ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

A Framework for Mining Public Health Using Hierarchical Latent Dirichlet Allocation (HLDA).

Vijeya Kaveri V*, and Maheswari V.

*Research Scholar , Sathyabama University, Chennai, Tamil Nadu, India.
Professor, Sathyabama University , Chennai, Tamil Nadu, India.

ABSTRACT

In traditional health recommending system the recommendations are not personalized according to the patient's, the recommendation extremely depends on physical , emotional and psychological matters of the user list is generated based on the diseases the patient navigates. In health era, personalized health recommendation helps us to extract personalized health content form overloading information's available on the web. the patient's , the initial step is to identify the patient interest in which health related issues they needs recommendation. In this paper we have used statistical topic modeling technique Hierarchical Latent Dirichlet Allocation (HLDA) to identify the user interest which provides robust and interpretable topic representation. After identifying the user interest, neighborhood selection is done based on ranking and finally recommendation is done according to user preference. In this model we have learned six parameters , in parameter(1) the topical distribution of each document is learnt , in parameter (2) the perspective distribution of each user is learnt , in parameter (3) the word distribution of each topic is learnt , in parameter (4) the tag distribution of each topic is learnt, in parameter(5) the tag distribution of each user perspective is learnt , in parameter(6) the probabilistic of each tag being generated from resource topics or user perspectives is learnt. Our Experimental results show that proposed model is better than State – of – Art than other models

Keywords: Personalized Health Recommendation ,topic modeling , user interest , HLDA .

Corresponding author

ISSN: 0975-8585

INTRODUCTION

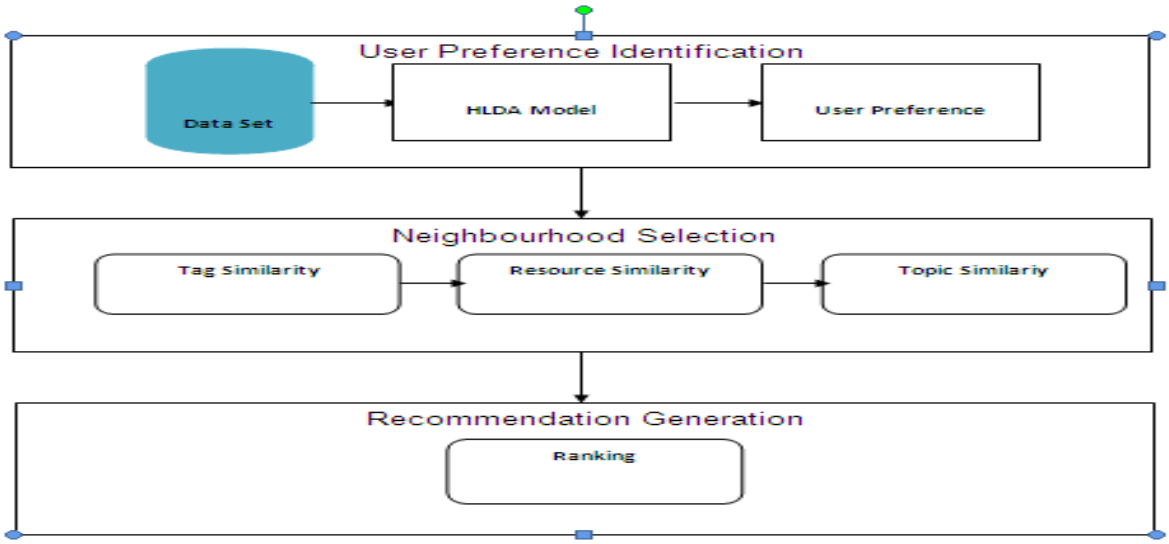
Social media is a place where people share, post, tag and comment. Twitter is one source of social media data (tweets) from whch we can analyze the events happeningaround us in the world . Tweets help us in getting updated new events , user sentiments , natural disaster and political opinions. Tweets help us analyzing what is currently happening in the world. Now a days researches are using the tweets to monitor and analyses health issues.Influenza is identifed to be one of the most common disease from social media feed. Generally from the tweets the common health issues like influenza or health topics such as diet and exercise are identified . In this paper we have proposed a Hierarchical Latent Dirichlet Allocation HLDA model for illness discovery and this model helps in identifying new illness without any prioir knowledge. In this paper, we have proposed a statistical topic modeling framework which identifies all general public health information from millions of health-related tweets.

