

# ОПТИМИЗАЦИЯ ПОКАЗОВ РЕКЛАМЫ В ПОИСКОВЫХ СИСТЕМАХ

А.Н. Корнетова, А.Я. Червоненкис

Предложен алгоритм отбора объявлений для каждого поискового запроса для показа рекламы по нему. Сформулирована и решена соответствующая оптимизационная задача с учетом ограничений уже существующей системы, таких как сумма денег, списанная со счетов рекламодателей, доля запросов, по которым реклама может быть показана. Критерием оптимальности принята эффективность показов рекламных объявлений, которая определяется удовлетворенностью пользователей и, соответственно, увеличением их внимания к рекламной выдаче на главной странице поиска.

**Ключевые слова:** коэффициент кликов баннера, реклама на поиске, интернет-реклама, оптимизация, построение алгоритма.

## ВВЕДЕНИЕ. СУЩЕСТВУЮЩАЯ СИСТЕМА ПОКАЗОВ РЕКЛАМЫ НА ПОИСКЕ

**Правила показа рекламного объявления.** Огромное число людей пользуются Интернетом для поиска информации. Обычно Интернет-пользователь делает некоторый запрос, на который он хочет получить ответ, и задача поисковой системы — ответить на запрос наилучшим образом. Чтобы иметь некоторую прибыль от поиска информации, поисковик размещает *рекламные объявления (баннеры)* на странице поисковой выдачи. Поскольку содержание объявлений зависит от запроса пользователя, такая реклама называется контекстной. По запросу составляется список *ключевых фраз*, к которым и привязываются рекламные объявления. Вообще говоря, не все баннеры, отобранные по запросу, могут быть показаны (об этом речь пойдет далее), т. е. отображены на странице поисковой выдачи. Баннеры содержат рекламную информацию о товаре или услуге рекламодателя, а также ссылку на соответствующий сайт. Если пользователь кликает на то или иное объявление, происходит переход на рекламируемый сайт, этот переход называется *кликом*. В случае клика со счета рекламодателя списываются деньги. Общая сумма списываемых со счетов рекламодателей денег оказывается существенной составляющей дохода поисковой компании. Рекламодатель для каждого объявления выставляет ставку, и происходит торг за показ его рекламного объявления на поиске. Другие схемы списаний денег со счетов рекламо-

дателей (например, за показы его объявления или за конверсию — покупку товара на сайте рекламодателя) можно увидеть в работе [1].

Рекламные объявления могут показываться в разных местах поисковой страницы (рис. 1): справа от результатов поиска (*гарантированные показы*), вверху страницы над результатами поиска (*спецразмещение*), внизу страницы и др.

Спецразмещение — наиболее привилегированное место для показа рекламы, оно приоритетно для рекламодателей [2] и самое прибыльное для поисковой системы. Зачастую, для того чтобы попасть в спецразмещение, из-за конкуренции рекламодателю приходится делать достаточно высокую ставку.

Настоящая работа посвящена разработке метода отбора рекламных объявлений для показа в блоке спецразмещения.

Выгодность показа рекламного объявления в спецразмещении для поисковой системы определяется двумя критериями:

— вероятность того, что пользователь кликнет на предъявленное объявление (*кликабельность CTR, Click Through Rate*); можно сказать, что это — степень эффективности показа объявления, его характеристика привлекательности для пользователя;

— *ставка Bid*, назначенная рекламодателем за показ его объявления в спецразмещении.

Детальное описание терминологии поисковой контент-рекламы можно найти в работе [3].

Для вероятности *CTR*, конечно, можно дать только оценку, поскольку как пользователи, так и объявления постоянно меняются. На оценку веро-



ятности клика влияет большое число факторов, которые учитываются в формуле оценки, порождаемой специальными программами машинного обучения. Однако в данной работе мы не будем касаться этого вопроса и примем, что оценка критерия *CTR* для каждой пары «запрос — рекламное объявление» нам дана.

Ожидаемая сумма денег, списываемых со счета рекламодателя за клик пользователя по конкретному баннеру, может быть оценена как *CPM* (Cost Per Million) =  $CTR \cdot Bid$ . Именно эта величина служит критерием, по которому отбираются кандидаты для показа в спецразмещении. В действительности деньги списываются несколько сложнее — согласно аукциону второй цены [4].

Существуют и другие критерии отбора объявлений:

- сортировка по критерию  $Bid \cdot quality + v$ , где *quality* — некоторое качество объявления (в нашем случае это предсказываемая оценка *CTR*), а *v* — количественная мера полезности объявления для поисковой системы в качестве обучающего материала и дополнительной информации о качестве (вообще говоря, это некоторое начальное предсказание значений *CTR* для новых баннеров); для новых баннеров значение *v* большое, для старых практически нулевое (так как там уже работает статистическая оценка вероятности *CTR* по истории показов объявления);

**Запрос пользователя**

Поиск Почта Карты Маркет Новости Словари Блоги Видео Картинки ещё

**Яндекс** Нашлось 165 млн ответов

toyota  в найденном  в Москве  расширенный поиск

Мои находки Настройка Регион: Москва Войти Помощь

**Спецразмещение**

**Все объявления**

**Такая Toyota нужна самому!**  
Спецпредложение от официального дилера! ТЦ Отрадное, ТЦ Коломенское. Спешите toyotanm.ru

**Пройдите Тест-драйв TOYOTA!**  
Выгода на новые TOYOTA до 500 т.р. в Тойота Центр Ясенево! (495) 7777-15 toyota-yasenevo.ru

**Автомобили Toyota в Major Auto**  
Уникальное предложение при покупке автомобиля в салонах официальных дилеров major-toyota.ru

Результаты **все** в рунете в мировом интернете

**"Тойота мотор" - автоконцерн**

RAV4 Toyota Trade-in  
Corolla LC Prado  
Camry Highlander

Продажа и техническое обслуживание автомобилей Toyota. Каталог автомобилей с описаниями и техническими данными. Новости компании. Контакты.  
Москва, МКАД автомагистраль, 84-й км, стр. 1, вл. 5 +7 (495) 258-34-65 toyota.ru Москва

