

Алгоритм размещения рекламных объявлений над результатами поиска, максимизирующий доход поисковой системы

А.Н.Сорокина

Национальный исследовательский университет Высшая школа экономики, Москва, Россия
Поступила в редколлегию 25.03.2014

Аннотация—Важной задачей для поисковой интернет-системы является отбор рекламных объявлений для показа над результатами поиска. От этого выбора зависит лояльность рекламодателей, которые эти рекламные объявления помещают в систему, а так же пользователей, которые реагируют на релевантность предоставляемой им информации. В статье ставится и решается задача оптимального выбора рекламных объявлений для показа, максимизирующего доход поисковой системы с ограничениями на среднюю кликабельность и долю запросов с рекламой над результатами поиска.

КЛЮЧЕВЫЕ СЛОВА: оптимизация, алгоритм размещения рекламы, поисковая система.

1. ВВЕДЕНИЕ

В настоящий момент огромная аудитория пользуется интернетом для поиска информации. Обычно Интернет-пользователь задает некоторый *запрос*, на который он хочет получить ответ, и задача поисковых систем — отвечать на эти запросы наиболее полно. Чтобы иметь *доход* от поиска информации для пользователя, поисковик использует рекламные объявления (*баннеры*), которые показываются на странице поисковой выдачи. Среди рекламных объявлений особое значение имеет так называемое “*спец-размещение*” (привилегированные позиции в показе объявлений, расположенные, как правило, сверху над результатами поиска). *Клик пользователя* по рекламному объявлению, расположенному в спец-размещении, приносит поисковой интернет-системе доход. Оптимальный подбор объявлений для показа в спец-размещении — важная практическая задача для современных поисковых интернет-систем. Ее решение — в интересах и других сторон: *рекламодателей* и самих *пользователей*, поскольку первые нуждаются в привлечении потребителей, а вторые — в релевантной их запросу информации.

Основными характеристиками рекламного объявления являются вероятность того, что пользователь кликнет на предъявленное объявление (кликабельность — *CTR, Click – Through – Rate*) [1] и ставка (*Bid*), назначенная рекламодателем за показ его объявления над результатами поиска.

В научной литературе проблематика выбора наилучшего критерия отбора рекламного объявления стала активно изучаться последние 10–15 лет. Встречается достаточно большое количество работ с различными критериями: классический *CPM (Cost – Per – Million) = CTR · Bid* [2], *bid · quality + v* [3], *релевантность* [4] – [6], информация о *продавце* и *бренде* [7], *позиционность* [8], информация о *пользователе* [9], *bid^k · CTR^l* [10].

Ни в одной из представленных в литературе работ не уделяется достаточное внимание проблеме разработки методов подбора порогов, математической постановке задачи оптимизации и её решению.

Таким образом, существует проблема усовершенствования метода отбора рекламных объявлений в спец-размещение. В целом, сложность состоит в том, что решение показа рекламного объявления на запрос пользователя требует учета большого количества факторов и переменных, которые сложно между собой взаимосвязаны. Решение этой важной и сложной научно-практической проблемы требует применения современного математического аппарата, а именно — постановки и решения задачи оптимизации, что и было сделано в представленной статье.

2. ПОСТАНОВКА ЗАДАЧИ ОПТИМИЗАЦИИ РЕКЛАМНЫХ ПОКАЗОВ

Поставим задачу следующим образом: как для заданной выборки запросов и заданного списка баннеров найти оптимальную расстановку этих баннеров в спец-размещении для показа по этим заданным запросам?

Считаем, что ставки (Bid) для всех баннеров известны, а также известны оценки вероятности клика (CTR) для каждой пары “запрос-баннер” (если данный баннер в принципе не может быть совмещен с определенным ответом на запрос, то полагаем, что оценка CTR равна нулю). Считаем также, что заданы: *минимальная средняя кликабельность* (CTR_{min}), *покрытие* — доля результатов поиска, сопровождаемых спец-размещением (Hit_{max}), а также *максимальное число баннеров* (k) в каждом хите по спец-размещению. В статье [11] рассмотрена математическая постановка задачи оптимизации в случае максимизации средней кликабельности при ограничении на суммарный доход, покрытие и количество объявлений в рекламном блоке. В данной статье поставим двойственную задачу, а именно: при этих ограничениях будем искать такую расстановку баннеров в спец-размещении, которая обеспечивает *максимум суммы денежных средств*, которая должна поступить от рекламодателей.