In our proposed framework we have discovered ailment, learn symptoms and treatment associated from tweets . Our first work is to build data set from 1.6 million health related tweets. we have developed a new topic model to create structured information from these data. The model organizes health terms into ailments, including associated symptoms and treatments. In our model we have seperated the ailment groups and more general topics by using explicit knowledge of symptoms and treatments .

Related Work

In the previous work of Tso-Sutter et.al [1] they have utilized the explict rating as well as tagging information as additional source for recommendation , but there performance of recommendation is affected by tag quality. In the work of Gemmis et al. [2] they have integrated the tagging information and proposed a content – based recommender system . Liang et al. [3] in his work has proposed a collaborative filtering items by extending user – item matrix to user – item – tag matrix. In the work of Sen et al. [4] a special tag rating function has been defined , which combines explicit user ratings with the predicted user preferences for items based on the preferences inferred for tags for recommendation . In the work of Niwa et al.[5] using the tf-idf term factor and inverse document frequency they have suggested the web pages . Shepitsen et al. [6] proposed a STS by using hierarchical clustering algorithm for tagged data to provide user interest based on resource . Bogers and Bosch[7] has surveyed a number of recommending approaches for social bookmarking website users that incorporated social tagging data. Zarli Htun et al.[8] has proposed Collaborative Filtering RS to extract the hidden topic from the resource collection and interested topics are identified by measuring the similarity of the users based on their browsing resource history. In this paper we have proposed Hierarchical LDA model for extracting hidden topics(tag) and their area of interest are measured by similarity measures form their browsing resource history .

Proposed System



ISSN: 0975-8585

The proposed Framework consits of three steps :

- User Preference Identification
- Similarity Selection
- Recommendation Generation

User Preference Identification

In this phase in order to identify the user preference we have used Hierarchical LDA (Latent Dirichlet Allocation) Topic Modelling approach . It is a probabilistic model in which the user interested topic (tag) is identified . It is a tree like arrangement where the more general tags are arranged near the root and at the leaves the specialized tags are arranged. Following formula is used for calculating the user level of interest on tag.

Each user 'u' has

- R (u) = {r_1… r_m} set of resources of his interest
- T(u) = {t_1, t_2… t_n} set of T personal tags - annotate resources
- rs(u,r_j) - interest weight of a user for a resource

$$rs(u, r_j) = \sum_{i \in Tu,r} ts(u, t_i)$$

Where

- rs(u,rj) = resource score of user 'u' for resource 'rj',
- Tu, r = tags used by user 'u' to annotate resource 'r',
- ts(u,ti) = tag score of user 'u' for tag 'ti' which is calculated as
-

$$ts(u, t_j) = \frac{freq(u, t_j)}{\sum_{t_i \in T(u)} freq(u, t_i)}$$

- ts(u,tj) = tag score of user 'u' for tag 'tj'
- freq (u,tj) = the number of times that user 'u' used tag 'tj'
- freq (tl) = total frequency of all tags used by user 'u'.

Once the resource interest weights are calculated, latent topics are derived based on user's interest. For each user 'u' user profile P is build which is represented as a vector of his interest topics with its weights,

$$P = \{(w_1, INF(u, w_1).....(w_k, INF(u, w_k)\}$$

INF

$$INF(u, w_k) = \max\{rs(u, r_1).....rs(u, r_s)\}$$

Similarity (Neighbourhood) Selection

In Neighbourhood selection Pearson correlation method is used to find users with similar topic interests . User topic profiles are matched to measure the interest similarity between users. In a non-rating environment topic similarity alone cannot give good neighbourhood selection . So in the proposed system the user similarity is calculated based on three similarity measures

1. Tag usage similarity
2. Resource item similarity
3. Interest factor similarity

ISSN: 0975-8585

Recommendation Generation

Recommendation list are generated 1depending on ranking of an tagged item by choosing the resources of similar neighbor as mentioned by the equation below

$$Rank(u,T_i) = \sum_{x \in Nei(u)} sim(u,x)$$

Where

- Rank(u,Ti) – Ranking of user u and tagged item
- Nei(u) - neighbors of user u
- Sim(u,x) - similarity value of user u and his neighbor x

EXPERIMENTS AND RESULTS

The dataset has been divided into two 80% of dataset for training and 20% of dataset for testing. The performance of the recommender system is measured by calculating Recall . For top-N RS, 'recall' is the number of items in the user's test set that also exists in the top-N recommended items. Therefore, recall is the ratio of hit set (HIT) size to the relevant set (REL) size (test set). Therefore, for all n tested users,

$$recall = \sum u| \frac{HITu}{RELu}$$

Where n is the number of user tested.

