International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 07 | July -2017 www.irjet.net p-ISSN: 2395-0072

Long Tail Keyword Suggestion for Sponsored Search Advertising

Ms. Rinki Talwar[1], Dr. Shuchita Upadhyaya[2]

[1]M.Tech Scholar, Dept. of Computer Science & Applications, Kurukshetra University, Haryana, India
[2]Professor, Dept. of Computer Science & Applications, Kurukshetra University, Haryana, India

---

Abstract - *Sponsored Search Advertising is a method of placing online advertisements along with organic results on the Search Engine Result Page that will be displayed when a user enters a query on search engine. Search advertising is sold and delivered on the basis of keywords. The main problem in Sponsored Search Advertising is of keyword suggestion. In past, the advertisers tend to bid for the keywords that have more search volume rather than that of having low search volume. Hence, the bidding price of former is more than that of latter. In this paper, an improved topic modeling based approach is proposed to suggest the related long tail keywords for advertisers that have low volume and are inexpensive but generates the equal amount of traffic cumulatively. Experimental results on AOL search data, 2006 shows that the proposed approach performs better than existing keyword suggestion method.*

Key Words: **Long Tail Keywords, LightLDA, Sponsored Search Advertising, Keyword Suggestion, AOL Search Data.**

1. INTRODUCTION

1.1 Background

Search engine advertising has nowadays become the major evolving technology. There are two kinds of advertisements: (a)Banner advertisements: It is a type of display ads usually used by the large advertisers, and (b)text advertisements or keyword advertisements or Sponsored search advertisements: It is a type of textual ads which are used mainly by all scale businesses and this type of ads contains title and a short description of the services offered by the advertisers, and also it contains the URL which will direct the user to the advertiser's website [**1**]. Sponsored search ads are based on the queries entered by the user online on search engine and deliver many relevant results so, they are considered less intrusive than that of banner advertisements or pop-up advertisements. In this research paper, the focus is on the text advertisements or Sponsored search advertisements. Textual ads are the major part of Internet-Marketing. While setting up a campaign with text ads, the advertisers are asked to associate the keywords with their advertisements to which it best describes. There are two types of text advertisements system: (i) Keyword Targeted Advertisement System, for ex. Google's AdWords which places the ads on the search result page [**1**] by matching the keywords entered by the user with the keywords of the advertiser's ads, and (ii) Content Targeted Advertisement System, for ex. Google's AdSense [**1**], which places the advertisements on the content-rich websites such as newspapers etc. There exist numerous tools that provide keyword suggestions to the advertisers for their advertisement for ex. Google's WordStream. There are various sources on the basis of which keywords can be extracted and suggested to the advertisers which are, (i) Query log based keywords suggestion, (ii) Proximity based keyword suggestion, and (iii) Meta-tag crawler-based keyword suggestion.

The nature of keyword suggestions for the sponsored search advertisements can be *short-tail* and *long-tail.* Long-Tail keywords are keyword phrases made up of 3-5 or more words. The price of the head queries i.e. short-tail queries are higher than that of the long tail queries b'coz the competition between the advertisers for the short-tail queries is higher than the long-tail queries. The long-tail query suggestions for advertisements shows the best results as long tail suggestions depicts the context of the user [**2**] [**3**]. According to NEIL PATEL [**4**], the long query suggestions for advertisers generate high traffic on their advertisements rather than using Head Terms only (Fig-1). That's why, advertisers prefer to use the long keywords because the customers who type descriptive keywords are more qualified than those who type only short queries and thus their advertisements reach to those customers directly. The Example is shown in (Fig-2), which conveys that the keyword 'shoes' generates HighCost and Competition but, is not depicting the user's intentions exactly whereas the keyphrase 'red Nike men's Running Shoes' depicts the user intentions very clearly and have Less Cost also.