Задачу будем решать в том приближении, когда рекламодатель платит за клик по своей ставке, а не по правилу аукциона второй цены.

Полагаем, что в нашей выборке имеется M запросов, которые мы будем нумеровать индексом j ($1 \leq j \leq M$), и задан список из N баннеров, нумеруемых индексом i ($1 \leq i \leq N$). Пусть B_j — количество баннеров-кандидатов на j -й запрос.

Обозначим далее:

Bid_i — ставка, назначенная рекламодателем за i -тый баннер, $Bid_i \geq 0$.

CTR_{ij} — оценка вероятности клика при размещении i -ого баннера над результатами поиска на j -тый запрос. Величины Bid_i и CTR_{ij} считаем заданными.

Введем далее бинарную переменную t_{ij} : $t_{ij} = 1$ означает, что i -й баннер размещен над результатом поиска на j -й запрос, и $t_{ij} = 0$ — в противном случае.

Тогда математическая задача запишется как:

$$\sum_{(i,j)} Bid_i \cdot CTR_{ij} \cdot t_{ij} = \sum_{(i,j)} CPM_{ij} \rightarrow \max \quad (1)$$

при ограничениях:

$$\sum_{(i,j)} CTR_{ij} \cdot t_{ij} / Events \geq CTR_{min}, \quad (2)$$

$$\sum_j \mathbb{1}[\sum_i t_{ij} \geq 0] / M \leq Hit_{max}, \quad (3)$$

$$\forall j \sum_i t_{ij} \leq k, \quad (4)$$

где $Events = \sum_{(i,j)} t_{ij}$ — общее количество показов рекламных объявлений в спец-размещении. Таким образом построена математическая задача оптимизации показов рекламных объявлений над результатами поиска с целью максимизации дохода поисковой системы при заданных ограничениях на долю запросов со спец-размещением, а так же ограничением на средний CTR показанных объявлений и количество объявлений в одном рекламном блоке.

3. ПОСТРОЕНИЕ ОГИБАЮЩЕЙ $CPM(CTR)$ ДЛЯ КАЖДОГО ЗАПРОСА

Поскольку функция (1), которую мы максимизируем, является суммой CPM_{ij} для всех запросов j , то целесообразно рассматривать каждый запрос отдельно. Предварительно проведём анализ одного запроса — определим зависимость суммарного дохода CPM_j и средней кликабельности CTR_j при всех возможных размещениях рекламных объявлений в рекламном блоке. В настоящий момент в большинстве поисковых систем параметр k в ограничении (4) принимает значение 3, т.е. на каждый поисковый запрос может быть показано не более чем три рекламных объявления. Тогда возможны следующие варианты размещения:

1. Все возможные варианты показа баннеров-кандидатов для показа по данному запросу при полном заполнении спец-размещения, т.е. $k_j = 3$ (везде где это возможно); таких вариантов будет $C_{B_j}^3$ (при $B_j = k_j$ это будет только один вариант).
2. Все возможные варианты размещения при неполном заполнении рекламного блока (два объявления в спец-размещении, $k_j = 2$); таких вариантов будет $C_{B_j}^2$.
3. Все возможные варианты размещения при показе только одного объявления, $k_j = 1$; таких вариантов будет B_j .

Таким образом, на один запрос всего необходимо будет перебрать максимум $V = \mathcal{O}(B_j^3)$ вариантов размещений рекламных объявлений над результатами поиска; будем нумеровать каждый вариант с помощью индекса v ($1 \leq v \leq V$); набор баннеров в варианте обозначим B_{jv} , и для них положим $t_{ij} = 1$ (i — номер банера из B_{jv}).

