# Comparative Study on Multi- Modal Topic Modeling with Dense block Detection

**Prajakta Sonone[1], Prof. A.V.Deorankar[2]**

[1]*P.G. Scholar, Department of Computer Science and Engineering, Government college of Engineering, Amravati, Maharashtra, India*

[2]*Associate Professor, Department of Information Technology Government college of Engineering, Amravati, Maharashtra, India*

---

***Abstract—*** There has been incredible growth of events over the internet in recent years. Google has become the giant source of knowledge for any event which has happened or happening over the internet. Some networking sites such as facebook, micro blogging sites such as twitter are evolved with time and became the highly used sites over the internet. Various E-commerce websites such as Amazon, Ebay, Flipkart etc are the widely used sites for online shopping. Above mentioned sites generates large amount of text data. In association with text data some images are also uploaded over the internet on these sites. The images associated with particular topic plays vital role for understanding the semantic relationship between the textual content and visual content which are related to images. To model this huge amount of data having both textual and visual contents multi-modal topic model is suggested in this paper. While dealing with multi-modality, study of semantic relationship between the images and text data is crucial part. This model also helps to study semantic relationship between them effectively. Topics which are trending, popular over the world can be seen on Social sites as well as micro blogging sites. In online shopping sites fake reviews, advertises, spam spreading information is posted. So to study this suspicious behavior dense suspicious block is also detected in this model. In this paper some and methods are presented and compared. The focus is to effectively model the textual and visual contents in multi-modal dataset having dense block for examining lockstep behaviors.

***Keywords—suspicious behavior, topic model, multi-modality, Event tracking, event detection.***

## 1. Introduction

With the immense growth of internet, more amounts of multimedia contents are generated over the internet. The social media sites like Face book, Flickr, Google News are the most popular sites among the users of internet. Micro blogging sites like Twitter and E-commerce webites contains trending topics, retweeting for given hashtag and ads, posts related to particular product respectively. While visiting these websites user may get various posts, various tweets from twitter, news related articles having associated images. So it is difficult for them to analyse and track those events with detailed summary. While modelling images and text data semantic relationship have less focus in previous studies. So Multi-modal event topic model helps to analyse, summarize and track those events and studies semantic relationships effectively. [1]

For studying suspicious behaviours in multimodal data,In case of E-commerce sites, financial activities of “who-trades-what stocks”, EBay’s “who-buys-from-whom” graph can be taken into consideration. Reviews, advertises can be misused for spreading wrong information or post fake advertises. So if we detect the dense block of information in given document then it becomes easier to study the lockstep behaviors because for example, one tweet with multiple retweet or multiple tweets with multiple retweets the difference can be clearly detected in case of suspicious behavior. Any behavioral factors which can spread the ads which are fake and which can spread the spam. So Many commercial products and approaches are attracted towards the behavior analysis of the problem so that this fraud containing information or any spam and advertising with URL hijacking can be detected. For this the use of adjacency matrix is becoming popular. While visualizing and summarizing the available data in the datasets various graphs are taken for analysis. The graphs are the graphical data of adjacency matrix. Detecting dense blocks in the adjacency matrix of graph data, and tensors of multimodal data has becoming popular. No method gives a exact and accurate way to spot the suspiciousness of dense blocks with different numbers of data modes and ranking that data is also difficult job according to people's choice. In this paper, we have defined some assignments related to multimodal topic model[14]. In previous works the modalities of the topic models are studied but in isolation. At the same time providing the results related to semantic relationships between text corpora was also studied but do not gives the perfect results. So the multi-modal data model and dense block detection can help to model and spot suspicious behaviors in datasets.

In this paper in first part the fundamental factors in multi-modal topic model are presented. The study of Existing work has been described in details which will gives scope for research work and finally we presented conclusion and future score for this paper.

## 2.FUNDMENTALS OF Multi-modal Events and Topic modelling

In case multi-modal events two things are very important while studying. First is textual information and second is visual information contained in the particular event which we want to study. So in multi-modal event model text data is non-visual representative topic and images are visual representative topic.

*Word Count:* In a multimedia document of particular event containing long text and related images is depicted as a pair of vectors of word counts which is corresponding to visual and textual contents.