1.2 Authors' Contribution

In this paper, the long-tail keyword suggestion system for sponsored search advertisements is proposed. In [**5**] Qiao proposed a keyword suggestion model for web advertisement using the famous Topic Modeling, LDA with Gibbs Sampling Approach, but its sampling complexity is more. In this paper, the LightLDA an improved topic modeling approach is used instead of using GibbsLDA [**5**], because sampling complexity of LightLDA is less than that of GibbsLDA as it uses the Metropolis-Hastings-Sampler that constructed proposals very carefully and also results in high convergence rate.

LightLDA also decreases the running time of the model proposed in [**5**]Qiao work that uses GibbsLDA. Proposed

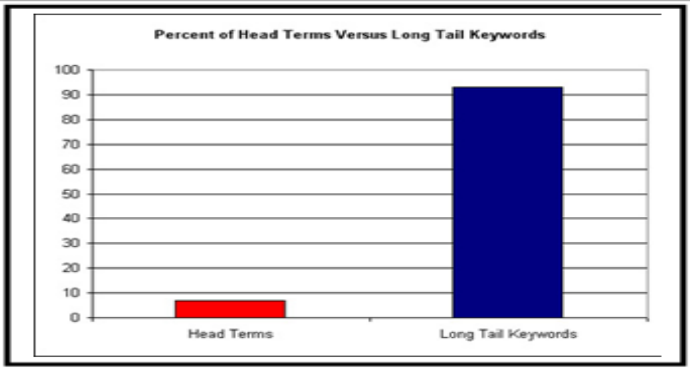

Fig-1 Comparison of Head Terms and Long Tail keywords

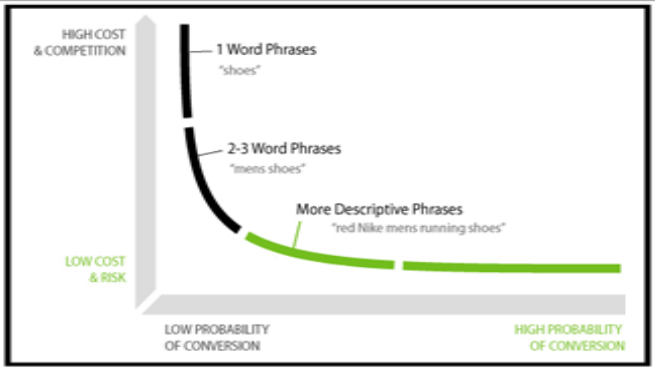

Fig-2 Competition of Word Phrases (short-tail) and more Descriptive Phrases (long-tail)

Work is targeted to increase the Precision, Recall, F1-measure and decrease the running time.

The objectives of this paper are:

- To increase the accuracy of Qiao et al. Model [**5**].
- To decrease the running time of the Qiao et al. Model [**5**].

The rest of this paper is organized as follows: in Sect.2 the literature review related to the keyword suggestion methods for the Sponsored Search Advertising is presented. In Sect.3 Lda using gibbs sampling is described. The proposed framework is discussed in Sect.4. The experimental setup and experimental results are described in Sect.5 & Sect.6 respectively and finally the conclusion is outlined in Sect.7.

2. RELATED WORK

Various methods have been proposed for suggesting the keywords for advertisements. According to various types of data sources, the keyword suggestion methods are broadly classified into 3 major categories:

- ***Keyword suggestion on basis of Query Log***: In this method, the keywords are suggested by using query log of search engine. In the query log based method, the keywords are suggested by conducting the co-occurrence analysis in search engine query logs [**6**] [**7**]. For example 'Nexus' is found to associate with 'Google' mainly, so the Nexus can be suggested as the competitive keyword of the 'Google' (i.e. seed keyword) to the advertisers. In [**8**], Zhang et al. proposed a relevant but less competitive keyword suggestions to the advertisers in order to boost the revenue of the search engine and to fill the empty ad slots. In [**1**] Sarmento et al. proposed the "synonymy" method of suggesting keywords by mining the previously submitted ads and thereby finding the relevant and irrelevant keywords and thus suggesting the relevant to the advertisers. In [**9**] DA et al. propose a new method which depicts that query logs timely measures the user's intentions and can be used widely in commercial advertising. In [**10**] Chuklin et al. proposed the good query expansion methods that will satisfy the user needs as the suggested query contains the seed term. In [**11**]Chen et al. proposed the "concept based suggestion method" in which the keywords are suggested according to the conceptual information extracted rather than the statistical occurrence in the query log. In [**12**], Jiang et al. proposed a framework on basis of which the queries in the log that matches best with the input query leading character is suggested to the users. In [**13**] Szpektor et al. proposed a new method of long-tail queries suggestions on the basis of Query-Flow-Graph and concentrated on the query templates rather than individual query-transitions. In [**14**], the author proposed a 'TermsNet' method in which the semantic relations between the terms and the neighbor of a term is found out using the directed graph and then the non-obvious words got rejected and important are suggested to the advertisers. In [**5**] Qiao et al. proposed a technique in which competitive keywords are generated as output where in first step the candidates are generated using indirect association analysis on query log and after that the topic modeling GibbsLDA approach is applied on the candidates set which returns the competitive i.e. related keyword suggestions to the seed keyword.
- ***Keyword suggestion on basis of Proximity with seed***: In this method, the keywords which are having the high proximity with the seed keyword are suggested. In [**15**] Abhishek and Hosanagar proposed a method which suggests the keyword by finding the semantic similarity between the terms by constructing the similarity graph and suggesting the similar but cheaper keywords. In [**16**]Broder et al. proposed a method of classifying the user's queries and then suggesting the keywords similar to the entered queries topics. In [**11**], some researches also uses the thesaurus/dictionary (corpus already constructed by the researcher) as a measure of proximity calculation. There are various suggestion methods of this kind, besides this it is not used much nowadays because it doesn't encounter the user real intentions which are the primary concern of the advertisers.
- ***Keyword suggestion on basis of Meta-tag crawlers***: In this method, the keywords are extracted from the

meta-tags of the similar type of advertisements, for ex. the seed keyword is entered to the search engine and the meta-tags from the relevant web-advertisements and web-pages got extracted and are suggested as keywords to the advertisers. But this kind of methods has some problems such as, firstly it doesn't concentrate on the advertiser's concerns and secondly advertisers may not be able to get many important keywords.

3. LDA USING GIBBS SAMPLING

LDA (Latent Dirichlet Allocation) is a topic modeling approach used to extract hidden patterns from the textual documents. It is an unsupervised machine learning technique. Latent dirichlet allocation is one type of topic model and was firstly presented as a graphical model by David Blei, Andrew Ng, Michael I. Jordan [**17**], and many variations of this has come till now. The GibbsLDA i.e. LDA using gibbs sampling assumes that each document is a distribution over topics and further each topic is distribution over words. LDA model is given in Fig-3. LDA uses 2 dirichlet distributions θ and Ø with two parameters α and β respectively, denoted by θ~Dirichlet (θ|α) and Ø~Dirichlet (Ø|β).

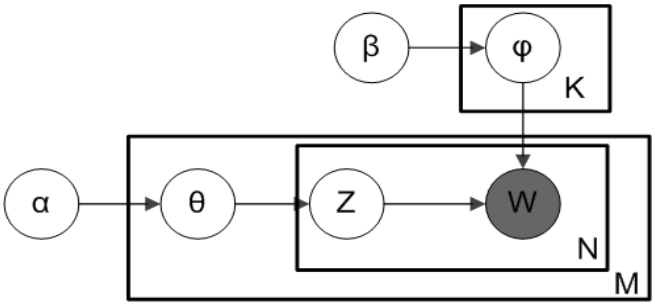

Fig-3 Plate Notation of LDA using gibbs sampling

Here, M is denoting the number of documents, N is denoting the number of words in the document and lastly K is denoting the number of topics. All, the other parameters are described below:

α is the parameter of the dirichlet prior on the per-document topic distributions,

β is the parameter of the dirichlet prior on the per-topic word distribution,

θ_M is the topic distribution for document *M*,

$Ø_K$ is the word distribution for topic *K*,

$Z_{M,N}$ is the topic for the *n*-th word in document *M*, and

$W_{M,N}$ is the specific word.

