



# УСОВЕРШЕНСТВОВАННЫЙ АЛГОРИТМ ОПТИМИЗАЦИИ ПОКАЗОВ РЕКЛАМЫ В ПОИСКОВЫХ СИСТЕМАХ И РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

А.Н. Сорокина, А.Я. Червоненкис

Описана модификация предложенного ранее авторами алгоритма оптимизации показов рекламных объявлений на главной странице поисковой выдачи по запросам пользователей в системе поиска в Интернете. Показано, что модификация алгоритма позволяет существенно сократить перебор и объем производимых операций. Выполнено экспериментальное тестирование нового алгоритма на реальных наборах данных, предоставленных компанией «Яндекс».

**Ключевые слова:** коэффициент кликов, баннер, интернет-реклама, алгоритм, оптимизация.

## ВВЕДЕНИЕ

В работе [1] поставлена задача оптимизации показов рекламы в спецразмещении в поисковых системах, предложен метод и описан алгоритм ее решения. Этот алгоритм в точности решает соответствующую оптимизационную задачу, однако он достаточно сложен по объему необходимых операций и потому неудобен в практической реализации. В настоящей работе предлагается новый алгоритм, который, как будет теоретически показано, решает ту же самую оптимизационную задачу, что и прежний, но позволяет существенно сократить перебор и объем производимых операций.

Напомним постановку задачи. Поисковая выдача на запросы пользователя может сопровождаться рекламными объявлениями (баннерами). Если пользователь кликает на рекламное объявление, то деньги со счета рекламодателя списываются в пользу поисковой компании. Списываемая сумма определяется ставкой, которую задает рекламодатель. Эти деньги составляют существенную долю дохода поисковой компании. Наиболее привлекательно для рекламодателя размещение объявлений над результатами поиска — здесь более высокая вероятность клика на баннер [2]. Это место называется спецразмещением. Соответственно конкуренция за попадание в спецразмещение максимальна, и ставки, предлагаемые рекламодателями, наиболее высокие. В дальнейшем нас будет интересовать только спецразмещение.

В поисковой системе число объявлений, размещаемых над результатами поиска, ограничено (в настоящее время в системе компании «Яндекс» не более трех). Блок объявлений, показанных на поисковый запрос, называется хитом, но не все поисковые выдачи могут сопровождаться рекламными хитами: задана доля запросов, ответы на которые могут содержать рекламные объявления. Эти ограничения создают конкуренцию рекламодателей за право попадания в спецразмещение, которая разрешается с помощью аукциона второй цены [3, 4]. Задача оптимизации показов рекламы в спецразмещении состоит в отборе баннеров для показа в зависимости от ставок, прогноза вероятности клика на данное объявление и заданных ограничений. Выделим два критерия, позволяющие оценить качество выбора: суммарный доход, поступающий от рекламодателей, и средняя кликабельность — насколько часто пользователь кликает на то или иное из предъявленных объявлений (что определяет среднюю удовлетворенность пользователя). В настоящей работе предлагается максимизировать среднюю кликабельность при заданном ограничении (снизу) на доход (хотя совершенно аналогично решается и задача максимизации дохода при заданном ограничении снизу на среднюю кликабельность).

Другие критерии отбора объявлений представлены в работах [5–11], а другие модели — в работах [12, 13].

Напомним формальную постановку задачи и принятые обозначения: *Bid* — ставка, назначенная рекламодателем за клик на показанное объ-

явление;  $CTR$  — оценка вероятности клика (считается заданной);  $k$  — максимальное количество баннеров в спецразмещении на один запрос,  $CPM = CRT \cdot Bid$  — ожидаемая сумма денег, списываемая со счета рекламодателя за клик пользователя по конкретному баннеру.

Требуется найти такое правило отбора баннеров, показываемых в спецразмещении, которое обеспечивает максимум математического ожидания среднего значения оценки  $CTR$  при ограничениях:

— сохранение среднего дохода поисковой компании за определенный период времени;

— сохранение доли поисковых выдач, содержащих спецразмещение;

— для каждого спецразмещения число баннеров в нем не должно быть больше заданной величины  $k$ .

Оказалось, что оптимальная стратегия отбора баннеров (при заданных внешних параметрах) определяется некоторыми четырьмя параметрами  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  и  $T_0$  [1]. Для каждого из баннеров известны его показатели  $Bid$  и  $CTR$ . Предполагается, что для каждого запроса задан список кандидатов баннеров, которые могут быть по нему показаны. Тогда правило оптимального отбора баннеров будет следующим:

1) для каждого из этих баннеров вычислить величину

$$F_{ij} = CTR_{ij}/T_0 + \lambda_1 Bid_i CTR_{ij} - \lambda_2;$$

2) составить список кандидатов для показа в спецразмещении, из баннеров, для которых величина  $F$  положительна (этот список может быть и пустым);

3) если число кандидатов больше  $k$ , то сократить список, оставив в нем только  $k$  баннеров с наибольшим значением  $F$  (усеченный список);

4) подсчитать значение  $R_j(\lambda) = \sum_i F_{ij}$ , где сумма берется по всем баннерам из усеченного списка;

5) если это значение меньше  $\lambda_3$ , то обнулить список и для данного запроса не показывать никаких баннеров в спецразмещении. Иначе показать в спецразмещении все баннеры из усеченного списка.