**Toyota — модельный ряд, комплектации, отзывы**  
Модели и цены 2012 Corolla Camry Land Cruiser  
6 298 объявлений о продаже Toyota в Москве  
Фотографии, комплектации и цены нового модельного ряда. Официальные дилеры. auto.yandex.ru > Toyota Москва

**Завод Toyota в Петербурге приступил к работе в две смены**  
Завод Toyota в Петербурге планирует удвоить производство по итогам 2012 года В данный момент автозавод выпускает одну модель – Toyota Camry в 12 модификациях и 6 цветах.  
REGNUM 14:09 ИА РБК Санкт-Петербург 13:28 Автостат 13:04 Все сообщения 16 news.yandex.ru 53 минуты назад

**TOYOTA | ...Центр Сокольники Новорижский Шереметьево - дилер Toyota...**  
Автомобили Комплектация Hilux Технические характеристики  
марка: Toyota; официальный дилер; тип авто: новые, с пробегом; тест-драйв; wi-fi  
Автомобили Toyota – Продажа Сервис Запчасти Toyota. ... 18.07.2012. Компания Тойота

**Яндекс Директ**

**Срочный выкуп авто! Звоните!**  
Хотите продать авто - быстро, с комфортом и за достойные деньги? Звоните! gemo-avto.ru

**Выкуп автомобилей Дорого! Срочно!**  
Выезд оценщика, снятие с учета, эвакуация бесплатно! 15 лет на рынке! mustangavto.com

**Успейте купить Toyota!**  
До 31 августа все автомобили Toyota по старым ценам в Тойота Центр Внуково. toyota-vnuково.ru

**Toyota от официального дилера!**  
Отличные предложения на все модели Toyota у дилера Тойота Центр Кунцево!  
Адрес и телефон toyota-kuntsevo.ru

Разместить объявление по запросу «toyota» — 898 887 показов в месяц

**Результаты поиска**

**Показы справа**

**ММАС**  
МОСКОВСКИЙ МЕЖДУНАРОДНЫЙ АВТОМОБИЛЬНЫЙ САЛОН

**НА СТЕНДЕ JAGUAR LAND ROVER ПАВИЛЬОН 2, ЗАЛ 7**

УЗНАЙТЕ БОЛЬШЕ



Рис. 1. Места для показов рекламы на поисковой выдаче

- сортировка по  $Bid \cdot quality + v$ , где  $v = cBid \cdot var(quality)$ ,  $c$  — константа, показывающая, насколько полезно обучение на данном объявлении;
- учет релевантности объявления запросу (с помощью ссылок на сайт рекламодателя) [5];
- в качестве характеристик, от которых зависит ранжирование объявлений (следовательно, и порог входа в спецразмещение), могут быть взяты следующие: насколько фраза, по которой показывается объявление, является широко- или узко-тематической, информация о продавце в розничной торговле, а также информация о бренде, представляемом объявлением [6];
- позиционные модели показа рекламы [7];
- ранжирование вида  $Bid^k CTR^l$  в зависимости от степеней  $k$  и  $l$  может порождать различные виды сортировки: например при  $l = 0$  получается сортировка только по ставке, а при  $l = k = 1$  — классическая сортировка по критерию  $CPM$  [8].

**Порог входа в спецразмещение.** Условия входа в спецразмещение определяются двумя ограничениями. Прежде всего, число мест для показа рекламных объявлений в спецразмещении ограничено, поэтому в действующей системе для одного запроса допускается показ не более *трех* баннеров. Важно ввести еще один термин: *рекламный хит* — непустое множество баннеров, показанных одновременно по одному запросу пользователя (соответственно, один хит по спецразмещению не может содержать больше трех объявлений). Другое ограничение: считается, что не по всем поисковым запросам нужно показывать спецразмещение, так как объявления не всегда содержат полезную для пользователя информацию и менее релевантны его запросу, чем результаты поиска. Поэтому в системе показов рекламы принято условие, что *покрытие* — доля рекламных хитов со спецразмещением среди всех рекламных хитов — должно быть ограничено сверху. Для этого вводится абсолютный порог входа в спецразмещение. Если  $CPM$  баннера (для данного запроса) превышает порог, то это объявление становится кандидатом на показ в спецразмещении, иначе баннер заведомо не будет показан. Порог подбирается так, чтобы часть результатов поиска не сопровождалась рекламой в спецразмещении вообще.

Если число кандидатов, прошедших порог, не больше трех, то все они будут показаны. Если же их число превышает три, то в спецразмещении показываются только три баннера с наибольшим значением  $CPM$ .

Таким образом, для показа нового баннера необходимо, чтобы его  $CPM$  превзошел абсолютный порог. А если уже имеются три кандидата, то  $CPM$

нового баннера должен превзойти таковой у наихудшего (по критерию  $CPM$ ) среди имеющихся кандидатов, чтобы вытеснить последнего из списка кандидатов. Это условие определяет *относительный порог* входа в спецразмещение.

## 1. ПОСТАНОВКА ЗАДАЧИ

Полезность показа рекламных объявлений в спецразмещении для поисковой системы определяется следующими критериями.

- Сумма денег, списываемых со счетов рекламодателей. В этом случае оптимизация правила отбора баннеров в спецразмещение представляется несложной задачей: нужно просто подобрать такое значение абсолютного порога входа, при котором покрытие будет в точности равно заданному ограничению.
- Средняя кликабельность — насколько часто в среднем пользователь кликает на одно из показанных объявлений, что определяет удовлетворенность пользователя, а также эффективность самой рекламы. Эта величина соответствует среднему значению  $CTR$  по всем запросам и всем объявлениям, попавшим в спецразмещение. Если эта величина мала, то пользователь просто уйдет в другую поисковую систему.

Подробнее о балансе между ставкой и качеством выдачи (в нашем случае  $CTR$ ) можно прочесть в работе [7].

Здесь мы ставим задачу максимизации среднего значения  $CTR$  при ограничении на суммарный доход поисковой системы, поэтому она ставится как поиск условного максимума средней кликабельности в спецразмещении при условии, что сумма денег должна быть не меньше заданной величины.

