Publisher Side Profit Optimization Using Adaptive Keyword Weighted Sponsored Search Technique

Shikha Gupta* and Atul Mishra

JC Bose University of Science and Technology, YMCA, Faridabad, Haryana, India
E-mail: shikha.0909@gmail.com
*Corresponding Author

Abstract

One of the most prominent fields of online advertising is Sponsored search and for various search engines, it acts as one of the main sources of revenue. This paper focuses on sponsored links displayed to the user along with search results when a query is fired by the user. Bidding on keywords is done by the advertiser for the expected future queries and accordingly, payment is done if clicked. A novel technique is proposed in this paper which aims to maximize the revenue earned by a search engine by using an Adaptive keyword weighted approach. Normally, the advertisers focus on keywords with a high frequency which leads to underexplored revenue of search engines. The approach proposed in this paper assigns weight to the keywords based on their winning probability. It also merges the assigned weight with the rarity factor leading to more revenue. With this approach, advertisers with relevant keywords which are rare are explored even if the bid value is low. Experimental results are shown in this paper for proving the improvements over the generalized balance algorithm.

Keywords: Real-time, keyword-based search, sponsored search, keywords, bid term, bid price, bid period, online advertisement.

1 Introduction

Online Marketing

Earlier, there were various sources of advertisements like TV, radio, newspapers, pamphlets which helps in building a connection between the advertiser and the user. Then, with the use of internet emails, display ads, social media were used for displaying the ads (Yuan et al., 2013; Deshwal, 2016). Now, as the use of the internet is increasing day-by-day the whole scenario has changed. Online medium now uses a new advertising model – The Sponsored Search Model (Shin, 2015), which is now used for setting up this connection which makes the relation more strong, trustful and healthy. The advertisers and the users now get connected in a much better way. Online marketing uses a set of powerful methodologies and tools for promoting the products and services of the advertiser and for creating the link. Now, with the increased use of the internet, online marketing is making a huge impact on the choice of users. Online marketing is helping the advertisers in targeting the right audience and is also helping the users to find the most relevant product or service at minimum cost and time. At the same time, it is also helping the publishers in making huge profits. The publishers help in setting the connection between the advertiser and the user (Gupta et al., 2021).

Sponsored Search Advertisement

Sponsored search advertisement, is a framework that involves the advertisers, publishers, and users. It is also called search ads or sponsored links. It helps in setting a stronger connection between these three. In sponsored search, whenever a query is fired or some keyword is searched by the user then, along with organic search results a set of sponsored links are also displayed on an internet search engine as a result of the user query (Qu et al., 2011). These sponsored links are of high relevancy with the query and thus draw the attention of the user towards them (McLaughlin et al., 2016). These sponsored links are of great use for the user and great profit for the publisher and the advertiser. They are one of the major sources of revenue for the publishers, i.e., the search engines (An et al., 2011) for example, Google's ad revenue for the year 2019 amounted to nearly 134.81 billion US dollars out of a total of 160.74 billion US dollars (Johnson, 2021) which is nearly 83%.

The sponsored links are displayed to the user as a result of their query based on the bidding done by them for those keywords (keywords entered in the query). An auction is carried out between the advertisers for selling the ad space (Tsan-Ming et al., 2015; Oberoi, 2013; Walrand, 2007). Moreover, the
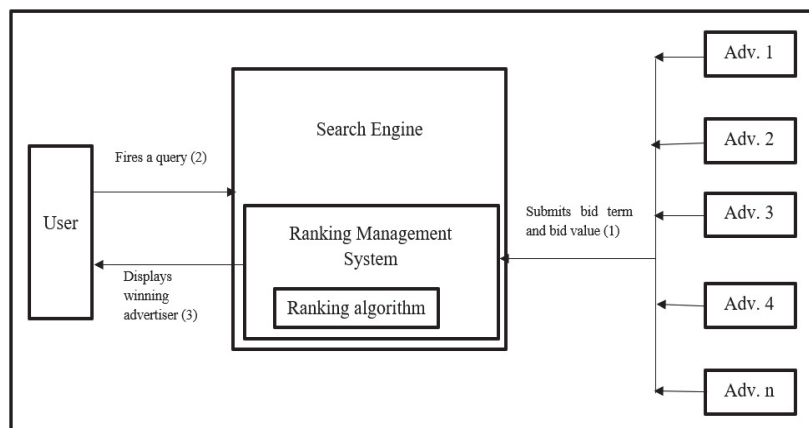
Figure 1 Schematic flow of sponsored search advertisement.

results are displayed by keeping the expected future queries in mind and at the same time by providing the best possible choice to the user (Gupta et al., 2020).

Figure 1 below shows the complete flow for the display of sponsored search. IT can be explained as:

(1) There are a set of advertisers who register themselves with the search engine by specifying the keywords or bid terms (in which they are interested to bid on) and maximum bid amount (maximum amount which they are willing to pay for their advertisement to be displayed for those keywords).

(2) A query is fired by the user on visiting the publisher's website or the search engine.

(3) The ranking algorithm then decides that which advertisement should be displayed to the user, for which various factors are considered like the quality score and the bid amount (Gupta et al., 2020).