Qiao et al. uses the GibbsLDA approach in their proposal [**5**] which has certain limitations.

3.1 Limitations of LDA using Gibbs Sampling

There are certain limitations of LDA:

- Firstly, the speed of gibbs sampling inference method is too slow for large dataset with many topics.
- Secondly, the sampling complexity of gibbs sampling is more.
- Thirdly, the topics learned by LDA sometimes are difficult to interpret by end users. (i.e. accuracy of LDA using Gibbs sampling is not so good).
- Fourthly, LDA suffers from instability problem, which occurs not only when there is new data arrives and the model needs to be update but also when the same Gibbs sampling method is run multiple times on the same data.

To remove these limitations of LDA using gibbs sampling, the LightLDA an improved topic modeling approach is used in Qiao et al. model [**5**] instead of LDA using gibbs sampling to enhance the performance of that system.

4. PROPOSED FRAMEWORK

The proposed approach has been given the name "LightLDA based keyword suggestion method" because it uses the LightLDA topic modeling approach. It consists of 2 steps: (i) Candidate Set Generation, and (ii) Improved Topic modeling.

The whole process is depicted in the model (Fig.-4).

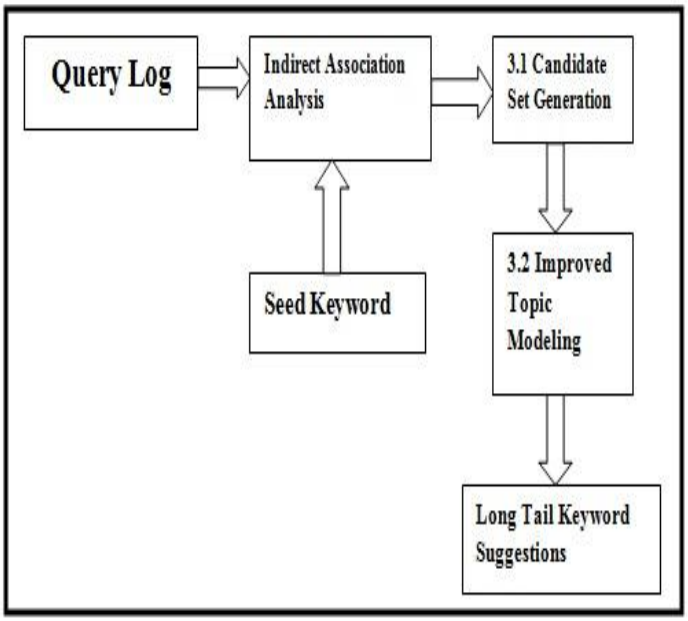

Fig-4 Architecture of LightLDA based keyword suggestion method

4.1 Candidate Set Generation

The very first step is to generate the candidate keywords for the seed keyword using the query log. Query Log used here is the **AOL Query Log, 2006** which is available online and consists of queries that user uses to search online on AOL search engine. The clustered form of AOL data is available from [**18**] which consists of 2 elements Query Keywords(i.e. represented by q.k) and its volume, that is how many times a user uses this keyword

for searching(i.e. represented by q.k_vol). Sample data of query log is shown in (Table-1).