Кроме того, сохраняются упомянутые ограничения на покрытие (долю показов со спецразмещением среди всех показов, содержащих рекламу) и на число рекламных объявлений в спецразмещении. Задача состоит в том, чтобы найти правило входа в спецразмещение, доставляющее максимум среднему значению  $CTR$  при упомянутых ограничениях. Но ситуация на поиске и рынке рекламы непрерывно меняется во времени, соответственно должны меняться и пороги, для подбора которых была разработана схема, представленная на рис. 2. Рассмотрим ее.

Периодически собирается случайная выборка запросов (см. блок 1 на рис. 2) из общего списка запросов из исторических логов данных за определенный период времени (неделя, месяц). По выборке запросов пользователей для каждого из запросов отбираются баннеры-кандидаты (блок 3) для показов (по фразам запроса). Считается, что

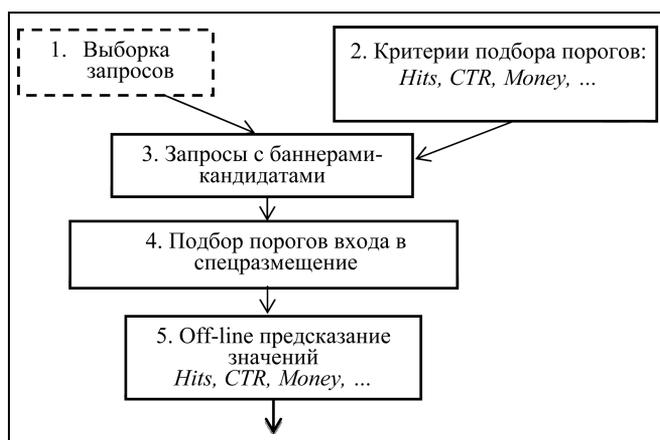


Рис. 2. Схема подбора порогов входа в спецразмещение

для каждого кандидата известна его ставка и предсказание кликабельности (отдельное для спецразмещения и для блока справа).

Критерии подбора порогов (блок 2): *Hits* — покрытие, о котором говорилось во Введении; *CTR* — кликабельность спецразмещения, т. е. число кликов по спецразмещению относительно числа показов рекламы в спецразмещении (считается суммарно по всей выборке запросов); *Money* — количество заработанных на спецразмещении денег (списанных со счетов рекламодателей), при подборе порогов оценивается как  $CTR \cdot Bid$ .

В блоке 4 реализуется отбор баннеров для их показа в спецразмещении или справа от результатов поиска. Не будем сейчас очень подробно останавливаться, как конкретно происходит отбор баннеров, важно, что после некоторого отсева, каждый баннер (чтобы он мог быть показан в спецразмещении) должен преодолеть порог входа и выдержать конкуренцию со стороны других кандидатов на показ в спецразмещении. Вид порога входа в спецразмещение, а также его подбор — сложная алгоритмическая и вычислительная задача, которую мы будем обсуждать далее.

Как только оптимальный порог получен, у нас есть off-line предсказание изменения системы по основным ее характеристикам. Также у нас появляется возможность запустить on-line эксперимент для проверки полученных off-line предсказаний. Если предсказания подтверждаются, то правило применяется ко всему потоку запросов, иначе выясняются причины расхождения. Пороги должны пересчитываться всякий раз при изменении формулы оценки *CTR*, при изменении требований на сумму денег или покрытие, а также по прошествии определенного времени.

Другие, не оптимальные модели, рассмотрены в работах [9, 10].

## 2. МАТЕМАТИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ

Сначала поставим задачу следующим образом: как для заданной выборки запросов и заданного списка баннеров найти оптимальную расстановку этих баннеров в спецразмещении 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 ограничения: минимальная сумма денег, которая должна поступить от рекламодателей, максимальное покрытие — доля результатов поиска, сопровождаемых спецразмещением, а также максимальное число баннеров (*k*) в каждом хите по спецразмещению. При этих ограничениях будем искать такую расстановку баннеров в спецразмещении, которая обеспечивает максимум средней кликабельности — среднего значения *CTR*.

Задачу будем решать в том приближении, когда рекламодатель платит за клик по своей ставке, а не по правилу аукциона второй цены (сами деньги будут слегка завышены, но разницу в характеристиках мы можем достаточно точно проследить).

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*,  $1 \leq i \leq N$ . Обозначим далее:

$Bid_i$  — ставка, назначенная рекламодателем за *i*-й баннер; все ставки неотрицательны, т. е.  $Bid_i \geq 0$ ;

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

Введем далее бинарную переменную  $t_{ij}$ , такую что  $t_{ij} = 1$  означает, что *i*-й баннер размещен над результатом поиска на *j*-й запрос, и  $t_{ij} = 0$  в противном случае. Именно оптимальные значения этой переменной мы будем искать.

Тогда  $TSUM = \sum_{i,j} t_{ij}$  будет означать суммарное число показов рекламных объявлений в спецразмещении по данной выборке запросов, а

$$CTR_0(t) = \sum_{i,j} CTR_{i,j} \cdot t_{i,j} / TSUM \quad (1)$$

— среднюю кликабельность — среднее значение по *CTR* по этим баннерам. Именно эту величину мы хотим максимизировать при заданных ограничениях.