Подбор параметров  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  и  $T_0$  предлагалось вести по представительному набору запросов, для каждого из которых известен список кандидатов баннеров и соответствующие значения  $Bid$  и  $CTR$ . Был предложен алгоритм, который для заданного пула решал задачу выбора оптимальных значений параметров  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  и  $T_0$  [1].

## 1. УСКОРЕННЫЙ АЛГОРИТМ ПОИСКА ПАРАМЕТРОВ $\lambda_1$ , $\lambda_2$ И $\lambda_3$

Полагаем, что в выборке имеется  $M$  запросов с индексами  $j$ ,  $1 \leq j \leq M$ , и задан список из  $N$  баннеров с индексами  $i$ ,  $1 \leq i \leq N$ . Обозначим:

$Bid_i$  — ставка, назначенная рекламодателем за  $i$ -й баннер, все  $Bid_i \geq 0$ ;

$CTR_{ij}$  — оценка вероятности клика при размещении  $i$ -го баннера над результатами поиска на  $j$ -й запрос. Значения  $Bid_i$  и  $CTR_{ij}$  считаем заданными.

Введем бинарную переменную  $t_{ij}$ , такую что  $t_{ij} = 1$  означает, что  $i$ -й баннер показан в спецразмещении на  $j$ -й запрос, и  $t_{ij} = 0$  в противном случае. Именно оптимальные значения этой переменной будем искать.

Далее заменим переменные  $t_{ij}$  на непрерывные, подчиненные дополнительному ограничению:  $0 \leq t_{ij} \leq 1$ . Корректность этой замены показана в работе [1]. Обозначим набор значений  $t_{ij}$  вектором  $\mathbf{t}$ , тогда суммарное число объявлений, показанных по всем запросам,  $TSUM = \sum_{i,j} t_{i,j}$ , а суммарный доход

$$MONEY(\mathbf{t}) = \sum_{i,j} Bid_i CTR_{ij} t_{ij}.$$

Суммарное количество запросов, для которых в спецразмещении показан хотя бы один баннер, обозначим  $Hit(\mathbf{t})$ . Это число не должно превосходить допустимый предел  $Hit_{max}$ .

В этих обозначениях целевая функция, подлежащая максимизации, имеет вид:

$$CRIT_0(\mathbf{t}) = \sum_{i,j} CTR_{ij} t_{ij} / TSUM. \quad (1)$$

Ограничения примут вид:

— по доходу:

$$MONEY(\mathbf{t}) \geq MON_{min}; \quad (2)$$

— по доле хитов:  $Hit(\mathbf{t}) \leq Hit_{max}$ ;

— по числу баннеров в спецразмещении на каждый запрос:

$$\forall j \sum_i t_{ij} \leq k;$$

— дополнительное ограничение  $0 \leq t_{ij} \leq 1$ .

Как уже отмечалось, алгоритм, предложенный в работе [1], в точности решал эту задачу. В настоящей работе предлагается его модификация (рис. 1), позволяющая существенно ускорить его работу.

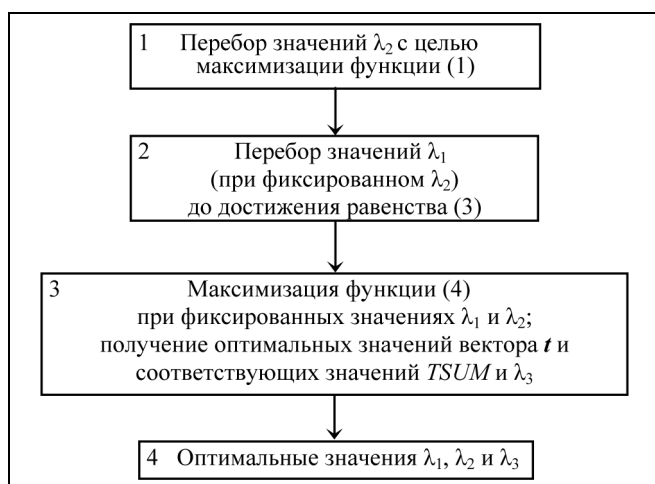


Рис. 1. Общая схема алгоритма подбора порогов входа в спецразмещение

Внешний цикл предлагаемой модификации алгоритма реализует перебор по переменной  $\lambda_2$ , в ходе которого ищется значение  $\lambda_{2\text{опт}}$ , доставляющее максимум (основной критерий) целевой функции (1). Вложенный цикл осуществляет перебор по переменной  $\lambda_1$ , в ходе которого ищется (при фиксированном значении  $\lambda_2$ ) значение  $\lambda_1$ , обеспечивающее равенство