Table -1: Query Log

Query Keyword(q.k)	*Query Volume(q.k_vol)*
Cingular	7344
Horoscope	2500
Colgate	500
Google	1500
Wikipedia	200

The seed keyword, say's' is the keyword regarding which the long-tail keyword suggestions are to be generated for the advertisers for their sponsored search advertisements. The candidates are indirectly associated keywords to the seed keywords. The method used to generate the candidates is *Candidate_Generation*. Here, in this method firstly the keywords directly associated with the seed is found out which is denoted by s.AK and then the keywords directly associated with s.AK and also indirectly associated with the seed are found out which forms the candidate set and is denoted by cCand. Corp denotes the corpus consisting of directly and indirectly associated keywords. For example, if a query 'Apple iPad' is taken then 'iPad' is co-occurring keyword with the keyword 'apple', it means keyword 'iPad' is associated with the keyword apple.

Candidate_Generation:

Input: Query-Log 'Q', Seed 'S'
Output: Candidates Keyword Set, cCand(c1,c2,c3)
Begin:
1. Initialize arrays 's.AK' , 'corp' , 'cCand' and 'list' to φ.
2. For each q in Q
3. if q contains S
4. add q to list
5. end
6. s.AK=s.AK U FindAssociativeKeyword(S,Q) //Method to find the associative keyword
7. For each Query 'q1' in s.AK
8. cCand=cCand U FindAssociativeKeyword(q1,Q)
9. end
10. For each Query 'q' in s.AK
11. add q to corp
12.end
13. For each Query 'q1' in ccand
14. add q1 to corp
15.end
16. return corp, cCand and list

4.2 Improved Topic Modeling

This is the second step of the proposed method in which the improved topic modeling approach 'LightLDA' is applied on the candidate Keywords set 'cCand' generated in first step and in the output of this step, the Long Tail Keyword suggestions are retrieved for the sponsored search advertisements.

Light LDA: It uses the cyclic Metropolis Hastings algorithm combined with alias tables for both document-topic and that of topic-word distribution. It uses the factorized strategy of proposal instead of using a single proposal like in GibbsLDA in [**5**]. LightLDA speeds up the process as well as reduces the computational complexity. It divides the proposal probability into 2 multiplicative terms document proposal pd(k) and word proposal pw(k). To construct the 2 proposals, document proposal as well as the word proposal the true conditional probability of the topic indicator z_{di} is given as:

$$p(z_{di} = k|rest) \propto \underbrace{\frac{n_{kw}^{-di} + \beta_w}{n_k^{-di} + \bar{\beta}}}_{\text{pw(k)}} \underbrace{\left(n_{kd}^{-di} + \alpha_k\right)}_{\text{pd(k)}}$$

where,
α and β are the 2 dirichlets used by standard GibbsLDA.
$\bar{\beta}$ = number of tokens in the document that are assigned to the topic k,
n_{kw}^{-di} = number of tokens with word 'w' that are assigned to topic k, and
n_k^{-di} = number of tokens assigned to topic k.

Main motive behind using LightLDA is to have high proposal acceptance rates, good space coverage and simplify the proposal generation complexity. For the document-proposal, it generates the proxy alias table which stores the number of times each topic appears in the document and For word-proposal the alias tables are generated for each word and algorithm cycles between these two proposals and also perform the MH-Tests for the acceptance and rejection of the proposals and alias table is computed every time the word is used. The acceptance probabilities of doc-proposal and word-proposal give the proposal where as the current topic is calculated with the help of LDA's update equation. Each keyword of cCand is tested to check with which probability it is related with the Seed keyword using LightLDA. The initial parameters set for the LightLDA algorithm are α=0.1, β=0.01, k be the number of topics=2 and mh-step=2.