**Ограничение по суммарному доходу.** Математическое ожидание суммы денег, списываемых со счета рекламодателя, равно вероятности клика на

данный баннер по заданному запросу, умноженной на ставку рекламодателя, выставленную для этого баннера, в случае его показа в спецразмещении. Если же он не показан, то и денег он принести не может. За оценку вероятности клика мы принимаем заданное значение  $CTR$ . Таким образом, сумма денег принимает значение  $Bid_k \cdot CTR_{i,j} \cdot t_{i,j}$ , где  $i$  — номер баннера, а  $j$  — номер запроса из нашей выборки. Тогда общая сумма денег, списываемых с рекламодателей по нашей выборке запросов, определяется как

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

Мы требуем, чтобы эта сумма была не меньше заданного значения  $MON_{\min}$ . Таким образом, ограничение принимает вид

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

**Ограничение на покрытие.** Напомним, что покрытием мы называем долю запросов, по которым в спецразмещении показано хотя бы одно рекламное объявление, среди всех результатов поиска, по которым была показана реклама. Формально, для этого введем еще одну бинарную переменную

$$\tau_j = \begin{cases} 1, & \text{если } \sum_i t_{ij} > 0, \\ 0 & \text{в противном случае,} \end{cases}$$

т. е.  $\tau_j = 1$ , если над результатом поиска размещен хоть один баннер, и  $\tau_j = 0$  в противном случае. Тогда ограничение на покрытие примет вид

$$Hit(t) = \sum_j \tau_j / M \leq Hit_{\max}. \quad (3)$$

**Ограничение на число баннеров,** размещаемых над каждым результатом поиска. Считаем, что это число не должно превышать  $k$ :

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

Итак, задача формально ставится следующим образом: максимизировать по бинарным переменным  $t_{ij}$  функцию (1) при ограничениях (2)—(4). Это задача дискретного программирования.

### 3. РЕШЕНИЕ ЗАДАЧИ ДЛЯ ЗАДАНОЙ ВЫБОРКИ ЗАПРОСОВ И БАННЕРОВ

#### 3.1. Переход от дискретной задачи к непрерывной

Сформулированная точная постановка задачи относится к классу задач дискретной оптимизации. Решать такие задачи трудно, обычно их ре-

шение требует комбинаторного перебора. Но, если такую задачу «погрузить» в непрерывное пространство, то часто для ее решения удается воспользоваться хорошо разработанными методами оптимизации в непрерывном пространстве искомых переменных. Мы также воспользуемся этим приемом и заменим бинарные переменные  $t_{ij}$  на непрерывные  $t_{ij}$ , подчиненные дополнительному ограничению:  $0 \leq t_{ij} \leq 1$ . Эта замена корректна, потому что, как мы увидим, при максимизации нашего критерия переменные все равно принимают крайние значения 0 или 1 (за исключением вырожденных случаев, которые будут отмечены далее). Поэтому решение непрерывной задачи совпадет с решением дискретной [1].

#### 3.2. Общий принцип — метод множителей Лагранжа

Согласно методу множителей Лагранжа, если требуется найти максимум некоторой функции  $F(x)$  при ограничениях  $\varphi_i(x) = 0$ , то можно эту задачу заменить на задачу поиска безусловного максимума функции  $F^*(x) = F(x) - \sum \lambda_i \varphi_i(x)$ , причем коэффициенты  $\lambda_i$  подбираются так, чтобы в точке максимума  $F^*(x)$  все ограничения  $\varphi_i(x) = 0$  выполнялись точно.

Если же ограничения имеют вид неравенств  $\varphi_i(x) \leq 0$ , то согласно теории Куна—Таккера [2] метод множителей Лагранжа модифицируется следующим образом. По-прежнему, ищется безусловный максимум функции  $F^*(x) = F(x) - \sum \lambda_i \varphi_i(x)$ , а коэффициенты подбираются так, чтобы выполнялись три условия:

- точка максимума функции  $F^*(x)$  должна удовлетворять ограничениям  $\varphi_i(x) \leq 0$ ,
- коэффициенты должны быть неотрицательны ( $\lambda_i \geq 0$ ),
- коэффициент  $\lambda_i$  должен быть равен нулю, если в точке максимума функции  $F^*(x)$  соответствующее ограничение не достигает предельного значения, т. е.  $\varphi_i(x) < 0$ .

На самом деле можно не все ограничения заменять слагаемыми в критерии, а часть из них оставить как ограничения [1], т. е. искать условный максимум функции

$$F^*(x) = F(x) - \sum_{i=1}^m \lambda_i \varphi_i(x)$$

при ограничениях  $\varphi_i(x) \leq 0$ ,  $i = m + 1, \dots, n$ . Требования к подбору коэффициентов остаются теми же, что были перечислены ранее.



Именно этим приемом, когда часть ограничений переводится в критерий (с последующим удовлетворением ограничений путем выбора соответствующих множителей Лагранжа), а часть сохраняется, мы и будем пользоваться в дальнейшем.

### 3.3. Применение метода множителей Лагранжа к решению поставленной задачи

#### 3.3.1. Добавление в критерий ограничения по деньгам

Итак, требуется найти максимум по  $t$  функции  $CRIT_0(t) = \sum CTR_{i,j} \cdot t_{i,j} / TSUM$  при ограничениях (2)–(4) и ограничении  $0 \leq t_{ij} \leq 1$ . Сначала переведем в критерий (с помощью множителя Лагранжа  $\lambda_1 \geq 0$ ) только ограничение (2). Запишем его в форме  $\varphi(x) \leq 0$ :

$$MON_{\min} - MONEY(t) \leq 0,$$

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

Получим новый критерий

$$CRIT_1(t) = \sum CTR_{i,j} \cdot t_{i,j} / \sum t_{i,j} - \lambda_1 (MON_{\min} - MONEY(t))$$

и будем искать его максимум при учете остальных ограничений.

#### 3.3.2. Дополнительное ограничение

Если бы наш критерий представлял собой сумму функций, каждая из которых зависит только от одной переменной  $t_{i,j}$ , то можно было бы оптимизировать каждую из этих функций по отдельности — провести декомпозицию задачи. Но в нашем случае это не так, поскольку знаменатель  $TSUM = \sum_{i,j} t_{i,j}$  в слагаемых  $\sum CTR_{i,j} \cdot t_{i,j} / \sum t_{i,j}$  зависит сразу от всех переменных. Для того чтобы все же добиться декомпозиции, предлагается следующее.

А. Сначала зафиксировать знаменатель  $TSUM = \sum_{i,j} t_{i,j}$  дополнительным требованием  $TSUM = T_0$ , где  $T_0$  — некоторая назначенная положительная константа, и найти максимум  $CRIT_1(t)$  с учетом этого дополнительного ограничения. Этот максимум будет зависеть от назначенного значения  $T_0$ .