*Image word:* Image word is denoted as a unit-basis vector v of size $D_v$ which contains exactly one non-zero entry for only one word in a dictionary of $D_v$ words.

*Text Word*: Like image word, text word $W_n$ is defined for a dictionary of $D_v$ words. An image is a collection of $N^w$ words and their occurrence are denotes as $V=\{v_1,v_2,v_3....,vN^{V}\}$ and text is a collection of $N^w$ words their occurrences are denoted as $W=\{w_1,w_2,w_3....wN^w\}$[1]

*Semantic Relationship:* Semantic relationship is a study of correlation or close association between the meanings of words in a document. Semantic relationship can be studied at the various levels. Such as Synonymy, Antonym, Homonymy, Hyponymy, Metonymy, Paraphrase, Ambiguity etc.

The words have similar, opposite meanings, ambiguous relationships, same grammatical category property or the one word which contains meaning of another word wrapped in it, So based on type of relationship between the words the semantic relationship gets formed.

*Topic Modeling:* In natural language processing, topic model is statistical model which can obtain the abstract patterns of topics in a collection of documents. It is widely used in text mining for discovery of hidden semantic structures in a document. The topics which are generated with the help of topic modeling are clusters of the word containing some semantic relationships between the words in a document. Topic models are statistical probabilistic models which help to discover the latent semantic structure of text body.

*Latent Dirichlet Allocation*: is an unsupervised machine learning technique which identifies latent topic information in large document collections. It uses a "bag of words" approach, which treats each document as a vector of word counts. Each document is represented as a probability distribution over some topics, while each topic is represented as a probability distribution over a number of words. LDA defines the following generative process for each document in the collection:

1. For each document, pick a topic from its distribution over topics.

2. Sample a word from the distribution over the words associated with the chosen topic.

3. The process is repeated for all the words in the document.

## 3. RELATED WORK

### 3.1 Social Event Detection and tracking:

A lot of work has been carried out in area of event tracking and topic detection. Among them most of the methods are based on single modality information or multi-modality information. However, these models studies visual and non-visual modalities in isolation to model the multimedia event data for social media analysis.

i. Diakopolas et al. have propose work for studying event visualization and social event analysis by using the twitter tweets related to particular event. Extracting information from large datasets and crawling the dataset information is included in this work for social event analysis [5].

ii. Makkonen et al. propose the model for extracting meaningful semantics such as names, tags, time references Based on a single clustering partition. A similarity metric have been proposed for these events [33].

iii. Hierarchical Hidden Markov model has been proposed by Xie et al. over the low-level audio-visual features for discovering the location and time based i.e. spatio-temporal patterns. For finding the clusters of text, the Latent Semantic Analysis is used [35].

iv. Non-negative Matrix Factorization framework was proposed by Lin et al. by using multi-relational structures for modelling the image stream data including images and short tags form social media events[36].

v. Zhang and Xu propose a CO-PMHT model to track event using cross domain knowledge and obtain their summary over time from social media events [3].

Vi. Michele Merler in 2012 proposes a Semantic model vectors representation. In this work video event detection has been studied, which combines semantic model vectors and other static or dynamic visual descriptors by extracting the information from various frames in videos [2].

### 3.2 Event Summarization:

In the area of event summarization, multi-document event summarization leads to solve the problem regarding overloaded information over the internet with respect to social media events.

Gong and Liu propose generic text summarization method that creates text summaries by ranking and extracting sentences from the original document [39]. Zhou et al. propose two-layer summarization framework for summarizing multiple disaster related documents.

### 3.3 Social Event Analysis:

Social event analysis based on sentiment analysis and sentiment classification has derived much attention. Timo Reuter in 2012 proposes a system that is able to perform pattern classification of a social media streams. Hu et al. explained whether social relations could help sentiment analysis with the help of social approach to deal with noisy type of data and short texts for classification based on sentiments.