Pseudo Code of LightLDA

1. For each documentd d ε D
2. For each word x in d
3. compute proposal by alias 'coinflip()' method;
4. compute w matrix which stores the unique words per document
5. compute k matrix which stores the topics per document
6. decrement d,w,k count matrix
7. if proposal==0 //doc proposal
8. choose index= random(0,number of words);
9. calculate probability p=z[d][index]

```
10.          calculate mh_acceptance=
             compute_doc_Acceptance(k,p)
11.      else //term proposal
12.           p=alias_sample(w)
13           calculate mh_acceptance=
              compute_term_acceptance(w,p)
14.      end
15.   end
16.      // MH- test
17.   mh_sample=random_float(0,1)
18.   if mh_sample<mh_Acceptance
19.           increment count matrix d,w,k //proposal is
      rejected
20.           revert to k
21.   else
22.           increment count matrix d,w,p //proposal is
      accepted
23.   end
24.end
```

By applying the LightLDA on candidate set cCand generated in first step, the competitive keywords are find out, the competitive keywords here defines are the keywords that are related to the seed keyword and then that competitive keywords are searched in the 'list' returned by the method 'Candidate_Generation' and those queries that contains the competitive keywords are returned as Long-Tail keyword suggestions.

5. EXPERIMENTAL SETUP

The implementation environment for performing experiments was a Windows 7 system on Dell Pc with a memory of 500GB & Intel core i5 processor and 4GB of RAM (Random Access Memory). The project is created and implemented in the java language using NetBeans IDE tool of version 8.1.

5.1 Dataset

The dataset which is used here is obtained online from AOL Query Log (2006) and consists of following fields <AnonId, Query, Query Time, Item Rank, ClickURL>. The clustered form of AOL data is available from [**18**] which consists of 2 elements Query Keywords(i.e. denoted by q.k) and its volume, that is how many times a user uses this keyword for searching(i.e. denoted by q.k_vol) which is shown in (Table-1).

Five seed keywords selected from different domains for experiments is shown in Table-2.

Table-2 Seed Keyword

Index	Seed
1	Colgate
2	Budweiser
3	Sony
4	Pantene
5	Skype

5.2 Evaluation Criteria

The relevancy criteria to evaluate the performance of the proposed method used here is threshold, which is described as:

If the volume of the query containing the seed keyword in query log>2, then it is relevant suggestion.

5.3 Measures

Metrics used for evaluation are: Precision, Recall, F1-measure, and running time which are discussed below one by one.

- **Precision**: It is defined as the fraction of relevant instances among the retrieved instances. It is also known as the number of correct results retrieved divided by the total number of retrieved results. Mathematically can be calculated as,

$$\text{Precision}=\frac{\text{Relevant Suggestions} \cap \text{Retrieved Suggestions}}{\text{Total Retrived Suggestions}}$$

- **Recall**: It is defined as the number of correct results retrieved divided by the total number of relevant results that should have been retrieved. Mathematically can be calculated as,

$$\text{Recall}=\frac{\text{Relevant Suggestions} \cap \text{Retrieved Suggestions}}{\text{Relevant Suggestions}}$$

- **F1-measure**: It is a measure of the total accuracy. It depends on the Precision and Recall. Mathematically can be calculated as,

$$\text{F1-measure}=\frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}$$

- **Running time:** It is defined as the time required in single execution of the proposed model. It is measured in ms (milliseconds).

6. EXPERIMENTAL RESULTS

In this section, the comparative results between proposed LightLDA based keyword suggestion method and Qiao et al. proposed GibbsLDA based keyword suggestion method [**5**] are shown below in Table-3

Table-3 Performance Comparison between GibbsLDA [**5**] and LightLDA

Performance metrics	GibbsLDA based keyword suggestion	LightLDA based keyword suggestion
Precision	0.85	0.883333
Recall	0.930657	0.967153
F1-Measure	0.888502	0.923345
Running Time(in ms)	790	391

Precision: In the proposed method, the precision got increased as shown in Fig-5 also. It is more than that of the GibbsLDA based keyword suggestion method of Qiao et al [**5**]. Precision shows the average probability of relevant retrieval. Increased values of precision indicate that the more relevant results are retrieved among the suggested results.