Теперь наш критерий приобретает вид:

$$CRIT_1(t) = \sum CTR_{i,j} \cdot t_{i,j} / T_0 - \lambda_1 \left( MON_{\min} - \sum_{i,j} Bid_i \cdot CTR_{i,j} \cdot t_{i,j} \right),$$

и действительно становится суммой функций, зависящих только от одной переменной  $t_{ij}$ .

Решая задачу, найдем значения переменных  $t_{ij}$ , доставляющие условный максимум критерия  $CRIT_1(t)$  при заданных  $\lambda_1$  и  $T_0$ .

Б. Перебором по  $T_0$  найти то значение  $T_{0(\text{опт})}$ , при котором достигается максимум критерия  $CRIT_1(t)$ .

Переведем с помощью множителя Лагранжа и новое дополнительное ограничение в критерий. Получим:

$$CRIT_2(t) = \sum CTR_{ij} \cdot t_{ij} / T_0 - \lambda_1 \left( MON_{\min} - \sum_{i,j} Bid_i \cdot CTR_{ij} \cdot t_{ij} \right) - \lambda_2 \left( \sum_{i,j} t_{ij} - T_0 \right)$$

при ограничениях

$$0 \leq t_{ij} \leq 1, \forall j \sum_i t_{ij} \leq k, \quad Hit(t) = \sum_j \tau_j / M \leq Hit_{\max}$$

#### 3.4. Оптимизация критерия $CRIT_2(t)$ при оставшихся ограничениях

Итак, на данном шаге мы должны максимизировать критерий

$$CRIT_2(t) = \sum CTR_{i,j} \cdot t_{i,j} / T_0 - \lambda_1 \left( MON_{\min} - \sum_{i,j} Bid_i \cdot CTR_{i,j} \cdot t_{i,j} \right) - \lambda_2 \left( \sum_{i,j} t_{i,j} - T_0 \right),$$

считая величины  $\lambda_1$ ,  $\lambda_2$  и  $T_0$  фиксированными.

Объединяя слагаемые, зависящие от  $t_{ij}$ , и вынося за скобки  $t_{ij}$ , получим

$$CRIT_2(t) = \sum_{i,j} t_{ij} \left( \frac{CTR_{ij}}{T_0} + \lambda_1 Bid_i \cdot CTR_{ij} - \lambda_2 \right) +$$

const,

где величина const от переменных  $t_{ij}$  не зависит. Видим, что наш критерий действительно представляет собой сумму функций, каждая из которых линейно зависит только от одной переменной  $t_{ij}$ .

*Шаг 1.* Теперь найдем максимум критерия  $CRIT_2(t)$  по переменным  $t_{ij}$  с учетом только одного ограничения:  $0 \leq t_{ij} \leq 1$ .

Ясно, что при этом переменная  $t_{ij}$  должна принять значение «ноль», если коэффициент при ней отрицателен. Действительно, в этом случае вклад в критерий  $CRIT_2(t)$  от слагаемого  $t_{ij} (CTR_{ij}/T_0 + \lambda_1 Bid_i \cdot CTR_{ij} - \lambda_2)$  будет отрицательным при любом положительном значении  $t_{ij}$ , а при  $t_{ij} = 0$  этот вклад будет нулевым.

Если же этот коэффициент положителен, то без учета остальных ограничений переменная  $t_{ij}$  должна принять значение 1. Действительно в этом слу-

чае вклад в критерий будет максимальным (и положительным) при  $t_{ij} = 1$ .

В случае же, когда этот коэффициент в точности равен нулю при некоторых значениях пары  $(i, j)$ , мы получаем, что  $CRIT_2(t)$  не зависит от  $t_{ij}$ . При естественных значениях входных параметров точное равенство коэффициента нулю возможно только в вырожденных случаях. Но при подборе значений коэффициентов Лагранжа может оказаться необходимым выбрать дробное значение  $t_{ij}$ . Однако на практике такое возможно только для очень малого числа пар  $(i, j)$ . Действительно, мы подбираем совсем небольшое число неопределенных коэффициентов, и обеспечить точное равенство коэффициента нулю можно тоже в очень малом числе случаев. Учитывая, что в нашем пуле содержатся тысячи элементов, замена значения  $t_{ij}$  на 0 или 1 в этих случаях практически не влияет на значение критерия.

Но остальные ограничения могут сделать предъявление  $i$ -го баннера на  $j$ -й запрос все же недопустимым, т. е. все же придется положить  $t_{ij} = 0$ .

Обозначим  $F_{ij}(\lambda, T_0)$  коэффициент при  $t_{ij}$ :

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

Тогда кандидатами на вход в спецразмещение будут те баннеры, для которых  $F_{ij} > 0$  или, иначе,

$$CTR_{ij}/T_0 + \lambda_1 Bid_i \cdot CTR_{ij} > \lambda_2.$$

Таким образом, величина  $\lambda_2$  может считаться абсолютным порогом на попадание в спецразмещение.

Составим теперь для каждого ( $j$ -го) запроса список кандидатов баннеров на вход в спецразмещение, упорядочив их в порядке убывания величин  $F_{ij}(\lambda)$ . Некоторые из списков могут оказаться пустыми, если для всех  $i$  окажется, что  $F_{ij}(\lambda) \leq 0$ .

**Шаг 2.** Теперь учтем ограничение на количество баннеров, размещаемых сверху от результатов поиска:

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

Если количество баннеров в списке кандидатов для  $j$ -го запроса меньше  $k$ , то это ограничение выполнится автоматически. Иначе, для максимизации нашего критерия  $CRIT_2(t)$  следует оставить в каждом списке ровно  $k$  кандидатов с максимальным значением  $F_{ij}(\lambda)$ , т. е. первые  $k$  элементов списка. Для остальных элементов положить  $t_{ij} = 0$ . Назовем этот укороченный список *списком для показа*.