Для каждого варианта размещения баннеров для показа в спец-размещении v вычисляем общий доход и среднюю кликабельность баннера запроса:

$$CPM_{jv} = \sum_{i \in B_{jv}} t_{ij} \cdot CTR_{ij} \cdot Bid_i,$$

$$CTR_{jv} = \sum_{i \in B_{jv}} t_{ij} \cdot CTR_{ij} / k_{jv},$$

где k_{jv} — количество объявлений, показанных в спец-размещении при выборе определённого варианта v размещения. После того, как для каждого из вариантов размещения баннеров получены значения CPM_{jv} и CTR_{jv} , можно для них построить огибающую по формуле:

$$\forall CTR_{jv} : eCPM_{jv}(CTR_{jv}) = \max_{CTR: \{CTR=CTR_{jv}\}} CPM_{jv}(CTR), \quad (5)$$

т.е. для каждого фиксированного значения $CTR = const$ выбирается соответствующее ему максимальное значение CPM_{jv} (см. Рис. 1).

Теперь для огибающих $CPM_{jv}(CTR_{jv})$ построим аппроксимацию (приближение полиномом); таким образом мы уходим от дискретной задачи к непрерывной: для каждого значения CTR можно найти соответствующее ему значение CPM , что важно для дальнейшего решения оптимизационной задачи.

Для каждого запроса можно вычислить $mCTR_{jv}$, соответствующее максимуму $mCPM_{jv}$, т.е. $mCTR_{jv}$ — кликабельность, при которой достигается максимальный доход по данному запросу Рис.2. $mCTR_{jv}$ нам понадобится для решения задачи оптимизации. Если максимум не единственный, то для запроса выбирается та точка, где $mCTR_{jv}$ максимальный.

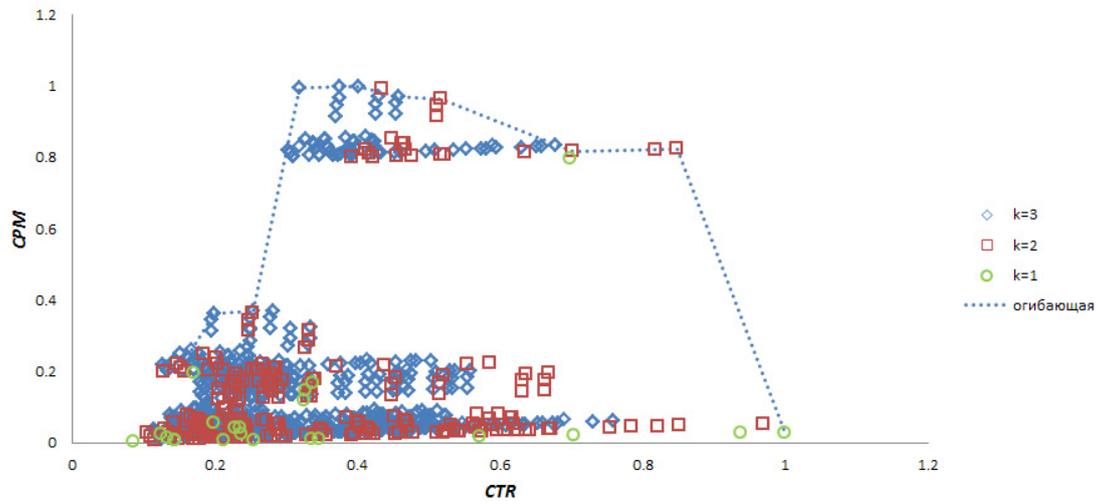


Рис. 1. Огибающая точек $CPM_{jv}(CTR_{jv})$ для различных вариантов отбора объявлений