### 3.4 Topic Model:

Topic models that are widely used for the topic modelling includes Latent Dirichlet Allocation (LDA) and probabilistic Latent Semantic analysis. These topics are extended further by introducing Supervised Latent Dirichlet allocation (SLDA). Yang in 2015 proposes a novel cross domain feature learning framework based on stacked denoising auto-encoder [28]. This algorithm helps to maximize correlations among various modalities and helps to extract semantic features at the same time. Al Sumait et al. propose online LDA method, which further extends Gibbs Sampling method, which derives topic-word distribution at next time slice [32]. Hong et al. propose a topic model, which is time-dependent and can be used for considering multiple text sources. However, these models fail to properly model the multi-model data. Therefore, Corr-LDA was proposed to capture correlations between image and annotations. The mm-LDA is proposed for social relation mining. The mm-LDA can be used for multi-modal information modelling which includes textual corpora and visual topics. Putthividhya et al. propose a topic-regression multi-modal Latent Dirichlet Allocation (LDA) Model. The model mm-LDA is the extension of the special words with background (SWB) model. The multi-modal event topic model is advanced level model. Which is applied for the social event detection and tracking. Blei and Lafferty Propose the dynamic model that makes the use of state space models based on the natural parameters.

McCallum et al. proposed a model to simultaneously discover groups among the entities and topics among the corresponding text. Zhang et al introduced a model to incorporate LDA into a community detection process.

### *3.5* Suspicious behavior detection:

A variety of research has found fraudulent behavior through mining multimodal relational data. These patterns of fraud have been found to show up in eBay reviews, opinion spam, and false accounts, among many others. Many methods have focused on labeling individual users, such as by using belief propagation (BP) or TrustRank scores [37][38]. These methods label suspicious nodes/users, but do not return suspicious grouping behaviors themselves. Later works found that adding additional modes of information aided in detecting suspicious behavior. CopyCatch found that suspicious patterns of Page Likes on Facebook correlated in time were good indicators of fraud. Many of the above methods return labels or scores for individual users or IP addresses but not blocks [17]. Even a human evaluation of the results is difficult. Finally, because they are operating on independent formulations, it is impossible to compare their effectiveness and measure progress in the field as a whole. However, none of them gives a "surprise" scoring function for a dense sub-tensor. Rather, in this paper we study and quantify this pattern in a principled manner.

### Multi-Modal Feature Learning

To deal with the multiple modalities of features, there are mainly three different kinds of learning methods,

*1) Feature Subspace:* The most widely used feature subspace method for multiple modalities is the canonical correlation analysis. This method can be seen as the problem of finding basis vectors for variables with different modalities. Thus, the correlation between the projected vectors of the variables along the basis vectors is mutually maximized. The basis vectors are decided by a set of linear transformations, one for each modality of the variables.

*2) Semantic Integration:* In, the query-by-example paradigm is extended to the semantic domain. A semantic feature space where each image is represented by the vector of posterior concept probabilities is defined. The semantic representation for each image is constructed based on the correlation space where the original features are mapped using CCA. The correlation between two modalities based on CCA and the semantic representation based on multi-class logistic regression are combined in this method.

### Multi-Objective Optimization

Multi-objective optimization is a field of optimization dealing with problems having not one but many, usually conflicting, objectives to be simultaneously optimized (minimized or maximized). Such problems arise frequently, especially in engineering and economics, for instance maximizing speed while minimizing fuel consumption, or minimizing the error in both the position and the orientation of a robotic arm. These objectives are conflicting, so that a solution that is optimal for either of them is usually suboptimal for the other. [7]

## 4. Proposed dense block detection and Multi-modal topic modeling overview:

The multi-modal event topic model (mm-ETM) is suggested in this paper which is used for finding the correlations between visual and textual modalities. This model can also used to track and detect the events according to time when the event is happened. The dependency relationships between textual and visual modalities are different for different semantic concepts. The event may contain text as well as images corresponding to particular topic. Some topics are represented using textual information and some using visual descriptors. For modelling such text corpora and multimedia contents, the multi-model topic is well suited [1].

### 4.1 Objectives of Dense block detection and Multi-modal topic modelling

1. In case of multi-modality of data text and images are mostly concerned. We have to model these two contents together. Multi-modal topic model works very well for modelling these contents together.
2. In case of Text data, dataset contains the information regarding each topic and images concerned with it.