$$MONEY(t) = MON_{\min}. \quad (3)$$

Замена неравенства (2) равенством (3) объясняется тем, что если бы было справедливо  $MONEY(t) > MON_{\min}$ , то по теории Куна — Таккера [14] должно быть справедливо  $\lambda_1 = 0$ , но тогда ограничение по деньгам вообще не учитывается, и максимум среднего значения  $CTR$  будет достигнут, если показать только баннеры с максимальным значением  $CTR$ . Но таковых будет один или совсем немного, а тогда (в практически интересных случаях) денег заведомо не хватит.

Можно показать, что оптимальное значение  $MONEY(t)$  (при фиксированном  $\lambda_2$ ) не убывает с ростом  $\lambda_1$ , и поэтому для нахождения корня уравнения (3) можно воспользоваться бинарным поиском.

Центральная часть алгоритма осуществляет выбор оптимальных значений переменных  $t_{ij}$ , доставляющих максимум функции

$$CRIT_1(t) = \sum_{i,j} t_{ij} \left( \frac{CTR_{ij}}{T_0} + \lambda_1 Bid_i CTR_{ij} - \lambda_2 \right). \quad (4)$$

при фиксированных значениях  $\lambda_1$  и  $\lambda_2$ . Значение переменной  $T_0$  можно выбрать любым, например равным 1 или текущему значению. Оптимизация идет тем же путем, что и описанном в работе [1].

## 2. ФОРМАЛЬНОЕ ОПИСАНИЕ АЛГОРИТМА

Запишем общий алгоритм подбора порогов для показа объявлений в спецразмещении.

**Обозначения:**  $M$  — общее число запросов,  $N$  — общее количество баннеров-кандидатов в выборке запросов,  $Hit_{\max}$  — максимальное допустимое покрытие,  $MON_{\min}$  — минимальные деньги, которые должны быть выручены от показа баннеров в спецразмещении по выборке запросов,  $T_0$  — количество показов баннеров при покрытии  $Hit_{\max}$  для текущего состояния системы. Для каждого запроса и баннера-кандидата дано:

$CTR_{ij}$  — прогноз  $CTR$  (кликабельности) объявления,

$Bid_i$  — ставка рекламодателя по  $i$ -му баннеру.

Алгоритм можно записать в следующем виде.

**Цикл по  $\lambda_2$ .**

Значение  $\lambda_2$  доставляет максимум (основной критерий) целевой функции (1).

**Цикл по  $\lambda_1$ .**

Вырученные деньги с ростом  $\lambda_1$  не убывают, т. е. можно считать что  $\lambda_1$  — параметр, регулирующий поступление денег от рекламодателей. Перебор по  $\lambda_1$  идет до достижения равенства (3).

**Цикл по запросам  $j$ .**

По всем баннерам-кандидатам  $i$  для запроса  $j$ :

— составить «список для показа», вычисляя для каждого баннера величину  $F_{ij}(\lambda, T_0) = CTR_{ij}/T_0 + \lambda_1 CPM_{ij} - \lambda_2$ , не включая в список те  $i$ , для которых  $F_{ij} \leq 0$ , и оставив  $k$  баннеров с наибольшим значением  $F_{ij}$ ;

— вычислить вклад в сумму критериев по всем баннерам, вошедшим в список для показа в спецразмещении:

$R_j = \sum F_{ij}$  (по баннерам, вошедшим в список).

Конец цикла по запросам  $j$ .

Упорядочить списки для запросов ( $j$ ) в порядке убывания  $R_j$ .

Оставить только первые  $Hit_{\max}$  из них (выполнение ограничения по покрытию), остальные обнулить (это в том случае, если хитов хватает, в противном случае оставить все хиты).

Запомнить  $\lambda_3 = \min_j R(j)$ .

Положить

$$t_{ij} = \begin{cases} 1, & \text{если пара } (i, j) \text{ в списке} \\ & \text{и оставлена для показа,} \\ 0 & \text{в противном случае.} \end{cases}$$

Изменять  $\lambda_1$ , пока не будет выполнено неравенство (3), следующим образом:

- если  $MONEY(\mathbf{t}) < MON_{\min}$ , то уменьшить  $\lambda_1$ ;
- если  $MONEY(\mathbf{t}) > MON_{\min}$ , то увеличить  $\lambda_1$ .

Положить  $\lambda_{1\text{опт}} = \lambda_1$ .

Запомнить  $\lambda_3, \lambda_{1\text{опт}}$ , списки  $t_{ij}$  (т. е. баннеры, отобранные для показа).

Конец цикла по  $\lambda_1$ .