In sponsored search advertisements, the ranking of advertisers is the most important concern. Several ranking algorithms are present and are used in various search engines. Earlier, only the bid amount was considered for ranking purposes but it sometimes led to upset users as the relevancy was not considered and it only gives more revenue to the search engine. With time, quality factor was introduced which was based on various factors like quality of landing page, CTR (Click-Through Rate), relevance with the query, and many more (Grbovic et al., 2016; Edelman et al., 2010; Rajaraman et al., 2014).

Motivation and Problem Definition

In sponsored search advertisement, both the advertisers and the publishers make a good profit while matching user queries and keeping in mind future queries as well. In this, the ranking of advertisers is one of the most important success factors. The efficiency depends on the approach used for ranking. While ranking there are several factors are to be considered. Although a lot of research has been done in this area, some research problems are still open. The main motivation behind the proposed algorithm is:

(a) **Keyword Weight:** Selection of keywords by the advertisers is one of the most challenging things. At the same time, all selected keywords are not equally relevant to the advertiser, their importance/ weight may vary (Gupta et al., 2020). So, some weight should be assigned to the keywords indicating their importance for the advertiser. Then, the mapping of the query and the advertiser can be done based on the weight of the matched keywords.

(b) **Explore versus Exploit:** It is one of the major research areas in Sponsored search advertising. Deciding between exploration and exploitation is a difficult task (Ruggiero et al., 2017). In existing approaches, high bid value or high CTR are considered for selecting the advertisers and it leads to exploitation of few advertisers only making it a problem of local maxima. As a result, for mapping future queries, sometimes no relevant advertiser is left with a balanced budget. Therefore, global maxima should be considered that is, the decisions should be taken keeping future queries in mind. Every advertiser must be given a chance to get mapped with the queries and the budget of every advertiser should be utilized in the best possible way. Also, for future queries, there should be relevant advertisers with a balanced budget so that mapping can be done. This can be done with the Exploration of advertisers, i.e., the advertisers with less bid value should also be selected. So, the selection should not be based on bid value only.

(c) **Return on Investment (ROI):** ROI is one of the most important concerns for the advertiser. It helps in calculating the return on the investment (Xia et al., 2016). The advertiser should be able to calculate that from the registered keywords, which is profitable and which is not, or which is the right time to change a keyword if the desired profit is not earned. This suggestion should be given to the advertiser.

(d) **Outliers:** Outliers are the rare or exceptional cases that are generally considered as noise but sometimes they can be useful and should not be ignored. Here, words having very low frequency can be considered as

an outlier and are ignored. However, they should also be considered in the designing of the algorithm.

(e) **Profit to the publisher:** The publisher can earn more and more profit when the number of queries that are getting mapped is high. Therefore, while designing the algorithm, it should be kept in mind.

Contribution and Paper Organization

This paper proposes an algorithm that tries to handle the problems identified in this paper. The proposed approach maps the queries with the advertisers keeping future queries and publisher's profit in mind, and providing the best possible mapping as well.

Section 2 of this paper discusses a brief literature review of sponsored search advertising. Section 3 proposes the new algorithm which will help in handling the above-discussed problems or challenges. Section 4 shows the experimental results of the proposed Keyword Weighted Ranking algorithm and shows its comparison with the existing generalized balance algorithm. In Section 5 a brief conclusion and future scope of the paper are given.

2 Related Work

The main concern in sponsored search advertising is the ranking of advertisers. Several approaches are available which helps in ranking. One of the most widely used models is Adwords (Nagarad, 2019). The advertisers with maximum profit should be selected keeping future queries in mind. The focus is on improving the search results (advertisers) for the queries and on increasing the profit of both advertisers and publishers. Many techniques are available but still, the results can be improved (Gupta et al., 2021).

All the existing approaches keep on exploiting the advertisers with the highest bid amount instead of exploring new promising advertisers (Ruggiero et al., 2017). As a result, the budget of few advertisers gets exhausted and few left untouched.

Tsan-Ming proposed an auction mechanism (bidding) for selling the sponsored links position to the advertisers. But it also suggests that an advertiser may be interested in more than one keyword phrase and the preference for them may vary. All the keywords may not be equally relevant to the advertiser (Tsan-Ming et al., 2015).

All the existing approaches have some issues or challenges that need to be addressed. This paper tries to handle them and propose an algorithm with better efficiency.

3 Proposed Work

This paper proposes an approach for improving the performance of the sponsored search advertising process and for addressing the problems discussed in the previous section. The whole process is divided into small episodes called as bid periods. The system will learn for future queries and in each episode the parameters are updated based on the performance in the previous episode.

Flow Diagram

The complete working of the keyword weighted ranking algorithm is as shown in Figure 2.

The complete workflow can be explained as:

Step 1: All the advertisers register themselves with the submission block by specifying their bid amount and keyword phrases.

Step 2: The registered keyword phrases and the bid amount are then entered into the system schema as shown in Table 2.

Step 3: The weights are then assigned to the stored keyword phrases by the Keyword Weight Assigner.

Step 4: A query is fired by the user in the search engine or on the publisher's site.

Step 5: The query terms are then forwarded to the extractor by the interface where the key terms are extracted from the query.