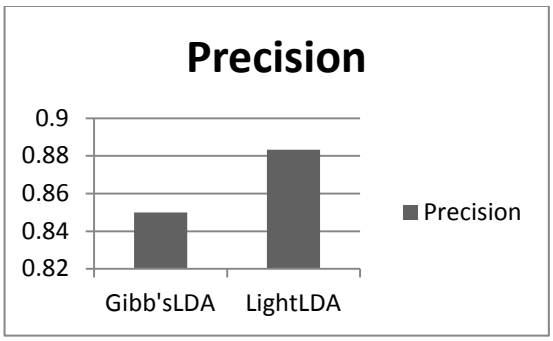

Fig-5 Result analysis of Precision

Recall: In the proposed method, the recall is more than that of GibbsLDA based keyword suggestion method of Qiao et al. [**5**] as shown in the Fig-6. Recall shows the average probability of complete retrieval. Increased values of recall will increase the F1-measure value directly.

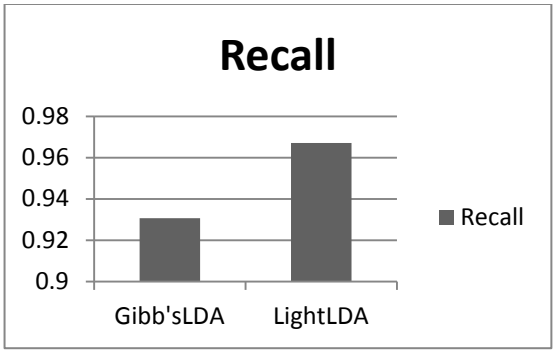

Fig-6 Result analysis of Recall

F1-measure: It is the harmonic mean of precision and recall. If both of them increase, F1-measure also get increased. The comparative graph between proposed LightLDA based keyword suggestion and GibbsLDA based keyword suggestion of Qiao et al. is shown in Fig.-7.

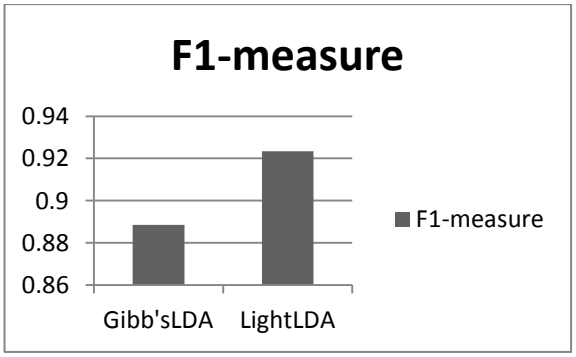

Fig-7 Result analysis of F1-Measure

It is the true indicator of accuracy. Higher values of the F1-measure of proposed method than that of GibbsLDA based keyword suggestion method of Qiao et al. [**5**] indicates the higher accuracy of keyword suggestion technique used in proposed method.

Running Time: The running time of the proposed method is less than that of the Qiao et al. proposed GibbsLDA based keyword suggestion method in [**5**]which is shown in the Fig.-8. The small running time of the proposed model shows that the speed of the proposed method is higher than the existing method [**5**].

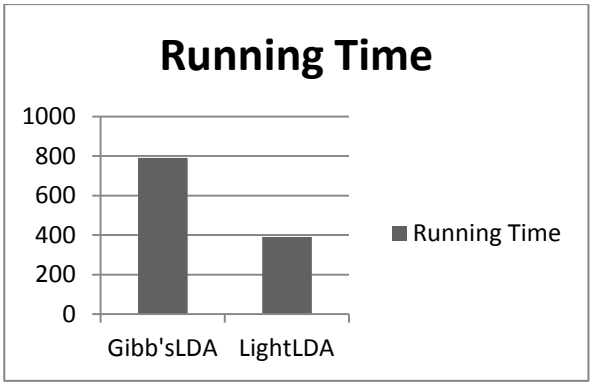

Fig-8 Result analysis of Running Time

7. CONCLUSIONS

In this paper, the keyword suggestion system for sponsored search advertisements has been proposed. The objective is to suggest Long-Tail keywords to the advertisers for sponsored search advertisements. We designed the system by using improved topic modeling on Query Log to suggest the Long-Tail keyword to advertisers. The experimental results indicate that the proposed method shows the better results than the topic modeling GibbsLDA based approach proposed by Qiao et al [**5**]. Also, the proposed system speeds up the process as the running time it takes is less than that of the GibbsLDA based approach proposed in [**5**].