Вычислить:

$$\begin{aligned} CRIT_1(\lambda_{1\text{опт}}, \lambda_2, T_0) &= \\ &= \sum_{i,j} t_{ij} \left( \frac{CTR_{ij}}{T_0} + \lambda_{1\text{опт}} Bid_i CTR_{ij} - \lambda_2 \right). \end{aligned}$$

Изменять  $\lambda_2$ , чтобы достигнуть максимума  $CRIT_0(\mathbf{t}) = \sum CTR_{ij} t_{ij} / TSUM$ .

Положить  $\lambda_{2\text{опт}}$  то значение  $\lambda_2$ , при котором достигнут  $\max CRIT_0(\mathbf{t})$ .

Конец цикла по  $\lambda_2$ .

В конце алгоритма получаем три значения  $\lambda_{1\text{опт}}, \lambda_{2\text{опт}}$  и  $\lambda_{3\text{опт}}$ , которые будут использоваться для работы с новыми запросами.

### 3. ДОКАЗАТЕЛЬСТВО ЭКВИВАЛЕНТНОСТИ ДВУХ АЛГОРИТМОВ

Согласно «старому» алгоритму, предложенному в работе [1], оптимизация идет в следующем порядке.

Цикл по  $\lambda_1$ .

Цикл по  $\lambda_2$ .

Цикл по  $T_0$ .

При заданных  $\lambda_1, \lambda_2$  и  $T_0$  выбор оптимального значения вектора  $\mathbf{t}$  идет по критерию максимума функции

$$CRIT_{\text{стар}}(\mathbf{t}) = \sum_{i,j} t_{ij} \left( \frac{CTR_{ij}}{T_0} + \lambda_1 Bid_i CTR_{ij} - \lambda_2 \right)$$

при дополнительных ограничениях.

Перебор по  $T_0$  до выполнения условия  $T_0 = TSUM$ .

Перебор по  $\lambda_2$  до выполнения условия  $MONEY = MON_{\min}$ .

Перебор по  $\lambda_1$  до достижения максимума функции  $CRIT_0(\mathbf{t})$ .

«Новый» (ускоренный) алгоритм отличается тем, что в нем отсутствует цикл по  $T_0$ , а значение  $T_0$  можно положить равным произвольной положительной константе. Для простоты положим  $T_0 = 1$ , случай произвольного  $T_0$  рассматривается аналогично. Соответственно изменятся и «внутренний» критерий, по которому находится вектор при заданных значениях  $\lambda_1$  и  $\lambda_2$ . Алгоритм состоит в следующем.

Цикл по  $\lambda_2$ .

Цикл по  $\lambda_1$ .

При заданных значениях  $\lambda_1$  и  $\lambda_2$ , выбор оптимального значения вектора  $\mathbf{t}$  идет по критерию максимума функции  $CRIT_{\text{нов}}(\mathbf{t}) = \sum_{i,j} t_{ij} (CTR_{ij} + \lambda_1 Bid_i CTR_{ij} - \lambda_2)$  при дополнительных ограничениях.

Перебор по  $\lambda_1$  до выполнения условия  $MONEY = MON_{\min}$ .

Перебор по  $\lambda_2$  до достижения максимума критерия  $CRIT_0(\mathbf{t}) = \sum_{i,j} CTR_{ij} t_{ij} / TSUM$ .

Покажем, что результат работы этих двух алгоритмов (выбор значения  $\mathbf{t}_{\text{опт}}$ ) совпадает. Допустим, что  $\mathbf{t}_{\text{опт}}$  — оптимальное значение вектора  $\mathbf{t}$ , выбранное «старым» алгоритмом. Пусть координаты этого вектора есть  $t_{ij\text{опт}}$ , а величина  $TSUM_{\text{опт}}$  равна  $\sum_{i,j} t_{ij\text{опт}}$ , подобранные при этом значения  $\lambda_1, \lambda_2$  и  $T_0$  равны  $\lambda_{1\text{опт}}, \lambda_{2\text{опт}}$  и  $T_{0\text{опт}}$ . Покажем, что значение  $\mathbf{t}_{\text{опт}}$  может служить и результатом «нового» алгоритма.

Действительно, если умножить  $CRIT_{\text{стар}}(\mathbf{t})$  на константу  $TSUM_{\text{опт}}$ , то получим:

$$\begin{aligned} CRIT_{\text{стар}}^*(\mathbf{t}) &= \sum_{i,j} t_{ij} (CTR_{ij} (TSUM_{\text{опт}} / T_0) + \\ &+ (\lambda_1 TSUM_{\text{опт}}) Bid_i CTR_{ij} - \lambda_2 TSUM_{\text{опт}}). \end{aligned} \quad (5)$$

При умножении функции (5) на константу значение вектора  $\mathbf{t}$ , доставляющее ей максимум, не меняется. В частности,  $\mathbf{t}_{\text{опт}}$  будет доставлять максимум и критерию

$CRIT_{\text{стар}}^*(\mathbf{t})$  при  $\lambda_1 = \lambda_{1\text{опт}}, \lambda_2 = \lambda_{2\text{опт}}$  и  $T_0 = T_{0\text{опт}}$ . Обозначим  $\lambda_1^* = \lambda_{1\text{опт}} TSUM_{\text{опт}}$  и  $\lambda_2^* = \lambda_{2\text{опт}} TSUM_{\text{опт}}$  и заметим, что по условию  $T_{0\text{опт}} = TSUM_{\text{опт}}$  и, значит,  $TSUM_{\text{опт}} / T_0 = 1$ .