Step 6: The key terms are then passed to the mapper.
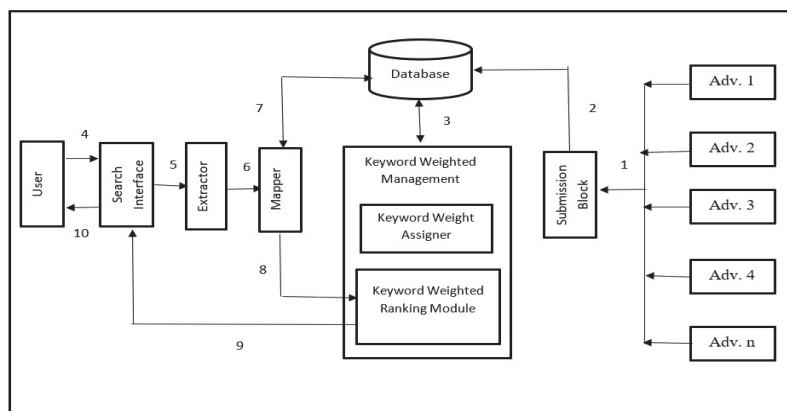
Figure 2 Proposed architecture of a sponsored search advertising.

Step 7: The mapper will then access the available system schema and a list of relevant advertisers is generated.

Step 8: The list generated in step 7 is then forwarded to the keyword weighted ranking module which will then rank those selected advertisers.

Step 9: The highest-ranked advertiser is then informed to the search engine.

Step 10: The advertisement of the winning advertiser is then displayed to the user.

The complete workflow can be explained with the help of pseudocode as in Algorithm 1.

Algorithm 1 Working of sponsored search advertising

Input: Bid terms and bid amount of each advertiser
Output: Most relevant and promising advertiser
1: Assign weights to keyword phrases:

$$w_{ki}^{aj} = O_{ki}^{aj}/S \quad //\text{for first episode only}$$

2: When a query is fired, extract key terms through extractor.
3: Mapper will give list of all relevant advertisers from the database.
4: Keyword weighted ranking algorithm will rank them using:

  4.1 Frequency of each keyword phrase among the advertisers is calculated.
  4.2 Weighted score is calculated for the matched keyword phrase of each relevant advertiser using the formula:

$$Ws_{ki}^{aj} = w_{ki}^{'aj}/f_{ki}$$

  4.3 Quality score of all selected advertisers is calculated by

$$\text{Quality Score} = \text{CTR} * \textit{Weighted Score} * \text{CVR}$$

$$* \text{Percentage of a balanced budget}$$

  4.4 Ranking Score is calculated for the ranking purpose using eq.

$$\text{Ranking Score} = \text{Bid amount} * \text{Quality Score}$$

5: Top most advertiser is selected and displayed to the user.
6: Update the value of Impression Coefficient.
7: Repeat step 4 to 8 for all the queries of that episode.
8: For i = 2 to number of episodes do   //weights are updated in all the episodes

$$\alpha_{ki}^{aj} = I_{ki}^{aj}/I_{t}^{aj}$$

$$w_{ki}^{'aj} = (\text{Weight in the previous episode} + \alpha_{ki}^{aj})/2$$

  goto step 9.

Notations

The various notations used in the work are:

Table 1   Notations used in the proposed approach

$k_i = \{b_i\}\ i = (1, 2, 3, \ldots)$	Keyword phrase which can be a single word or a set of words which an advertiser will specify as a Bid Term where 'i' represents the index of the keyword amongst specified Bid Terms.
w_{ki}^{aj}	Initial weight of k_i i,e., weight for the first episode only for an advertiser$_j$
α_{ki}^{aj}	Impression Coefficient i.e., the contribution of k_i in generating the impressions for a given episode for an advertiser$_j$
w'^{aj}_{ki}	Updated weight of k_i i.e., the new weight of keyword phrase after each episode for an advertiser$_j$
O_{ki}^{aj}	Frequency of occurrence of k_i in the advertisement of a particular advertiser$_j$
$S = \Sigma\, O_{ki}^{aj}\ i = \{1, 2, 3, \ldots\}$	Sum of occurrence of all k_i of a given advertiser$_j$ in his ad
I_{ki}^{aj}	Impressions generated for a given advertiser$_j$ due to k_i
$I_t^{aj} = \Sigma\, I_{ki}^{aj}\ i = \{1, 2, 3, \ldots\}$	Total no. of impressions generated by all the k_i of an advertiser$_j$
$fki = \Sigma k_i^{aj}\ j = \{1, 2, 3, \ldots\}$	Frequency of occurrence of k_i as a Bid Term among stall the advertisers i.e., count of advertisers having k_i as a Bid Term.
Ws_{ki}^{aj}	Weighted Score, which is the ratio of weight and frequency of the keyword phrase k_i because of which an advertiser$_j$ is selected, it is used during the ranking of advertisers.

Working

The complete work can be explained in three parts: Keyword Weight Management, Query Management, Keyword Weighted Ranking Module.

(a) *Keyword Weight Management:* At the beginning, all the advertisers register themselves with the search engine/ publisher by specifying their keyword phrases (bid terms) and the bid amount (the maximum amount that they are willing to pay for the display of their advertisement). Then, the system will assign weights to all the registered keyword phrases.