**Шаг 3.** Остается ограничение на покрытие:

$$Hit(t) = \sum_j \tau_j / M \leq Hit_{max}$$

или

$$\sum_j \tau_j \leq Hit_{max} \cdot M,$$

т. е. в спецразмещении должно присутствовать не более, чем в  $Hit_{max} \cdot M$  результатах поиска. Если и так количество запросов с непустым списком для показа меньше, чем  $Hit_{max} \cdot M$ , то ограничение выполнится автоматически. В противном случае часть списков следует обнулить.

Обозначим  $R_j(\lambda)$  вклад в критерий для каждого запроса от баннеров, вошедших в список:

$$R_j(\lambda) = \sum_i (CTR_{ij}/T_0 + \lambda_1 Bid_i \cdot CTR_{ij} - \lambda_2),$$

где сумма для каждого запроса берется только по баннерам, вошедшим в список для показа.

Ясно, что максимум критерия будет достигнут, если мы оставим только  $Hit_{max} \cdot M$  списков с наибольшим значением  $R_j(\lambda)$ . Иными словами, если мы упорядочим списки в порядке убывания  $R_j(\lambda)$ , то следует оставить только  $Hit_{max} \cdot M$  первых списков (если, конечно их хватит, иначе оставляем все). Обозначим  $\lambda_{3\text{ опт}}$  минимальное значение  $R_j(\lambda)$  по всем оставшимся спискам. Тогда будут оставлены только те списки, для которых  $R_j(\lambda) \geq \lambda_3$ . Разумеется, величина  $\lambda_3$  будет зависеть от принятых значений  $\lambda_1, \lambda_2$  и  $T_0$ .

Теперь можно окончательно зафиксировать оптимальные значения  $t_{ij}$ :

$$\begin{cases} t_{ij} = 1, \text{ если } i\text{-й баннер вошел в оставленный} \\ \text{усеченный список для } j\text{-го запроса,} \\ t_{ij} = 0 \text{ в противном случае.} \end{cases}$$

Этот результат мы получили для заданных значений  $\lambda_1, \lambda_2$  и  $T_0$ . Обозначим через  $t_{\text{опт}}(\lambda_1, \lambda_2, T_0)$  вектор, составленный из значений  $t_{ij}$  этого решения.

### 3.5. Выбор необходимого значения $\lambda_2$

Теперь мы должны найти то значение коэффициента  $\lambda_2$ , при котором действительно выполнится наше дополнительное ограничение  $TSUM = T_0$ .

Заметим, что зная значения  $t_{ij}$ , мы можем подсчитать величину  $TSUM = \sum_{i,j} t_{ij}$ , причем значение  $TSUM$  (как и значения  $t_{ij}$ ) будет зависеть от зафиксированных значений  $\lambda_1, \lambda_2$  и  $T_0$ . Но нас сейчас бу-



дет интересоваться только зависимостью  $TSUM$  от лагранжева коэффициента  $\lambda_2$ , считая по-прежнему значения  $\lambda_1$  и  $T_0$  фиксированными. Перебирая все значения  $\lambda_2$ , мы можем попытаться добиться равенства  $TSUM(\lambda_2) = T_0$ . Конечно, может случиться, что ни при каком значении  $\lambda_2$  это равенство не выполнится (в силу целочисленности  $T_0$  или по иной причине, но такие значения объявим недопустимыми). Найденное значение  $\lambda_{2\text{опт}}(\lambda_1, T_0)$  (абсолютного порога) будет зависеть от  $\lambda_1$  и  $T_0$ . Ему будет соответствовать решение  $t_{\text{опт}}(\lambda_1, \lambda_{2\text{опт}}(\lambda_1, T_0), T_0)$ .

### 3.6. Выбор оптимального значения $T_0$ путем максимизации критерия $CRIT_1(t)$

Теперь нам нужно избавиться от дополнительного ограничения  $TSUM = T_0$ , поскольку оно не входит в число ограничений исходной постановки задачи. Сделать это можно с помощью максимизации критерия

$$CRIT_1(t, \lambda_1, T_0) = \sum CTR_{ij} t_{ij} / T_0 - \lambda_1 (MON_{\min} - MONEY(t))$$

путем перебора по всем допустимым значениям  $T_0$  при фиксированном значении  $\lambda_1$ . При этом в качестве  $t_{ij}$  должны браться значения, найденные при  $\lambda_2 = \lambda_{2\text{опт}}(\lambda_1, T_0)$ . Это значение заведомо обеспечивает равенство  $TSUM(\lambda_1, \lambda_2, T_0) = T_0$  в силу операции, описанной в п. 3.5.

Обозначим  $T_{0\text{опт}}(\lambda_1)$  то значение  $T_0$ , при котором достигается максимум критерия  $CRIT_1$ . Оно будет зависеть от значения  $\lambda_1$ . Ему будут соответствовать значения  $\lambda_{2\text{опт}}(\lambda_1, T_{0\text{опт}})$ ,  $\lambda_{3\text{опт}}$  и решение  $t_{\text{опт}}(\lambda_1, \lambda_{2\text{опт}}(\lambda_1, T_{0\text{опт}}), T_{0\text{опт}}(\lambda_1))$ .

### 3.7. Оптимизация критерия $CRIT_0(t)$ : выбор требуемого значения $\lambda_1$

Найденные значения  $T_{0\text{опт}}$ ,  $\lambda_{2\text{опт}}$ ,  $\lambda_{3\text{опт}}$  и  $t_{\text{опт}}$  будут зависеть от параметра  $\lambda_1$ . Согласно теории Куна—Таккера, для максимизации нашего основного критерия  $CRIT_0(t)$  остается выбрать правильное значение параметра  $\lambda_1 \geq 0$ . Его нужно выбрать так, чтобы выполнилось ограничение  $MONEY(t) \leq MON_{\min}$ . Он может быть отличен от нуля, если только неравенство по деньгам переходит в равенство:

$$MONEY(t) = MON_{\min}. \quad (5)$$

При  $\lambda_1 = 0$  в критерии  $CRIT_1(t)$  деньги вообще не учитываются, и если при этом вырученных де-

нег хватает, то неравенство по деньгам выполнится автоматически. Если же денег не хватает, то следует увеличивать  $\lambda_1$ , пока не выполнится равенство  $MONEY(t) = MON_{\min}$ . Вырученные деньги с ростом  $\lambda_1$ , во всяком случае, не убывают, так как величина  $MONEY$  входит в критерий  $CRIT_1(t)$  все с большим весом. Если же при сколь угодно большом значении  $\lambda_1$  денег все равно не хватает, то задача вообще неразрешима. В нормальном случае перебором по  $\lambda_1$  найдем то значение  $\lambda_{1\text{опт}}$ , при котором выполнится равенство (5). Соответственно определяются значения  $T_{0\text{опт}}(T_{0\text{опт}}(\lambda_{1\text{опт}}), \lambda_{2\text{опт}}(\lambda_{1\text{опт}}, T_{0\text{опт}}))$  и решение  $t_{\text{опт}}(\lambda_{1\text{опт}}, \lambda_{2\text{опт}}, T_{0\text{опт}})$ , и задача будет решена полностью.

