from
client
options file)
Port : (default or obtained from
client
 options file)
Schedule Log :
E:\users\mikedile\logfiles\
dsmsched.log
Error Log :
E:\users\mikedile\logfiles\
dsmerror.log
MS Cluster Mode : (default or
obtained
from client options file)
```

XIV. Copy Journal based backup service events

```
4097: Informational message
4098: Warning message
4099: Error message
4100: Journal Based Backup service
file monitor parameters
4101: Journal Based Backup service
database parameters
```