Тогда функция  $CRIT_{\text{стар}}^*(\mathbf{t})$  при оптимальных значениях  $\lambda_1, \lambda_2$  и  $T_0$  примет вид:

$$CRIT_{\text{стар}}^*(\mathbf{t}) = \sum_{i,j} t_{ij} (CTR_{ij} + \lambda_1^* Bid_i CTR_{ij} - \lambda_2^*).$$

Но эта функция совпадает с функцией, используемой в «новом» алгоритме. Действительно, функция

$$CRIT_{\text{нов}}(\mathbf{t}) = \sum_{i,j} t_{ij} (CTR_{ij} + \lambda_1 Bid_i CTR_{ij} - \lambda_2)$$

совпадает с функцией

$$CRIT_{\text{стар}}^*(\mathbf{t}) = \sum_{i,j} t_{ij} (CTR_{ij} + \lambda_1^* Bid_i CTR_{ij} - \lambda_2^*).$$



Поэтому и значение  $t_{\text{опт}}$  доставит максимум новой функции  $CRIT_{\text{нов}}(t)$  при  $\lambda_1 = \lambda_1^*$  и  $\lambda_2 = \lambda_2^*$ . При этом значения  $\lambda_1^*$  и  $\lambda_2^*$  таковы, что выполняются условия подбора, так как при  $t = t_{\text{опт}}$   $MONEY = MON_{\text{min}}$  и достигается максимум функции  $CRIT_0(t) = \sum_{i,j} CTR_{ij}t_{ij}/TSUM$ .

Таким образом, выбор оптимального значения вектора  $t$  будет одинаковым для обоих алгоритмов, но значения параметров  $\lambda_1$  и  $\lambda_2$  будут разными, и это нужно учесть при работе с новыми парами «запрос — баннер».

#### 4. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНОГО ТЕСТИРОВАНИЯ НОВОГО АЛГОРИТМА

Последовательно рассмотрим каждый из этапов выбора значений параметров  $\lambda_1$ ,  $\lambda_2$  и  $\lambda_3$  и убедимся, что алгоритм отбирает объявления нужным нам способом. Тестирование производилось на 100 тыс. поисковых запросах, реально встретившихся при работе Яндекса за некоторую неделю во времени, и баннеров, предложенных рекламодателями.

Для каждого из этих запросов были заранее отобраны баннеры-кандидаты на показ в спецразмещении — примерно по 50 баннеров на запрос. Также, для каждой пары «запрос — баннер-кандидат» были заранее подсчитаны оценки вероятности клика  $CTR_{ij}$ , для каждого баннера была известна его ставка  $Bid_i$ . Значение параметра  $T_0$  можно выбрать любым, и мы положим его равным текущему значению суммарного числа показанных баннеров. Просмотр будем вести, начиная с внутренних операций при фиксированных значениях  $\lambda_1$  и  $\lambda_2$ . Затем посмотрим результат работы перебора по параметру  $\lambda_1$  при фиксированных значениях  $\lambda_2$ . Наконец, рассмотрим работу внешнего цикла по  $\lambda_2$ , в результате которого будет выбрано оптимальное значение параметра  $\lambda_2$ , а вместе с тем и требуемые значения параметров  $\lambda_1$  и  $\lambda_3$ .

##### 4.1. Значения параметров $\lambda_2$ и $\lambda_1$ фиксированы

Рассмотрим ту часть алгоритма, которая на схеме (см. рис. 1) представлена блоками 2 и 3). Поскольку здесь все параметры известны, то для каждой пары (запрос — баннер-кандидат) можно вычислить величину

$$F_{ij}(\lambda, T_0) = CTR_{ij}/T_0 + \lambda_1 CRM_{ij} - \lambda_2.$$

В предварительный список баннеров для показа на  $i$ -й запрос включаются только те баннеры, для которых величина  $F_{ij}(\lambda, T_0)$  положительна. Список мог оказаться и пустым. Из этого списка отбирались  $k = 3$  лучших баннеров по критерию

максимума  $F_{ij}(\lambda, T_0)$ . Если же число баннеров в списке оказывалось меньше трех, то отбирались все баннеры списка. Таким путем формируются усеченные списки (хиты) для всех запросов нашего пула.