The keyword weight shows the usefulness of the keyword phrase or contribution in winning for the advertiser. Also, for a given advertiser, the

sum of weights of all the registered keyword phrases should be 1. In the beginning, the weight to each keyword phrase of an advertiser is assigned by dividing the frequency of occurrence of that particular keyword in the advertisement by the sum of the frequency of occurrence of all the keywords in the advertisement, as shown in Equation (1).

$$w_{ki}^{aj} = O_{ki}^{aj}/S \qquad (1)$$

After running an episode, the weight of each keyword phrase is updated for the next episode based on their performance. The updation is done by using the Impression Coefficient, which is calculated for each keyword phrase of every advertiser by using the ratio of impressions because of that particular keyword phrase divided by total impressions of that advertiser as shown in Equation (2):

$$\alpha_{ki}^{aj} = I_{ki}^{aj}/I_{t}^{aj} \qquad (2)$$

The updated weight is calculated for the next episode using Equation (3):

$$w_{ki}^{'aj} = (\text{Weight in the previous episode} + \alpha_{ki}^{aj})/2 \qquad (3)$$

The pseudocode for the above discussed module can be explained in Algorithm 2.

Algorithm 2 Keyword weighted management algorithm

Input: Registered keyword phrases of each advertiser
Output: Weights of all the registered keyword phrases.
1: Assign weights to keyword phrases using:

$$w_{ki}^{aj} = O_{ki}^{aj}/S \quad //\text{for first episode only}$$

2: For i = 2 to number of episodes do   //weights are updated at the beginning of all the episodes
3: Update weights of all the keyword phrases

$$\alpha_{ki}^{aj} = I_{ki}^{aj}/I_{t}^{aj}$$

$$w_{ki}^{'aj} = (\text{Weight in the previous episode} + \alpha_{ki}^{aj})/2$$

4: End for loop

(b) *Query Management*: After assigning the weight to each keyword phrase, an episode starts. During an episode, several queries will arrive. Queries

are then mapped to the advertiser keeping in mind the revenue earning for the publisher/search engine. While mapping queries, advertisers are explored who are having rare keywords or outliers. The weighted score, Ws is calculated for all relevant advertisers, using the ratio of weight and rarity of the matched keyword phrase. The weight of the keyword phrase is already discussed in the previous section. Preference is given to the advertisers having the least common terms as keyword phrases therefore, rarity is used. This will help in assigning more queries in the future as the advertisers having more frequent keyword phrases can be mapped later as well. The weighted score is calculated as in Equation (4):

$$\boxed{\mathrm{Ws}_{ki}^{aj} = w_{ki}^{\prime aj}/f_{ki}}$$

(4)

This can be explained below in Algorithm 3.

Algorithm 3 Query management algorithm

Input: (i) Weight of keyword phrases calculated in Algorithm 1.
    (ii) Keyword phrases of all the advertisers.
Output: Weighted Score of all the relevant advertisers.
1: Frequency of each keyword phrase among the advertisers is calculated.
2: A query is fired and relevant advertisers are extracted.
3: Weighted score is calculated for the matched keyword phrase of each relevant advertiser using the formula:
$$\mathrm{Ws}_{ki}^{aj} = w_{ki}^{\prime aj}/f_{ki}$$

(c) *Keyword Weighted Ranking Module*: The outputs of the above discussed algorithms are then used for the ranking purpose. The relevant advertisers are ranked by the Keyword Weight Ranking module using the ranking score calculated for each advertiser (Yuan et al., 2013; Deshwal, 2016) as discussed below in Algorithm 4:

$$\boxed{\mathrm{Ranking\ Score} = \mathrm{Bid\ amount} \ast \mathrm{Quality\ Score}}$$

(5)

where,

$$\boxed{\begin{aligned}\mathrm{Quality\ Score} = \mathrm{CTR} \ &\ast\ \textit{Weighted Score} \ \ast\ \mathrm{CVR} \\ &\ast\ \mathrm{Percentage\ of\ a\ balanced\ budget}\end{aligned}}$$

(6)

Here,

CVR: Conversion Rate
CTR: Click-Through Rate

Also, this approach uses the percentage of the balanced budget while ranking the advertisers. It was introduced in the Generalized Balance Algorithm (Nagarad, 2019).

Using the Ranking Score value, the advertisers are ranked and top advertisers are then displayed to the user for the query. Also, whether the advertisement is clicked or not, whether the product is purchased or not, are tracked and updations are done in the schema.

Algorithm 4 Keyword weighted ranking algorithm

Input: (i) Bid terms and bid amount of each advertiser
      (ii) Weighted Score of all the relevant advertisers.
Output: Most relevant advertiser.
1:   Quality score of all selected advertisers is calculated by

$$\text{Quality Score} = \text{CTR} * \textit{Weighted Score} * \text{CVR}$$
$$* \text{ Percentage of a balanced budget}$$

2:   Ranking Score is calculated for the ranking purpose using eq.

$$\text{Ranking Score} = \text{Bid amount} * \text{Quality Score}$$

3:   Top most advertiser is selected and displayed to the user.
4:   Update the value of Impression Coefficient.

The process discussed above repeats for every query and the database is updated. After execution of an episode, weights are updated for the next episode based on their impressions in the previous one using Equations (1) and (2) discussed above. The proposed algorithm also helps in identifying the keyword phrases whose ROI (Return on Investment) is not good enough. It does so by calculating the impression count, if it is less than a particular threshold value then, it suggests the advertiser to change it.