Заметим, что если уже выбраны параметры (пороги)  $\lambda_{1\text{опт}}$ ,  $T_{0\text{опт}}$ ,  $\lambda_{2\text{опт}}$ , и  $\lambda_{3\text{опт}}$ , то для заданного ( $j$ -го) запроса определить, какие баннеры должны войти в спецразмещение по этому запросу (определить  $t_{j\text{опт}}$ ), можно, не обращая внимания на другие запросы. Для этого нужно:

1) вычислить значение

$$F_{ij} = CTR_{ij} / T_{0\text{опт}} + \lambda_{1\text{опт}} Bid_i \cdot CTR_{ij} - \lambda_{2\text{опт}}$$

для всех баннеров, совместимых с данным запросом;

2) составить список кандидатов для показа в спецразмещении, в него поместить только те баннеры, для которых

$$CTR_{ij} / T_{0\text{опт}} + \lambda_{1\text{опт}} Bid_i \cdot CTR_{ij} > \lambda_{2\text{опт}}$$

(этот список может быть и пустым);

3) если число кандидатов больше  $k$ , то сократить список, оставив в нем только  $k$  баннеров с наибольшим значением  $F_{ij}$ ; это и будет список для показа;

4) подсчитать значение

$$R_j(\lambda) = \sum_i (CTR_{ij} / T_{0\text{опт}} / T_{0\text{опт}} + \lambda_{1\text{опт}} Bid_i \cdot CTR_{ij} - \lambda_{2\text{опт}}),$$

где сумма берется по баннерам из списка кандидатов.

5) если эта величина меньше  $\lambda_{3\text{опт}}$ , то обнулить список, и для данного запроса не показывать никаких баннеров в спецразмещении. Иначе показать все баннеры из списка для показа, т. е. положить  $t_{j\text{опт}} = 1$ , остальные значения  $t_{ij}$  обнулить.

В этой возможности определить значение  $t_{j\text{опт}}$ , не обращая внимания на другие запросы, и есть смысл декомпозиции.

**4. ФОРМАЛЬНОЕ ОПИСАНИЕ АЛГОРИТМА ПОДБОРА ПОРОГОВ ВХОДА В СПЕЦРАЗМЕЩЕНИЕ**

Запишем общий алгоритм подбора порогов для показа объявлений в спецразмещении.

Положим:  $M$  — 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 нашей выборке запросов.

Для каждого запроса и баннера-кандидата дано:

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

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

Цикл по  $\lambda_1$  (например, от 0 вверх).

Вырученные деньги с ростом  $\lambda_1$  не убывают, т. е. можно сказать, что  $\lambda_1$  — параметр, регулирующий поступление денег от рекламодателей.

Цикл по  $T_0$

$T_0$  — общее число показов рекламы в спецразмещении по всей выборке запросов.

Цикл по  $\lambda_2$ .

Величина  $\lambda_2$  может считаться абсолютным порогом на попадание в спецразмещение, она регулирует число показов в хите по одному запросу.

Цикл по запросам  $j$ .

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

обнулить списки;

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

если  $F_{ij} > 0$ , то попытаться включить пару  $(i, j)$  в  $j$ -й список:

Если в  $j$ -м списке объектов меньше чем  $k$ , то включить баннер  $i$  в список, иначе попытаться вытеснить худший по  $F_{ij}$  и вставить  $i$ -й, а если текущий баннер хуже всех по критерию  $F_{ij}$ , то не включать его.

Вычислить вклад в суммарный критерий баннеров, вошедших в список для показа в спецразмещении:

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

Положить:

$$\tau_j = \begin{cases} 1, & \text{если список кандидатов на показ не пустеет,} \\ 0 & \text{в противном случае.} \end{cases}$$

Конец цикла по запросам  $j$ .

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

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

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

Положить

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

Вычислить  $TSUM = \sum_{i,j} \tau_{ij}$  (величина  $TSUM$  будет зависеть от параметров  $\lambda_1, \lambda_2, T_0$ ).

Менять  $\lambda_2$ , пока не будет достигнуто  $TSUM = T_0 \pm \varepsilon$ . Положить  $\lambda_{2\text{опт}} = \lambda_2$ .

Запомнить  $\lambda_3, \lambda_{2\text{опт}}$ , списки  $t_{ij}$  (т. е. баннеры, отобранные для показа)

Конец цикла по  $\lambda_2$ .

Вычислить

$$CRIT_1(\lambda_1, \lambda_{2\text{опт}}, T_0) = \sum_{i,j} t_{ij} (CTR_{ij}/T_0 - \lambda_1 (MON_{\min} - MONEY(t))).$$

Менять  $T_0$ , чтобы достигнуть  $\max_{T_0} CRIT_1(t)$ .

Положить  $T_{0\text{опт}}$  то значение  $T_0$ , при котором достигнут  $\max_{T_0} CRIT_1$ .

Запомнить  $\lambda_3, \lambda_{2\text{опт}}, T_{0\text{опт}}, MONEY(t)$ .

Конец цикла по  $T_0$ .

Менять  $\lambda_1$ , пока  $MONEY(t) < MON_{\min}$ .

Конец цикла по  $\lambda_1$ .