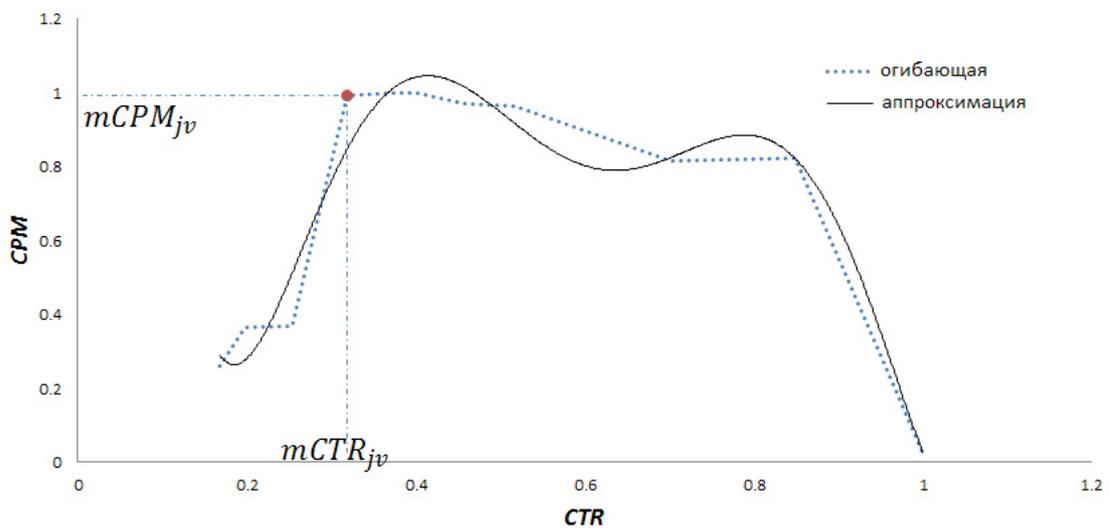


Рис. 2. Аппроксимация огнибающей точек $CPM_{jv}(CTR_{jv})$, нахождение $mCTR_{jv}$

4. РЕШЕНИЕ ЗАДАЧИ ОПТИМАЛЬНОГО РАЗМЕЩЕНИЯ РЕКЛАМНЫХ ОБЪЯВЛЕНИЙ

Работа алгоритма начинается со следующих начальных характеристик системы:

- максимальная доходность для спец-размещения:

$$CPM_{max} = \sum_{j=1}^M mCPM_{jv},$$

- соответствующая максимальной доходности средняя кликабельность:

$$CTR(CPM_{max}) = \sum_{j=1}^M mCTR_{jv}/M.$$

Если $CTR(CPM_{max}) \geq CTR_{min}$ (2), то мы нашли оптимальное размещение рекламных объявлений в спец-размещении. Однако численные эксперименты показывают, что обычно $CTR(CPM_{max})$ значительно меньше нужного значения CTR_{min} . Тогда мы хотим увеличивать среднюю кликабельность с шагом ΔCTR и искать такие запросы, по которым при изменении на ΔCTR доход CPM меняется минимально. Остаётся выбрать значение шага ΔCTR ; от него будет зависеть точность выхода на ограничение (2). Выберем шаг следующим образом:

$$\Delta CTR = \sum_{j=1, mCTR_{jv} < CTR_{min}}^M (mCTR_{jv} - CTR_{min})/M. \quad (6)$$

Алгоритм 1. Алгоритм оптимальной расстановки баннеров.

1. $\forall j, (1 \leq j \leq M)$:
 - (a) $CTR_j = mCTR_{jv}$
 - (b) $CPM_j = mCPM_{jv}$
2. $\forall j, (1 \leq j \leq M)$:
 - (a) $\Delta CPM_j = CPM(CTR_j) - CPM(CTR_j - \Delta CTR)$
 - (b) находим $j_{min} : \Delta CPM_{min} = \min_{CPM_j \in eCPM_{jv}} CPM_j$
 - (c) для запроса j_{min} :
 - (d) $CTR_{j_{min}} = CTR_{j_{min}} + \Delta CTR$
 - (e) $CPM_{j_{min}} = CPM_{j_{min}} - \Delta CPM_{j_{min}}$
3. для всего набора запросов:
 - (a) $CTR = \sum_{j=1}^M CTR_j/M$
 - (b) $CPM = \sum_{j=1}^M CPM_j$
4. если $CTR \geq CTR_{min}$, то:
 - (a) максимальное значение доходности при заданных ограничениях найдено и равно CPM
 - (b) запоминаем значение $\Delta CPM_{min}/\Delta CTR$
 - (c) выход из цикла
5. иначе повторяем цикл с п. 2

Как только достигнуто значение CTR_{min} , для каждого из запросов мы знаем соответствующие значения CPM_j и номера баннеров, которые будут показаны. Для того чтобы достигнуть ограничения (3), необходимо отсортировать все запросы по убыванию их CPM_j и оставить только Hit_{max} из них, запомнить пороговое значение $CPM_{Hit_{max}}$.