3. With the help of some algorithms and mathematical derivations Multi-modal data can be modelled.
4. After processing the model user need to import two datasets on the JSP page as meta-dataset and reviews-dataset. Both datasets will be internally processed.
5. On View Reviews page, you can search and sort the reviews, products, etc
6. On Search Page, you can write text by your own, or copy any review from the dataset file, and search it. It will give the images by mining according to the text searched. Also it gives corresponding textual result data.
7. Dense block detection module is suggested to detect malicious behaviour. Where data block is dense then that information is always remains worth for inspecting.
8. After studying the models mentioned in this paper, proposed model can efficiently model text and an image containing dense block which also comes under inspection.

## 5. Conclusion

In this paper, a novel approach is suggested for social media event analysis. Multi-model topic model have been used for event tracking and evolution. It also helps in generating effective summaries of events over the time where with each coming day new events occur in the world. For differentiating the visual representative topics and non-visual representative topics this model can model the correlations between textual and visual modalities. Incremental updating strategies will be used in this method. Model inference and parameter inference mechanisms helps to process online data with new updating strategies.

## 6. Future Scope

In the future, more investigation can be done under this framework, such as event summarization and event attribute mining in social media. With images frames of videos and audios can also be studied. It will be big task to explore whether the tracking performance can be improved by considering different domains, such as, Flickr, Google News, YouTube.

### *References*

[1] Shengsheng Qian, Tianzhu Zhang, Changsheng Xu, and Jie Shao, "Multi-modal event topic model For social event analysis" *IEEE transactions on multimedia*, vol. 18, no. 2, february 2016

[2] M. Merler, B. Huang, L. Xie, G. Hua, and A. Natsev, "Semantic model vectors for complex video event recognition," IEEE Trans. Multimedia, vol. 14, no. 1, pp. 88–101, Feb. 2012.

[3] T. Zhang and C. Xu, "Cross-domain multi-event tracking via CO-PMHT," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 10, no. 4, pp. 31:1–31:19, 2014.

[4] X. Yang, T. Zhang, C. Xu, and M. S. Hossain, "Automatic visual concept learning for social event understanding," *IEEE Trans.Multimedia*, vol. 17, no. 3, pp. 346–358, Mar. 2015. N.

[5] Diakopoulos, M. Naaman, and F. Kivran-Swaine, "Diamonds in the rough: Social media visual analytics for journalistic inquiry," in *Proc. IEEE Symp. Vis. Analytics Sci. Technol.*, Oct. 2010, pp. 115–122.

[6] X. Wu, C.-W. Ngo, and A. G. Hauptmann, "Multimodal news story clustering with pairwise visual near-duplicate constraint," *IEEE Trans. Multimedia*, vol. 10, no. 2, pp. 188–199, Feb. 2008.

[7] I. Kalamaras, A. Drosou, and D. Tzovaras, "Multi-objective optimization for multimodal visualization," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1460–1472, Aug. 2014.

[8] W. Liu, T. Mei, and Y. Zhang, "Instant mobile video search with layered audio-video indexing and progressive transmission," *IEEE Trans.*

[9] C. Kang, S. Xiang, S. Liao, C. Xu, and C. Pan, "Learning consistent feature representation for cross-modal multimedia retrieval," *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 370–381, Mar. 2015.

[10] S. Qian, T. Zhang, R. Hong, and C. Xu, "Cross-domain collaborative learning in social multimedia," in *Proc. ACM MM*, 2015, pp. 99–108.

[11] T. Zhang, C. Xu, G. Zhu, S. Liu, and H. Lu, "A generic framework for video annotation via semi-supervised learning," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1206–1219, Aug. 2012.

[12] X. Yang, T. Zhang, and C. Xu, "Cross-domain feature learning in multimedia," *IEEE Trans.Multimedia*, vol. 17, no. 1, pp. 64–78, Jan. 2015.

[13] M. Jiang, P. Cui, and C. Faloutsos, "Suspicious Behavior Detection:Current Trends and Future Directions," in IEEE Intelligent Systems, no. 1, pp. 31–39, 2016.

[14] Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi, Shiqiang Yang Senior Member, IEEE, Christos Faloutsos,IEEE Trans. knowledge and Data Engineering, Manuscropt ID "Spotting Suspicious Behaviors in Multimodal Data: A General Metric and Algorithms".

[15] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang, "Catchsync: catching synchronized behavior in large directed graphs," in ACM SIGKDD, pp. 941-950, 2014.