Теперь для каждого усеченного списка считается суммарный критерий:  $R_i = \sum F_{ij}$ , где сумма берется по всем баннерам списка для  $i$ -го запроса. Далее просматриваются все запросы пула с непустым списком и отбираются из них первые  $Hit_{\text{max}}$  в порядке убывания суммарного критерия  $R_i$ . Остальные запросы будут оставлены без рекламных объявлений. Тем самым будет выполнено ограничение по покрытию. Минимальное значение  $R_i$ , относящееся к отобраным запросам, принимается за  $\lambda_3$ . Оно, конечно, зависит от заданных значений  $\lambda_1$  и  $\lambda_2$ . Может случиться так, что все запросы с непустым списком будут отобраны, а значение  $Hit_{\text{max}}$  все равно не будет достигнуто. Это значит, что ограничение на покрытие автоматически выполняется благодаря требованию  $F_{ij}(\lambda, T_0) > 0$ . В этом случае полагаем  $\lambda_3 = 0$ . Теперь подсчитывается число  $TSUM$  как суммарное число пар «запрос — баннер», оставленных для показа. Оно нам понадобится при вычислении основного критерия. Одновременно вычисляется критерий  $CRIT_1$  как сумма значений  $F_{ij}(\lambda, T_0)$  по всем отобраным парам «запрос — баннер».

##### 4.2. Значение параметра $\lambda_2$ фиксировано, варьируется параметр $\lambda_1$

Для нахождения требуемого значения параметра  $\lambda_1$  рассматривается зависимость величины  $\sum CPM$

$$\sum CPM = \sum Bid_i CTR_{ij}$$

от параметра  $\lambda_1$  при фиксированном значении параметра  $\lambda_2$ . Сумма берется по всем отобраным парам запрос-баннер, процедура отбора которых описана в предыдущем разделе. Величину  $\sum CPM$  в рассматриваемом случае принимаем как аппроксимацию суммарного дохода  $MONEY(t)$ .

На рис. 2 показаны кривые изменения величины  $\sum CPM$  в зависимости от изменения параметра  $\lambda_1$  для различных значений параметра  $\lambda_2$ . Одно из этих значений ( $\lambda_1 = 7,46$ ) соответствует точке оптимума основного критерия, но оно может быть найдено только после проведения оптимизации по этому параметру.

Согласно предлагаемому алгоритму, оптимальное значение параметра  $\lambda_1$  достигается при выпол-

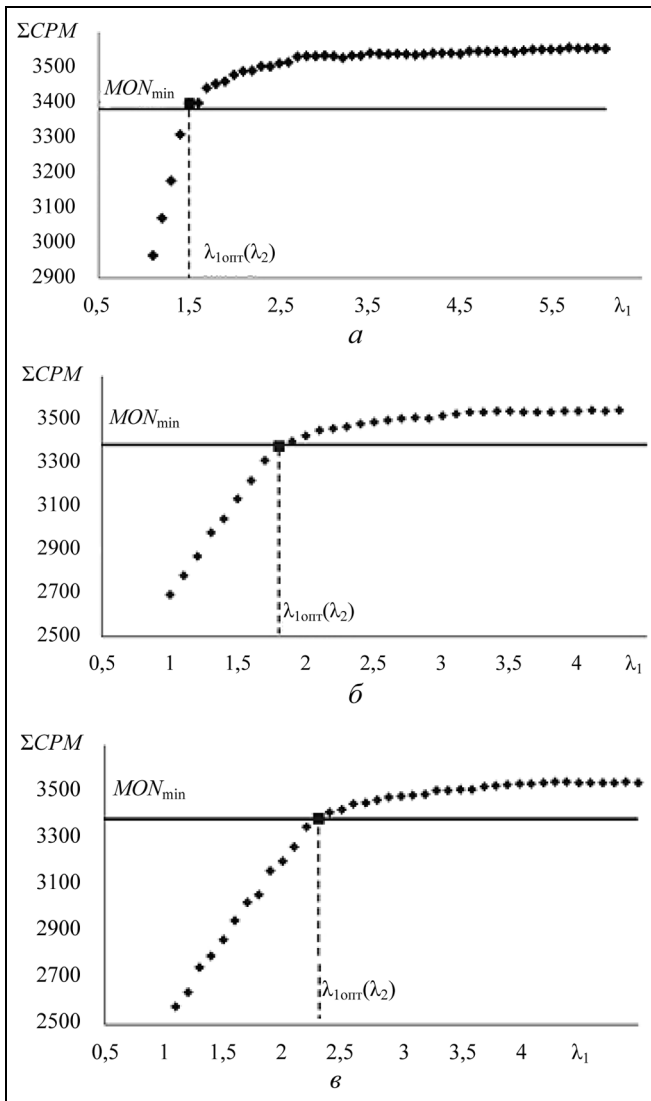


Рис. 2. Динамика изменения величины  $\Sigma CPM$  в зависимости от варьирования параметра  $\lambda_1$  при фиксированном значении параметра  $\lambda_2$ : а —  $\lambda_2 = 7,46$ ; б —  $\lambda_2 = 8,58$ ; в —  $\lambda_2 = 10,26$

нении равенства (3) или в нашем приближении  $\Sigma CPM = MON_{min}$ . Видно, что величина  $\Sigma CPM$  монотонно растет с увеличением значения параметра  $\lambda_1$ . Поэтому для нахождения корня равенства можно использовать тот или иной ускоренный алгоритм.