Сходимость Алгоритма 1 достигается за счёт того, что средняя кликабельность в формуле (2) монотонно убывает.

Теперь, когда мы “обучили” алгоритм, есть возможность использовать его для работы с новыми поступающими от пользователей запросами.

Алгоритм 2. Работа с новыми запросами.

1. Для нового запроса строится огибающая $eCPM_{jv}$.
2. На огибающей находится точка с “производной” $\Delta CPM_{min}/\Delta CTR$, по которой оцениваются CPM_j и CTR_j .
3. Оцененное CPM_j сравнивается с порогом $CPM_{Hit_{max}}$ и принимается решение о показе рекламных объявлений в спец-размещении по данному запросу.
4. Если по запросу решено показывать рекламу в спец-размещении, то по построенной огибающей для нового запроса и оцененному CTR_j определяются номера баннеров, которые будут показаны в спец-размещении.

Для одного запроса построение огибающей производится за $\mathcal{O}(B_j^3)$ операций, где B_j – количество баннеров-кандидатов для соответствующего запроса.

Предложенный алгоритм имеет следующие достоинства:

1. Для каждого запроса рассматриваются все варианты размещения рекламных объявлений в спец-размещении. Это позволяет:
 - учитывать взаимное влияние объявлений друг на друга (за счёт учёта этого при прогнозировании CTR) [13];
 - учитывать дополнительные характеристики объявлений при отборе, такие как релевантность [4]–[6], конверсионность [14] и т.п.;
 - считать off-line прогноз дохода исходя из аукциона второй цены [15] (т.к. нам известен порядок показа объявлений в рекламном блоке);
 - менять ранжирующую функцию внутри рекламного хита [16].
2. Имеется возможность учитывать дополнительные ограничения; от этого сложность алгоритма не изменится. Дополнительным ограничением может быть, например, средняя релевантность объявлений в спец-размещении.
3. Как побочный результат можно рассматривать некоторую классификацию запросов по различным характеристикам при максимизации дохода или любого другого выбранного критерия. Например, по виду огибающей можно разделять запросы на следующие:
 - монотонные виды огибающих: возрастающая (доход с запроса увеличивается с его средней кликабельностью), убывающая (доход убывает с увеличением средней кликабельности);
 - функция с одним максимумом: это значит, что для запроса есть один набор баннеров для показа, для которого доход максимальный. Это может говорить об отсутствии конкуренции, то есть что определённые рекламные объявления с большими ставками практически всегда показываются над результатами поиска;
 - у огибающей имеется несколько максимумов: запрос сильно конкурентный, причём в разных случаях по нему показываются те или другие наборы объявлений. Это может говорить о некоторой периодичности запроса или об его многозначности.

5. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНОГО ТЕСТИРОВАНИЯ АЛГОРИТМА

Данные для тестирования выбирались следующим образом:

- Взяты логи показов рекламы на поисковой странице компании “Яндекс” за промежуток времени 25–31 января 2013 года.
- Из данных по поисковому логу выделялись запросы, которые пользователи задавали за эту неделю.
- Из всего множества запросов порядка 10^9 было выбрано случайно 10^5 запросов (выборка запросов репрезентативная).

- Для каждого запроса составлен список соответствующих ему рекламных объявлений.
- Для каждого объявления известно:
 - CTR_{ij} — предсказание кликабельности для пары баннер-фраза;
 - Bid_i — ставка для данного баннера и соответствующей фразы.

Далее для каждого из запросов строились всевозможные варианты показов баннеров: из всех кандидатов составлялись различные возможные хиты. В одном хите могло быть три, два или один баннер. Также различными считаются варианты, когда состав рекламного хита одинаковый, однако объявления могут менять свои позиции. На *Рис. 3* представлен пример характеристик различных наборов объявлений для одного и того же запроса.