The schema for maintaining the database of the above-proposed system is shown below in Table 2. It collects the record of all the registered advertisers by storing their keyword phrases (bid terms) and the bid amount. Along with it, the CTR, total budget, a balanced budget, Quality score, Effective click-through rate, and CVR are also recorded. After executing an episode, the values are updated for making decisions for future queries in the next episode.

Table 2   Database of keyword weighted ranking management system

List of Advertised	List of k_i for Each Advertiser	Weight of Each k_i	Frequency of k_i	Impressions of Each k_i	Bid Price of Each Advertiser	Total Budget of Each Advertiser	Click-Through Rate	Balance Budget	CVR
A1	k1	w1	F_{k1}	I1	Bid1	B1	CTR1	BB1	CVR1
	k2	w2	F_{k2}	I2					
	k3	w3	F_{k3}	I3					
A2	k4	w4	F_{k4}	I4	Bid2	B2	CTR2	BB2	CVR2
	k5	w5	F_{k5}	I5					
	k6	w6	F_{k6}	I6					

4 Experimental Results and Discussions

The efficiency of the above-proposed approach can be proved using an experimental setup. Also, the results of the generalized balance algorithm and the proposed keyword-weighted ranking algorithm are compared to show the effectiveness.

Dataset

For the experimental purpose, a dataset is created with 10 different items of sale. The items of sale considered are computers, cellphone, furniture, fashion, gift packs, travel, home décor, baby products, kitchen appliances, and cosmetics. A set of 100–150 keywords for each item of sale is taken as shown in Table 3. To each advertiser, relevant keywords in the range 1–4 are randomly assigned. The database of each advertiser is created which includes fields like their keywords, the weight of the keywords, frequency of keywords among the advertisers, impressions of each keyword, bid price, total budget, click-through rate, balance budget, and conversion rate of each advertiser as discussed below in Table 4.

Evaluation Methodology

Since the goal of our proposed work is to increase the profit of the publisher by mapping more and more queries, we used Keyword Weighted Ranking Algorithm to do so. Keeping some features the same (which have the same impact for both the approaches), the algorithm is run for the generalized balance algorithm and the keyword weighted ranking algorithm. Both the approaches were parallelly run using the same set of data (common features) and the results are compared. The complete run can be explained as:

Step 1: Randomly a set of keyword phrases (in the range 1–4) for all the advertisers is taken from the set of associated keywords.

Table 3   List of keywords in various items of sales

Items of Sale	Associated Keywords
Computer	Laptop, pc, desktop, laptop price, laptop deals, desktop computer, laptop for sale, gaming pc, remote desktop, computer shop near me, computer price, apple mac, apple imac, imac pro, hp laptop price, apple pc, sony laptop, dell desktop, computer store, computer shop and many more.
Cell Phone	Cell phone store near me, phone accessories, cellular near me, mobile phones near me, best mobile phone 2020, best cellphone, Samsung cell phones, cell phones for sale, cell phones near me, best cell phone 2020, best buy cell phones, cheap cell phones, mobile phone shop near me, cell phone stores, LG cellphone, cellphone price, mobile phone shop, new cell phones, motorolla cell phones, cellphone price and many more.
Furniture	Sofa, furniture stores near me, table, furniture stores, bed, couch, office chair, tv stand, sofa bed, chair, dining table, computer desk, coffee table, sofa set, dressing table, desk chair, study chair, study table, desk, king size bed and many more.
Fashion	Clothing stores near me, nykaa fashion, plus size clothing, outfits, street style store, fashion designer, wholesale clothing, Sabyasachi, fast fashion, fashionista, mens fashion, online clothing stores, womens clothes, mens clothes, clothing brands, fashion house, cheap clothes, cute clothes, winter clothes, fashion clothing and many more.
Gift Packs	Gift sets, hot chocolate gift sets, gift set for women, baby gift sets, gift set for men, gift packing ideas, gift packing, birthday care package, chocolate box packaging, mens skincare gift set, Christmas gift set, newborn baby gift set, gift box packaging, dry fruits gift pack, gift packing box, gift set for him, couple gift set, fragrance gift set, birthday gift sets, gift sets for women and many more.
Travel	Trip, places to visit near me, road trip, luggage, train ticket booking, bus booking, tourist places near me, google travel, international travel, travel agency, travel bag, vacation packages, my trip, beautiful places, air india flight booking, online bus booking, train ticket booking online, railway booking, travel agency near me, places to go and many more.
Home Decor	Wall décor, décor, wall mirror, home décor stores, home interior design, Ganpati decoration, home décor items, living room décor, interior design styles, home interior, decorating, framed wall art, interior decoration, cheap home décor, office décor, house décor, home décor near me, wall art painting, modern home décor, home goods décor, home office décor and many more.
Baby Products	Baby shampoo, diaper bag bagpack, crib, baby swing, baby toys, baby carrier, pacifier, baby bouncer, baby wipes, baby food, bibs, baby cot, baby crib, swaddle, baby oil, baby dress, baby bed, baby car seat, diaper bag, baby stuff and many more.