В конце работы алгоритма мы получаем четыре величины  $\lambda_{1\text{опт}}, T_{0\text{опт}}, \lambda_{2\text{опт}}$  и  $\lambda_{3\text{опт}}$ , которые будут использоваться для работы с новыми запросами.



## 5. РАБОТА С НОВЫМИ ЗАПРОСАМИ

Рассмотрим, каким образом мы будем использовать полученные с помощью описанного в § 4 алгоритма величины  $\lambda_{1\text{опт}}$ ,  $T_{0\text{опт}}$ ,  $\lambda_{3\text{опт}}$ ,  $\lambda_{2\text{опт}}$  для обработки нового запроса, поступившего от пользователя в систему показов рекламных объявлений.

Ранее было отмечено, что если данные величины уже получены, то для каждого запроса из выборки запросов можно определить, какие баннеры должны быть показаны в спецразмещении, не обращая внимания на другие запросы. Мы рассчитываем на то, что используемая выборка запросов достаточно велика, чтобы результат подбора порогов применялся в тех же условиях, в которых собиралась эта выборка. Тогда, в силу закона больших чисел, если мы будем применять те же правила к новым запросам, то интегральные критерии — среднее значение  $CTR$ ,  $CPM$  и покрытие — будут достаточно близки к значениям, полученным по выборке запросов. В то же время параметры были подобраны так, чтобы наш главный критерий — среднее значение  $CTR$  по спецразмещению — достиг максимума при заданных ограничениях на среднее значение  $CPM$  и покрытие. Поэтому и на новых данных с достаточной точностью в среднем критерий будет достигать максимума, а ограничения выполняться.

Итак, получив новый запрос, нужно выполнить следующие действия.

1. Отобрать баннеры-кандидаты для возможных показов (по фразам запроса). Для каждого из этих баннеров известно значение ставки  $Bid_i$  и прогноз кликабельности  $CTR_{ij}$ , где  $i$  — индекс баннера, а  $j$  — индекс запроса.

2. Вычислить значение  $F_{ij} = CTR_{ij}/T_{0\text{опт}} + \lambda_{1\text{опт}} \cdot Bid_i \cdot CTR_{ij} - \lambda_{2\text{опт}}$  для всех полученных в п. 1 баннеров.

3. Составить список кандидатов на показ в спецразмещении. В этот список для показа поместить только те баннеры, для которых  $F_{ij} > 0$ . Этот список может быть и пустым (в этом случае в спецразмещении не будет показано ни одного баннера).

4. Если число кандидатов больше  $k$ , то сократить список, оставив в нем только  $k$  баннеров с наибольшим значением  $F_{ij}$ .

5. Подсчитать значение  $R_j = \sum_i F_{ij}$ , где сумма берется по всем баннерам из усеченного списка.

6. Если  $R_j < \lambda_{3\text{опт}}$ , то обнулить список, и для данного запроса не показывать никаких баннеров в спецразмещении. Иначе показать в спецразмещении все баннеры из усеченного списка.

## ЗАКЛЮЧЕНИЕ

Итак, рассмотрена проблема отбора баннеров-кандидатов для показа в спецразмещении для заданных ограничений по деньгам, а также по покрытию с целью максимизации средней клика-

бельности баннеров в спецразмещении. Предложен алгоритм отбора, а также формула входа в спецразмещение. Однако в действительности разработанный алгоритм отбора показателей сложнее, чем описано в данной работе. Например, рассмотрены такие модификации, как:

— сортировка по ставке  $Bid_i$  при показе объявлений;

— упрощение и ускорение самого алгоритма подбора порогов для входа в спецразмещение (изложенный в данной статье алгоритм достаточно сложен по объему переборов);

— разные модели для клика по объявлению (в том числе учет зависимости вероятности клика от позиции баннера).

На данный момент предложенный алгоритм уже применяется для некоторой части рекламных показов в спецразмещении. В результате покрытие увеличилось на 0,5 %, доход на 0,9 %, а значение  $CTR$  увеличилось на 4,53 %. Но все это будет изложено в отдельной публикации.

## ЛИТЕРАТУРА

1. Fain D.C., Pedersen J.O. Sponsored search: A brief history. — URL: <http://www.asis.org/Bulletin/Dec-05/pedersen.html> (дата обращения 17.01.2013).
2. Kumar R., Mahdian M., Li E., et al. Systems and methods for exploring new sponsored search listings of uncertain quality. — 2009.
3. Graepel T., Candela J.Q., Borchert T., Herbrich R. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine // Proc. of ICML. — 2010. — P. 13–20.
4. 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. — Iss. 1. — P. 242–259.
5. Feng J., Bhargava H.K., Pennock D.M. Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms // Inform Journal on Computing. — 2007. — Vol. 19, N 1. — P. 137–148.
6. Ghose A., Yang S. An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets // Net Institute Working Paper. — 2009. — May 25.
7. Graepel H., Candela J.Q., Borchert T., Herbrich R. Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine // Proc. of the 27th Intern. Conf. on Machine Learning ICML 2010.
8. Lahaie S., M. Pennock D.M. Revenue analysis of a family of ranking rules for keyword auctions // Proc. of the 8th ACM conf. on Electronic commerce, 2007. — P. 50–56.
9. Richardson M., Dominowska E., and Ragno R. Predicting clicks: Estimating the click-through rate for new ads // Proc. of the 16th Intern. World Wide Web Conf (WWW — 2007). — ACM Press, 2007.
10. Zhu Z. A., Chen W., Minka T., et al. A novel click model and its applications to online advertising // In Proc. Of WSDM, 2010. — P. 321–330.
11. Поляк Б.Т. Введение в оптимизацию. — М.: Наука, 1983. — 384 с.
12. Линейные неравенства и смежные вопросы / Сб. статей под ред. Г.У. Куна, А.У. Таккера. — М.: Иностранная литература, 1959. — 469 с.

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

Анна Николаевна Корнетова — разработчик, компания «Яндекс»; аспирант, Национальный исследовательский институт — Высшая школа экономики, г. Москва, ✉ akornet@yandex-team.ru,

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