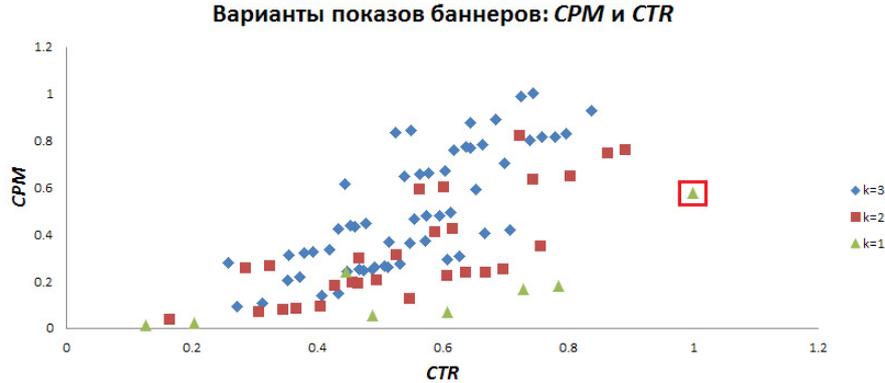


Рис. 3. Варианты показов баннеров в одном рекламном хите.

Из приведенного примера видно, что для максимизации суммарного дохода CPM невыгодно показывать одно объявление в спец-размещении (точка на графике, обрaмлённая квадратом), а для среднего CTR это вполне целесообразно. На *Рис. 4* построена огибающая с условием: выбрать $\max CPM_j(CTR_j)$ при $CTR_j = CTR_{min}$. По огибающей можно получить значение $\Delta CPM_j / \Delta CTR_j$.

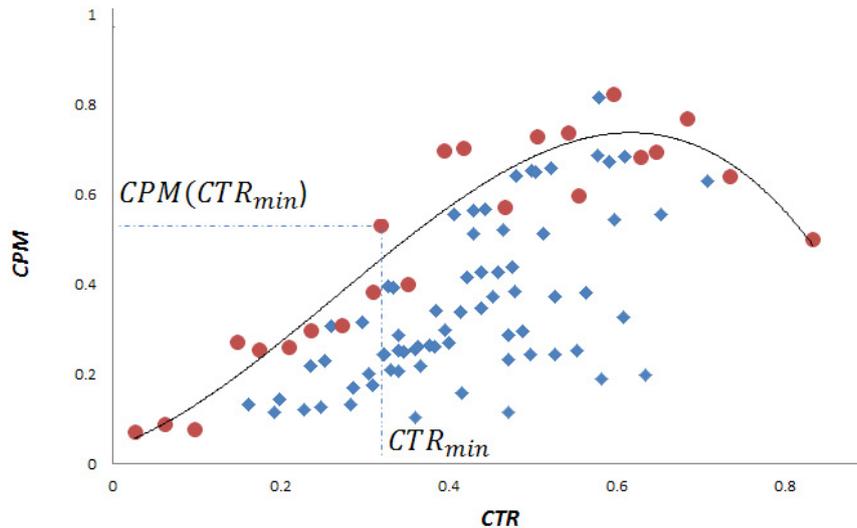


Рис. 4. Огибающая для максимальных значений дохода.

Такие огибающие строятся для каждого запроса, после этого по алгоритму подбирается оптимальное значение CPM_n и Hit_{max} запросов. На наборе из 10^5 запросов алгоритм показал увеличение дохода на 4%. Как показывают другие эксперименты по увеличению доходности поисковых систем [17], [15], [18], [19], увеличение доходности на 4% — это хороший результат, учитывая, что мы никак не влияем на механизм аукциона и на ставки непосредственно.

ЗАКЛЮЧЕНИЕ

В работе поставлена математическая задача оптимизации показов рекламных объявлений над результатами поиска: максимизация общего дохода от показа рекламы при ограничении на долю запросов, по которым реклама может быть показана, а так же на среднюю кликабельность объявлений в спец-размещении и максимальное количество рекламных объявлений на один показ (не более трёх). Полученный алгоритм отличается гибкостью использования ограничений по различным характеристикам: учёт дохода исходя с аукциона второй цены и любых других видов аукционов, возможность учёта взаимного влияния рекламных объявлений в одном рекламном хите, возможность использования различных ранжирующих функций. В дальнейшем планируется использовать метод для оптимизации конверсионности и релевантности рекламных объявлений, показываемых над результатами поиска.