Интересно отметить, что суммарный доход  $CPM$  по системе растет сильнее при варьировании  $\lambda_1$  от 1,5 до 2,7, дальше же, видимо, происходит некоторое «насыщение», т. е. объявлений с большими значениями  $CPM$  уже не остается среди кандидатов на показ, и приходится добирать все более дешевые объявления, и суммарный  $CPM$  начинает расти медленнее.

### 4.3. Варьирование параметра $\lambda_2$

Теперь после того, как для каждого значения  $\lambda_2$  найдено требуемое значение  $\lambda_1$ , рассмотрим, как ищется оптимальное значение параметра  $\lambda_2$ . Для каждого значения  $\lambda_2$  одновременно с нахождением требуемого значения  $\lambda_{1opt}(\lambda_2)$ , т. е. для пары  $(\lambda_2, \lambda_{1opt}(\lambda_2))$  мы нашли оптимальное значение переменных  $t_{ij}$ , удовлетворяющее ограничениям, а по ним среднее значение  $CTR$ , суммарный доход  $CPM$ , величину  $TSUM$  для нашего пула данных и требуемое значение параметра  $\lambda_3$ , контролирующего покрытие. По этим данным алгоритм вычисляет значение основного критерия

$$CRIT_0(t) = \sum_{i,j} CTR_{ij} t_{ij} / TSUM,$$

график зависимости которого от  $\lambda_2$  показан на рис. 3.

Каждая точка на этом графике представляет собой результат работы внутренних циклов, представленных на рис. 1 (блоки 2 и 3). Видно, что зависимость основного критерия от параметра  $\lambda_2$  имеет достаточно отчетливый максимум. Точка, в которой достигается этот максимум, принимается за оптимальное значение  $\lambda_1 = \lambda_{2opt}$  для которого мы получаем также соответствующие значения  $\lambda_{1opt}(\lambda_2)$  и  $\lambda_{3opt}(\lambda_2)$ . Эти значения составляют выход программы оптимизации и будут использованы для дальнейшей работы с новыми запросами и баннерами-кандидатами на показ в спец-размещении.

Полный цикл подбора порогов для пула запросов занимает от 6 до 9 ч в зависимости от шага перебора по параметрам  $\lambda_1$  и  $\lambda_2$ . Экспериментальное тестирование предложенного алгоритма показало, что процедура работает успешно.

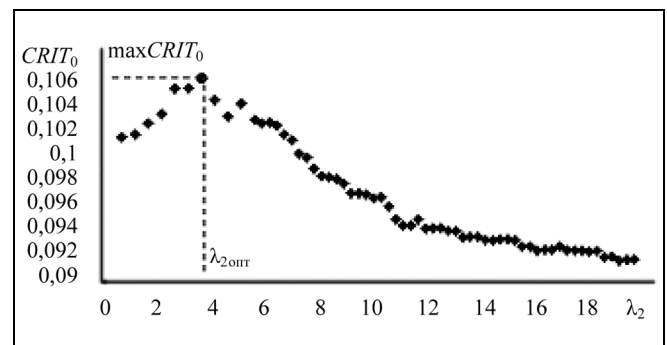


Рис. 3. Зависимости величины  $CRIT_0$  от параметра  $\lambda_2$



## 5. РАБОТА АЛГОРИТМА НА РЕАЛЬНЫХ ДАННЫХ

Схема показов баннеров в интернет-системе «Яндекс» очень сложна, поэтому подбор параметров on-line на данном этапе мало возможен. В настоящее время алгоритм подбирает значения  $\lambda_{1\text{опт}}$ ,  $\lambda_{2\text{опт}}$  и  $\lambda_{3\text{опт}}$  на пуле запросов, состоящем из 100 тыс. случайных запросов из поисковых логов, собранных за неделю данных. В среднем на каждый запрос предварительно отбирается 50 баннеров-кандидатов, и по этим данным получают оптимальные значения параметров  $\lambda_{1\text{опт}}$ ,  $\lambda_{2\text{опт}}$  и  $\lambda_{3\text{опт}}$ , процедура выбора которых описана в § 4.

После того, как эти параметры найдены, работа с новыми (текущими) запросами идет on-line и выбор баннеров, предъявляемых пользователю, осуществляется очень быстро. На каждый запрос предварительно составляется список кандидатов-баннеров. (Этот отбор идет по простым признакам, его описание выходит за рамки данной работы.)