[16] N. Shah, A. Beutel, B. Gallagher, and C. Faloutsos, "Spotting suspicious link behavior with fbox: An adversarial perspective," in IEEE ICDM, pp. 959–964, 2014.

[17] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos, "Copycatch: stopping group attacks by spotting lockstep behavior in social networks," in WWW, pp. 119-130, 2013.

[18] S. Pandit, D. H. Chau, S. Wang, and C. Faloutsos, "Netprobe: a fast and scalable system for fraud detection in online auction networks," in WWW, pp. 201-210, 2007.

[19] K. Maruhashi, F. Guo, and C. Faloutsos, "Multiaspectforensics: Pattern mining on large-scale heterogeneous networks with tensor analysis," in ASONAM, pp. 203-210, 2011.

[20] H.-H. Mao, C.-J. Wu, E. E. Papalexakis, C. Faloutsos, K.-C. Lee, and T.-C. Kao, "Malspot: Multi2 malicious network behavior patterns analysis," in PAKDD, pp. 1-14, 2014.

[21] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in WWW, pp. 191-200, 2012.

[22] L. Akoglu, R. Chandy, and C. Faloutsos, "Opinion fraud detection in online reviews by network effects," in ICWSM, 2013

[23] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang, "Inferring strange behavior from connectivity pattern in social networks," in PAKDD, pp. 126-138, 2014.

[24] J. Pei, D. Jiang, and A. Zhang, "On mining cross-graph quasicliques," in ACM SIGKDD, pp. 228-238, 2005.

[25] J. Chen and Y. Saad, "Dense subgraph extraction with application to community detection," in IEEE TKDE, vol. 24, no. 7, pp. 1216-1230, 2012.

[26] A. Sun and S. S. Bhowmick, "Quantifying visual-representativeness of social image tags using image tag clarity," in *Proc. Social Media Modeling Comput.*, 2011, pp. 3–23.

[27] C. S. Firan, M. Georgescu, W. Nejdl, and R. Paiu, "Bringing order to your photos: Event-driven classification of Flickr images based on social knowledge," in *Proc. CIKM*, 2010, pp. 189–198.

[28] T. Zhang, C. Xu, G. Zhu, S. Liu, and H. Lu, "A generic framework for video annotation via semi-supervised learning," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1206–1219, Aug. 2012.

[29] X. Yang, T. Zhang, and C. Xu, "Cross-domain feature learning in multimedia," *IEEE Trans.Multimedia*, vol. 17, no. 1, pp. 64–78, Jan. 2015.

[30] X. Yang, T. Zhang, C. Xu, and M. Yang, "Boosted multifeature learning for cross-domain transfer," *TOMCCAP*, vol. 11, no. 3, pp. 35:1–35:18, 2015.

[31] S. Qian, T. Zhang, C. Xu, and M. S. Hossain, "Social event classification via boosted multimodal supervised latent Dirichlet allocation," *TOMCCAP*, vol. 11, no. 2, pp. 27:1–27:22, 2014.

[32] Y.-R. Lin, H. Sundaram, M. D. Choudhury, and A. Kelliher, "Discovering multirelational structure in social media streams," *TOMCCAP*, vol. 8, no. 1, p. 4, 2012.

[33] L. AlSumait, D. Barbará, and C. Domeniconi, "On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking," in *Proc. ICDM*, 2008, pp. 3–12.

[34] J. Makkonen, H. Ahonen-Myka, and M. Salmenkivi, "Simple semantics in topic detection and tracking," *Inform. Retrieval*, vol. 7, no. 3–4, pp. 347–368, 2004. L.

[35] Xie *et al.*, "Discovering meaningful multimedia patterns with audio-visual concepts and associated text," in *Proc. ICIP*, 2004, pp.

[36] Y.-R. Lin, H. Sundaram, M. D. Choudhury, and A. Kelliher, "Discovering multirelational structure in social media streams," *TOMCCAP*, vol. 8, no. 1, p. 4, 2012.

[37] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," in 1999.

[38] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos, "Neighborhood formation and anomaly detection in bipartite graphs," in IEEE ICDM, p. 8, 2005.

[39] Y Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," in *Proc. SIGIR*, 2001, pp. 19–25.