We have compared the proposed system with user-based collaborative filtering system s UI-IDF-CF and UT-IDF-CF and with Kullback Libler Divergence KL-CF .

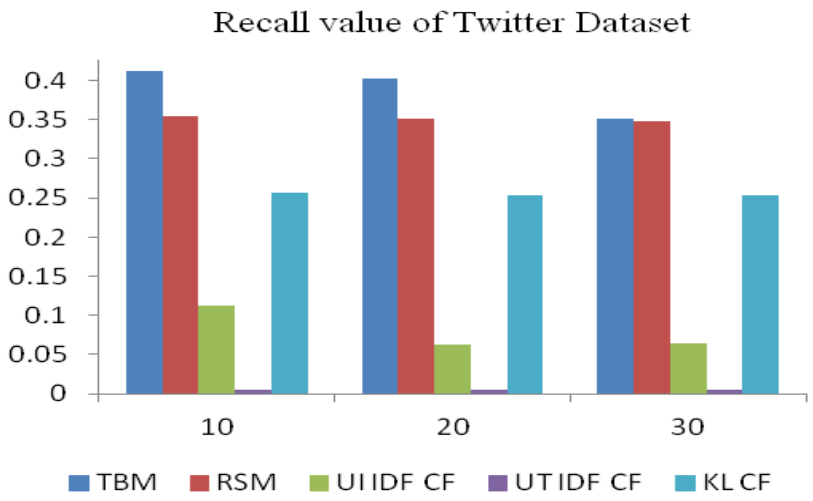


Fig. 3 shows that the proposed approaches can perform better than comparison approaches in both datasets.

CONCLUSION

In this paper we have proposed a framework for recommendation system based on tagging information provided by the user in STS. We have adopted HLDA model to derive the user preferred topics . Once the topic is extracted , then user – topic rating matrix is generated as implicit rating matrix from a non-rating environment like social book marking . In the second phase neighbourhood selection is done based on tag similarity , resource similarity and interest factor similarity . In the third phase top – N recommendation are given to the user by the user – topic rating matrix. From the above, experimental result we conclude that our proposed system achieves better performance than the other systems.

ISSN: 0975-8585

REFERENCES

[1] Tso-Sutter, K.H.L., Marinho, L.B. and Schmidt-Thieme, L.(2008) "Tag-aware Recommender Systems Fusion of Collaborative Filtering Algorithms", Proceedings of the 2008 ACM symposium on Applied computing,ACM,USA.

[2] Gemmis, M. D., Lops, P., Semeraro, G., & Basile, P. (2008) "Integrating Tags in a Semantic Contentbased Recommender", Proceedings of ACM Conference on Recommender Systems, 163-170.

[3] Liang, H., Xu, Y., Li, Y., Nayak, R., & Weng, L. T. (2009) "Personalized recommender systems integrating social tags and item axonomy", Proceedings of the Joint Conference on Web Intelligence and Intelligent Agent Technology.

[4] Sen, S., Vig, J., & Riedl, J. (2009) "Tagommenders: Connecting Users to Items through Tags", Proceedings of the 18th International Conference on World Wide Web, 671-680.

[5] Niwa, S., Doi, T. and Hon'Iden, S. , (2006) "Web Page Recommender System Based on Folksonomy Mining", Transactions of Information Processing Society of Japan, 47(5).

[6] Shepitsen, A. , Gemmell, J. ,Mobasher, B. and Burke, R., (2008) "Personalized recommendation in social tagging systems using hierarchical clustering", In Proc. Of the 2008 ACM conference on Recommender systems.

[7] Bogers, T. and Bosch, A.V.D. (2009) "Collaborative and Content-based Filtering for Item Recommendation on Social Bookmarking Websites", ACM RecSys '09 workshop on Recommender Systems and the Social Web.

[8] Zarli Htun and Phyu Phyu Tar , "A resource recommender system based on Social tagging data", Machine Learning and Applications: An International Journal (MLAIJ) Vol.1, No.1, September 2014.