REFERENCES

[1] Luís Sarmento, Paulo Trezentos, João Pedro Gonçalves, and Eugénio Oliveira, "Inferring local synonyms for improving keyword suggestion in an on-line advertisement system," in ADKDD '09 Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising, Paris, France, 2009, pp. 37-45.

[2] Amar Budhiraja and P. Krishna Reddy, "An Approach to Cover More Advertisers in Adwords," in 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, France, 2015.

[3] Amar Budhiraja and P. Krishna Reddy, "An Improved Approach for Long Tail Advertising in Sponsored

Search," in International Conference on Database Systems for Advanced Applications, vol. 10178, Suzhou, China, 2017, pp. 169-184.

[4] [Online]. http://neilpatel.com/blog/

[5] Dandan Qiao, Jin Zhang, Qiang Wei, and Guoqing Chen, "Finding competitive keywords from query logs to enhance search engine advertising," vol. 54, no. 4, pp. 531-543, 2017.

[6] Wenhong Luo, David Cook, and Eric J. Karson, "Search advertising placement strategy: Exploring the efficacy of the conventional wisdom," vol. 48, no. 8, pp. 404-411, 2011.

[7] Anton Schwaighofer, Joaquin Quiñonero Candela, Thomas Borchert, Thore Graepel, and Ralf Herbrich, "ADKDD '09 Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising," in Knowledge Discovery and Data Mining, Paris, France, 2009, pp. 27-36.

[8] Ying Zhang, Weinan Zhang, Bin Gao, Xiaojie Yuan, and Tie-Yan Liu, "Bid keyword suggestion in sponsored search based on competitiveness and relevance," vol. 50, no. 4, pp. 508-523, 2014.

[9] ZHI DA, JOSEPH ENGELBERG, and PENGJIE GAO, "In Search of Attention," vol. LXVI, no. 5, pp. 1461-1499, 2011.

[10] Aleksandr Chuklin and Pavel Serdyukov, "How query extensions reflect search result abandonments," in SIGIR '12 Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, Portland, Oregon, USA, 2012, pp. 1087-1088.

[11] Yifan Chen, Gui-Rong Xue, and Yong Yu, "Advertising keyword suggestion based on concept hierarchy," in WSDM '08 Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, California, USA, 2008, pp. 251-260.

[12] Jyun-Yu Jiang, Yen-Yu Ke, Pao-Yu Chien, and Pu-Jen Cheng, "Learning user reformulation behavior for query auto-completion," in SIGIR '14 Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, Gold Coast, Queensland, Australia, 2014, pp. 445-454.

[13] Idan Szpektor, Aristides Gionis, and Yoelle Maarek, "Improving Recommendation for Long-tail Queries," in WWW '11 Proceedings of the 20th international conference on World wide web, Hyderabad, India, 2011, pp. 47-56.

[14] Amruta Joshi and Rajeevi Motwan, "Keyword Generation for Search Engine Advertising," in Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), Hong Kong, China, 2006, pp. 490-496.

[15] Vibhanshu Abhishek and Kartik Hosanagar, "Keyword generation for search engine advertising using semantic similarity between terms," in International Conference on E-Commerce, Minneapolis, MN, USA, 2007, pp. 89-94.

[16] Andrei Z. Broder et al., "Robust classification of rare queries using web knowledge," in SIGIR '07 Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, Amsterdam, The Netherlands, 2007, pp. 231-238.

[17] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet Allocation," vol. 3, pp. 993-1022, 2003.

[18] http://boston.lti.cs.cmu.edu/Data/web08-bst/AOLQs.txt.