К этому моменту оценка кликабельности объявления для данного запроса  $CTR_i$  и его ставка  $Bid_i$  считаются уже известными, и нужно произвести всего лишь ряд элементарных математических операций, чтобы вычислить величину

$$F_i = CTR_i + \lambda_{1\text{опт}} Bid_i CTR_i - \lambda_{2\text{опт}}$$

В предварительный список для показа включаются все баннеры, для которых  $F_i > 0$ . Из него отбираются  $k = 3$  баннеров с наибольшим значением  $F_i$  (если их хватит, иначе все). Они составляют полный хит для показа. Теперь считается суммарный критерий  $R = \sum F_i$  (сумма берется по всем баннерам хита), который сравнивается с параметром  $\lambda_{3\text{опт}}$  если  $R < \lambda_{3\text{опт}}$ , то рекламные объявления на данный запрос в спец-размещении не показываются, иначе — показывается весь хит.

Для экспериментального тестирования использованы данные за январь 2013 года, которые предоставила компания «Яндекс».

В ходе работы алгоритма полученный выигрыш относительно базового уровня составил 8 %. Предложенный способ отбора баннеров в спец-размещение применяется в настоящее время в системе показов рекламных объявлений компании «Яндекс».

## ЗАКЛЮЧЕНИЕ

Предложена модификация алгоритма оптимизации показов рекламных объявлений, представленного в работе [1]. Доказана эквивалентность первоначального и улучшенного алгоритмов. Но-

вый алгоритм существенно сокращает перебор и объем производимых операций. Экспериментальное тестирование нового алгоритма на реальных наборах данных, предоставленных компанией «Яндекс», показало, что его применение дает выигрыш в 8 % максимизируемой средней оценки вероятности клика.

## ЛИТЕРАТУРА

1. Корнетова А.Н., Червоненкис А.Я. Оптимизация показов рекламы в поисковых системах // Проблемы управления. — 2013. — № 1. — С. 40—49.
2. Kumar R., Mahdian M., Li E., et al. Systems and methods for exploring new sponsored search listings of uncertain quality. — 2009.
3. Edelman B., Ostrovsky M. and Schwarz M. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords // American Economic Review. — 2007. — Vol. 97, N 1. — P. 242—259.
4. Fain D.C., Pedersen J.O. Sponsored search: A brief history // Bulletin of the American Society for Information Science and Technology. — 2006. — Vol. 32, N 2. — P. 12—13.
5. Graepel T., Candela J.Q., Borchert T. and Herbrich R. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine // ICML 2010 — Proc. of 27th Intern. Conf. on Machine Learning. — P. 13—20.
6. 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. — Vol. 19, N 1. — P. 137—148.
7. Ghose A., Yang S. An empirical analysis of search engine advertising: Sponsored search in electronic markets // Management Science. — 2009. — Vol. 55, N 10. — P. 1605—1622.
8. Graepel T., et al. Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine // ICML 2010 — Proc. of 27th Intern. Conf. on Machine Learning. — P. 13—20.
9. Lahaie S., Pennock D.M. Revenue analysis of a family of ranking rules for keyword auctions // Proc. of the 8th ACM Conf. on Electronic Commerce. — ACM 2007. — P. 50—56.
10. Radlinski F., et al. Optimizing relevance and revenue in ad search: a query substitution approach // Proc. of the 31st Annual Intern. ACM SIGIR Conf. on Research and Development in Information Retrieval. — ACM 2008. — P. 403—410.
11. Feuerstein E., et al. New Stochastic Algorithms for Scheduling Ads in Sponsored Search // Web Conference, 2007. LA-WEB 2007. Latin American. — IEEE 2007. — P. 22—31.
12. Richardson M., Dominowska E., Ragno R. Predicting clicks: estimating the click-through rate for new ads // Proc. of the 16th Intern. Conf. on World Wide Web. — ACM 2007. — P. 521—530.
13. Zhu Z. A., et al. A novel click model and its applications to online advertising // Proc. of the Third ACM Intern. Conf. on Web Search and Data Mining. — ACM 2010. — P. 321—330.
14. Кун Г.У., Таккер А.У. Линейные неравенства и смежные вопросы. — М.: Изд-во иностр. литературы, 1959.

Статья представлена к публикации членом редколлегии Ф.Т. Алескеровым.

**Сорокина Анна Николаевна** — ст. исследователь-разработчик, компания «Яндекс»; аспирант, Национальный исследовательский университет — Высшая школа экономики, г. Москва, ✉ anna.n.sorokina@gmail.com,

**Червоненкис Алексей Яковлевич** — канд физ.-мат. наук, вед. науч. сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ☎ (495) 334-88-20 ✉ chervnks@ipu.ru.