(*Continued*)

Table 3   Continued

Items of Sale	Associated Keywords
Kitchen Appliances	Air fryer, microwave, dishwasher, food processor, oven, microwave oven, best air fryer, juicer, bread maker, hand blender, electric stove, smoothie maker, double door fridge, juicer machine, oven price, home appliances, Samsung microwave, electric oven, electric cooker, best food processor and many more.
Cosmetics	Makeup, makeup brushes, eye makeup, makeup set, primer makeup, makeup products, best makeup brushes, makeup palette, permanent makeup, natural makeup, makeup shop, highlighter makeup, kids makeup, face powder, foundation makeup, cheap makeup, fashion cosmetics, cosmetic shop near me, makeup shop, makeup pouch and many more.

Table 4   Generation of various fields of dataset

Fields	Explanation
List of k_i for each advertiser	To each advertiser, relevant keywords in the range 1–4 are randomly assigned.
Weight of each k_i	Initially, equal weight is assigned to all keywords. After an episode, the weight of the keywords is updated using Keyword Weighted Management Algorithm.
Frequency of k_i	The frequency of the keyword phrases among the advertisers is calculated using Query Management Algorithm.
Impressions of each k_i	Impressions of all the keywords are also recorded and updated with every query.
Bid Price, Total budget, CTR, CVR of each advertiser	These fields are randomly assigned to the advertisers because their impact is the same for both cases.
Balance Budget	Updated after every selection of the advertiser for a query.

Step 2: Weights are assigned to them using Algorithm 2. Initially, equal weights are assigned to them but later the weights are updated using Equation (3).

Step 3: After assigning the weights, Algorithm 3 is called. In this, the frequency of each keyword phrase among the advertisers is calculated.

Step 4: Now, random queries are generated and mapping is done for both approaches.

Step 5: A weighted score is calculated for the advertisers using Equation (4). This weighted score is then used for the ranking purpose in Keyword Weighted Ranking Algorithm as discussed above in Equations (5) and (6).

Step 6: Algorithm 4 then gives the highest ranked advertiser using Keyword Weighted Ranking Algorithm which is then displayed to the user. The highest-ranked advertiser using the Generalized Balance Algorithm is also calculated parallelly using Equations (7) and (8).

$$\boxed{\text{Ranking Score} = \text{Bid amount} * \text{Quality Score}} \quad (7)$$

where,

$$\text{Quality Score} = \text{CTR} * \text{CVR} * \text{Percentage of a balanced budget}$$
$$(8)$$

This process continues for each query and, a few factors like impressions and a balanced budget are updated. After the completion of an episode, the weights of the keyword phrases are updated based on the impressions using Equation (3).

The dataset is run for various cases like 10 episodes of 10 queries each, 10 episodes of 50 queries each, 10 episodes of 100 queries each, 100 episodes of 10 queries each, 100 episodes of 50 queries each, and 100 episodes of 100 queries each. The results show the difference in performance because of newly added features, i.e., weight, frequency, and impression of each keyword phrase.
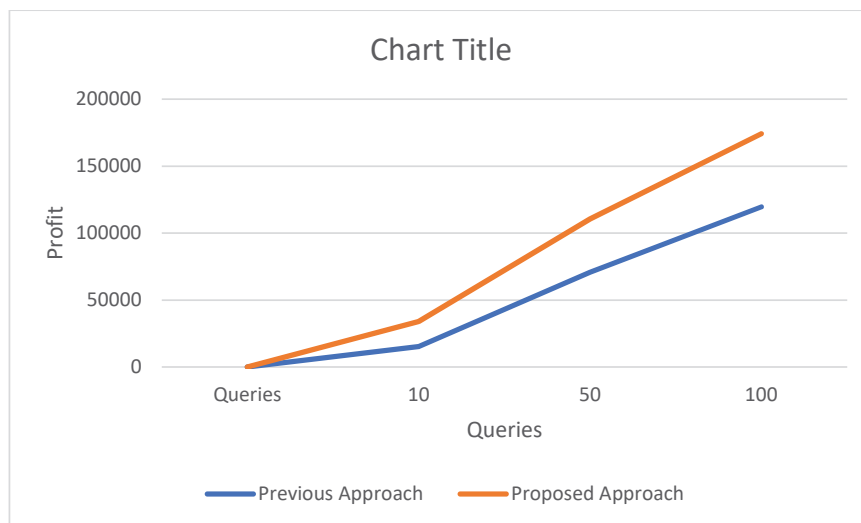
Performance Comparison

Table 5 below shows the difference in profit earned using the Generalized balance algorithm and, Keyword weighted ranking algorithm while running ten episodes of 10, 50, and 100 queries each.

Figure 3 clearly shows that there is a marginal increase in profit using the Keyword weighted ranking algorithm. The graph clearly shows that initially there is not much difference as in both the approaches the relevant advertisers with the balanced budget were available. But slowly, when the budget starts

Table 5   Profit earned using Generalized balance algorithm and Keyword weighted ranking algorithm in 10 episodes

Episodes	Queries	Profit from Generalized Balance Algorithm	Profit from Keyword Weighted Ranking Algorithm
10	10	15242	33961
10	50	70855	110498
10	100	119523	174170

Figure 3   Comparison of Generalized balance algorithm and Keyword weighted ranking algorithm while running 10 episodes.

getting exhausted then in the Generalized balance algorithm few relevant advertisers with the remaining budget were left. As a result, fewer queries get mapped leading to less profit. Whereas in the case of the proposed Keyword weighted ranking algorithm, future queries were considered while ranking. Therefore, many relevant advertisers with the remaining budget were left. As a result, more queries get mapped to increase the profit earned.