СПИСОК ЛИТЕРАТУРЫ

1. К. Е. Бауман А.Н. Корнетова, В.А. Топинский, Д.А. Хакимова Оптимизация прогноза вероятности посещения контекстной рекламы в поисковой системе “Яндекс” *Научно-техническая информация. Сер. 2, Информационные процессы и системы.* – 2013. – № 4. – С. 1–8.
2. Bhumkar S., Mehta M., Shoemaker A. Methods and systems for enhancing internet experiences using previews : заяв. пат. 11/751,585 США. – 2007.
3. Agarwal D. Prediction of click through rates using hybrid kalman filter-tree structured markov model classifiers : пат. 7680746 США. – 2010.
4. Broder A. Z. et al. Search advertising using web relevance feedback *Proceedings of the 17th ACM conference on Information and knowledge management.* – ACM, 2008. – P. 1013–1022.
5. Ciaramita M., Murdock V., Plachouras V. Online learning from click data for sponsored search *Proceedings of the 17th international conference on World Wide Web.* – ACM, 2008. – P. 227–236.
6. Lacerda A. et al. Learning to advertise *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval.* – ACM, 2006. – P. 549–556.
7. Ghose A., Yang S. An empirical analysis of search engine advertising: Sponsored search in electronic markets *Management Science.* – 2009. – V. 55. – № 10. – P. 1605–1622.
8. Zhang W. V., Jones R. Comparing click logs and editorial labels for training query rewriting *WWW 2007 Workshop on Query Log Analysis: Social And Technological Challenges.* – 2007.
9. Schroedl S., Kesari A., Neumeyer L. Personalized ad placement in web search *Proceedings of the 4th Annual International Workshop on Data Mining and Audience Intelligence for Online Advertising (Ad-KDD)*, Washington USA. – 2010.
10. Lahaie S., Pennock D. M. Revenue analysis of a family of ranking rules for keyword auctions *Proceedings of the 8th ACM conference on Electronic commerce.* – ACM, 2007. – P. 50–56.
11. Корнетова А. Н., Червоненкис А. Я. Оптимизация показов рекламы в поисковых системах *Проблемы управления.* – 2013. – Т. 1. – №2. – С. 40–49.
12. Gupta S., Bilenko M., Richardson M. Catching the drift: learning broad matches from clickthrough data *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.* – ACM, 2009. – P. 1165–1174.

13. Reiley D. H., Li S. M., Lewis R. A. Northern exposure: A field experiment measuring externalities between search advertisements *Proceedings of the 11th ACM conference on Electronic commerce*. – ACM, 2010. – P. 297–304.
14. Sculley D. et al. Predicting bounce rates in sponsored search advertisements *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. – ACM, 2009. – P. 1325–1334.
15. Edelman B., Ostrovsky M., Schwarz M. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. – *National Bureau of Economic Research*, – 2005. – №w11765.
16. Feng J., Bhargava H. K., Pennock D. M. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms *INFORMS Journal on Computing*. – 2007. – V. 19. – №1. – P. 137–148.
17. Graham P. Search engine using sales and revenue to weight search results: пат. 6631372 США. – 2003.
18. Bhargava H. K., Feng J. Paid placement strategies for internet search engines *Proceedings of the 11th international conference on World Wide Web*. – ACM, 2002. – P. 117–123.
19. Zhu Y. et al. Optimizing search engine revenue in sponsored search *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. – ACM, 2009. – P. 588–595.

Algorithm of ads placement above the search results, maximizing search engine revenue

Sorokina A.

An important problem for the Internet search system is the selection of advertisements to show above the search results. This choice depends on the loyalty of advertisers that provide these ads to placed in, as well as users who react to relevant information provided to them. In the article formulated, mathematically described and solved the problem of the optimal choice of advertisements to display for maximizing search engine revenue with restrictions on the average clickthrough rate and part of queries with ads above the search results.

KEYWORDS: optimization, ad placement algorithm, sponsored search.