Table 6 and Figure 4 below shows the increase in profit while running 100 episodes of 10, 50 and 100 queries.

Figures 3 and 4 shows the efficiency of the proposed algorithm over the existing generalized balance algorithm. The profit can be increased by using the above proposed algorithm.

Table 6   Profit earned using Generalized balance algorithm and Keyword weighted ranking algorithm in 100 episodes

Episodes	Queries	Profit from Generalized Balance Algorithm	Profit from Keyword Weighted Ranking Algorithm
100	10	52843	96362
100	50	69221	103977
100	100	93827	155566
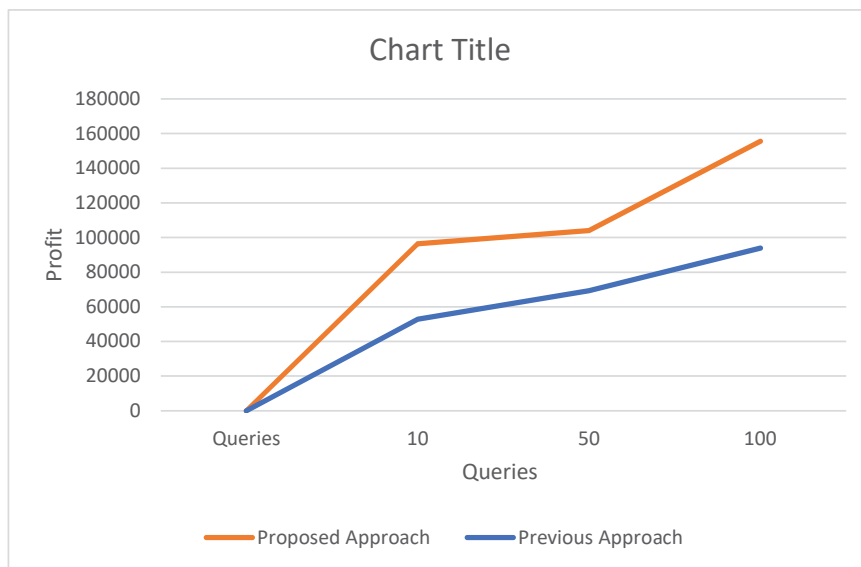
Figure 4 Comparison of Generalized balance algorithm and Keyword weighted ranking algorithm while running 100 episodes.

Conclusion

Sponsored Search Advertisement is one of the most important area in Online Marketing. It is gaining lots of interest from the researchers in recent years. A lot of research has been done but still the performance can be improved. In this paper, a novel approach has been proposed to provide better results to the users and at the same time to increase the profit of the publishers. The Keyword Weighted Ranking Algorithm uses two additional factors while ranking. Firstly, the weight of keyword phrases which is calculated using Keyword Weight Management Algorithm. It helps in identifying the relevance or importance of the registered keyword phrases for the advertiser. Secondly, the frequency of keyword phrases calculated using Query Management Algorithm. It gives preference to advertisers having rare words as keyword phrases, which helps in mapping more queries in the future. The efficiency of the proposed algorithm is evaluated and compared with the existing generalized balance algorithm approach and remarkable improvement is seen. Both mathematical and experimental results show the increased efficiency over the Generalized Balance Algorithm.

In the future, the effectiveness can be checked on various other datasets as well. Also, the performance can be compared by integrating other parameters as well.

References

An, W., Park, C., Han, X., Pattipati, K.R., Kleinman, D.L., and Kemple, W.G. (2011). Hidden Markov Model and Auction-Based Formulations of Sensor Coordination Mechanisms in Dynamic Task Environments. *IEEE Transactions on Systems, Man, and Cybernetics*, Part A. 41. 1092–1106. Doi: 10.1109/TSMCA.2011.2114342.

Bateni, M.H., Feldman, J., Mirrokni, V. and Wong, S.C. 2014. Multiplicative bidding in online advertising. In *Proceedings of the fifteenth ACM conference on Economics and computation (EC '14). Association for Computing Machinery*, New York, NY, USA, 715–732. DOI: https://doi.org/10.1145/2600057.2602874

Deshwal,P. (2016). Online advertising and its impact on consumer Behavior. *International Journal of Applied Research*; 2(2): 200–204, ISSN Print: 2394–7500.

Devanur,N.R. and Hayes,T.P. 2009. The adwords problem: online keyword matching with budgeted bidders under random permutations. In *Proceedings of the 10th ACM conference on Electronic commerce (EC '09). Association for Computing Machinery*, New York, NY, USA, 71–78. DOI: https://doi.org/10.1145/1566374.1566384

Edelman, B. and Schwarz, M. 2010. Optimal Auction Design and Equilibrium Selection in Sponsored Search Auctions. American Economic Review, *American Economic Association*, vol. 100(2), pages 597–602.

Ghose, A. and Yang, S. (2007). An Empirical Analysis of Search Engine Advertising: Sponsored Search and Cross-Selling in Electronic Markets. NET Institute, *Working Papers.*

Grbovic, M., Djuric, N., Radosavljevic, V., Silvestri, F., Baeza-Yates, R., Feng, A., Ordentlich, E., Yang, L. and Owens, G. (2016). Scalable Semantic Matching of Queries to Ads in Sponsored Search Advertising. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR '16). Association for Computing Machinery*, New York, NY, USA, 375–384. DOI: https://doi.org/10.1145/2911451.2911538

Gupta, S. and Mishra, A. (2020). A novel approach for improving Real Time Bidding Strategy. *Materials Science and Engineering Conference Series*. Vol. 804, no. 1, p. 012043. doi: 10.1088/1757-899X/804/1/012043.

Gupta, S. and Mishra, A., Big Data Analytics: Approaches and Applications in Online Marketing (July 12, 2021). Available at *SSRN*: https://ssrn.com/abstract=3884656 or http://dx.doi.org/10.2139/ssrn.3884656

Gupta, S., and Mishra, A. (2021). Evolution of Online Marketing Tools, Approaches, and Strategies With Associated Challenges: A Survey. *International Journal of Technology Diffusion (IJTD)*, 12(3), 61–82. http://doi.org/10.4018/IJTD.2021070104

Hang, L. (2011). A Short Introduction to Learning to Rank. *IEICE Transactions*. 94-D. 1854–1862. 10.1587/transinf.E94.D.1854.

Johnson,J. (2021) *Advertising revenue of Google from 2001 to 2020.* https://www.statista.com/statistics/266249/advertising-revenue-of-google/

Klapdor,S., Anderl, E.M., Wangenheim, F.V. and Schumann, J.H. 2014. Finding the Right Words: The Influence of Keyword Characteristics on Performance of Paid Search Campaigns. Journal of Interactive Marketing, Elsevier, vol. 28(4), pages 285–301. https://doi.org/10.1016/j.intmar.2014.07.001.

McLaughlin, K. and Friedman, D. (2016). Online ad auctions: An experiment. *ESI Working Paper* 16-05. http://digitalcommons.chapman.edu/esi_working_papers/183/

Nagarad,S. (2019). *Google Adwords*. https://www.digitalvidya.com/blog/what-is-google-adwords/

Oberoi A. (2013) *The History of Online Advertising*. https://www.adpushup.com/blog/the-history-of-online-advertising/

Qu, Y. and Cheng, G. (2011). Falcons Concept Search: A Practical Search Engine for Web Ontologies. *IEEE Transactions on Systems, Man, and Cybernetics*, Part A. 41. 810–816. 10.1109/TSMCA.2011.2132705.

Rajaraman, A., Leskovec, J. and Ullman, J. (2014). *Mining of Massive Datasets*. 10.1017/CBO9781139058452.

Ribeiro-Neto, B., Cristo, M., Golgher, P.B. and Moura, E.S. 2005. Impedance coupling in content-targeted advertising. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05). Association for Computing Machinery*, New York, NY, USA, 496–503. DOI: https://doi.org/10.1145/1076034.1076119

Ruggiero, C., Prabhakar, K., Maxim, S. and Christopher,A. (2017). Sponsored Search Auctions with Rich Ads. ACM 978-1-4503-4913-0/17/04. *International World Wide Web Conference Committee (IW3C2)*, WWW 2017, April 3–7, 2017, Perth, Australia.

Rusmevichientong, P. and Williamson, D.P. 2006. An adaptive algorithm for selecting profitable keywords for search-based advertising services. In *Proceedings of the 7th ACM conference on Electronic commerce* (EC '06). *Association for Computing Machinery*, New York, NY, USA, 260–269. DOI: https://doi.org/10.1145/1134707.1134736

Shin, W. (2015). Keyword Search Advertising and Limited Budgets. Marketing Science, *INFORMS*, vol. 34(6), pages 882–896. DOI: 10.1287/mksc .2015.0915.

Tsan-Ming, C., Xun, L. and Cheng, M. (2015). Search-Based Advertising Auctions With Choice-Based Budget Constraint. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 45, No. 8.

Walrand, J. (2007). An efficient Nash-implementation mechanism for network resource allocation. Automatica. 46. 1276–1283. 10.1016/j.auto matica.2010.05.013.

Yuan, S., Wang, J. and Zhao, X. (2013). Real-time bidding for online advertising: measurement and analysis. https://arxiv.org/abs/1306.6542

Yuan, S., Wang, J., Chen, B., Mason,P. and Sam Seljan. 2014. An empirical study of reserve price optimisation in real-time bidding. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14). Association for Computing Machinery*, New York, NY, USA, 1897–1906. DOI: https://doi.org/ 10.1145/2623330.2623357

Biographies

Shikha Gupta is an Assistant Professor in the Department of Computer Engineering, JC Bose UST, YMCA, Haryana, India and is currently pursuing her Ph.D. She holds a Bachelors's and Master's degree in Computer Engineering. Her areas of interest include Big Data, Online Marketing, and Page Ranking. She has published various research papers in this area in international journals and Conferences.

Atul Mishra is working as a Professor in the Department of Computer Engineering, JC Bose UST, YMCA, Haryana, India. He holds a Ph.D. in Computer Science and Engg. He has more than16 years of work experience in the Optical Telecommunication Industry specializing in Optical Network Planning and Network Management tools development in North America. He has a USA patent in his name in optical network management protocol. His research interests include SOA, Network Management & Mobile Agents, Big data, and Cloud computing. He has published more than 30 research papers in national and international journals.