

Санкт-Петербургский Государственный Политехнический Университет
Институт прикладной математики и механики
Кафедра прикладной математики

Диссертация допущена к защите
Зав. кафедрой

"__" _____

**ДИССЕРТАЦИЯ
на соискание степени МАГИСТРА**

Тема: метод ранжирования разнородных результатов поиска

Направление: 01.04.02 - Прикладная математика и информатика
Магистерская программа: системное программирование

Выполнил студент гр. 63601/2 _____ Толмачев А.С.

Руководитель, к.ф.-м.н., доц. _____ Иванков А.А.

Консультанты:

по вопросам информационного поиска, к.ф.-м.н. _____ Кураленок И.Е.

по вопросам охраны труда, к.т.н., доц. _____ Монашков В.В.

Санкт-Петербург
2015

Содержание

Введение	3
1 Обзор литературы	9
1.1 Классическая задача ранжирования	9
1.1.1 Формулировка задачи	9
1.1.2 Обзор методов решения	9
1.2 Ранжирование специализированных ответов	10
1.2.1 Формулировка задачи	10
1.2.2 Обзор методов решения	10
2 Описание метода	11
2.1 Неформальная постановка задачи	11
2.2 Основные идеи	11
2.3 Критерий полезности поисковой выдачи	12
2.4 Формальная постановка задачи	13
2.5 Построение модели оценки полезности поисковой выдачи	14
2.6 Алгоритм ранжирования	15
2.6.1 Базовый алгоритм	16
2.6.2 “Жадный” вариант алгоритма	16
3 Программная реализация	18
3.1 Схема системы ранжирования	18
3.2 Уменьшение числа обращений к поисковым источникам	18
4 Оценка качества работы метода	19

4.1	Методы оценки качества поиска	19
4.1.1	Методы, основанные на экспертных оценках	19
4.1.2	Методы, основанные на поведении пользователей	21
4.2	Описание результатов	21
5	Вопросы охраны труда	22
5.1	Общая характеристика санитарно-гигиенических условий труда	22
5.2	Эргономические требования	22
5.3	Микроклиматические условия	22
5.4	Уровень шума	22
5.5	Системы освещения	22
5.6	Излучения	22
5.7	Электробезопасность	22
5.8	Инженерно-технические мероприятия по созданию благоприятных условий труда	22
5.9	Методика и приборы контроля параметров среды	22
Заключение		23
Список литературы		24

Введение

С развитием информационных систем и ростом их популярности растет и количество информации, производимой с их помощью. Так, по данным аналитической компании IDC (International Data Corporation) общий объем цифровой информации в мире составил на 2013 год примерно 4.4 зеттабайт¹, он увеличивается каждый год примерно на 40% и к 2020 году составит приблизительно 44 зеттабайт [23]. Существенная доля этой информации – информация, размещенная во всемирной сети Интернет. Она большей частью неструктурирована и очень разнообразна. Несомненно, без помощи поисковых систем ориентироваться в этом огромном информационном пространстве не представляется возможным. Поэтому системы веб-поиска на сегодняшний день играют очень важную роль в нашей жизни.

С момента своего возникновения веб-поисковые системы активно развиваются. Одно из современных направлений их развития касается смешивания в результатах поиска разнотипной информации. Первые системы поиска в интернете в ответ на запрос выдавали список ссылок на веб-страницы (рис. 1) [24]. Такой вид результатов поиска для того времени был естественным, поскольку представляемая в них информация была однотипной. Но по мере того, как развивались интернет-технологии и увеличивалось число интернет-пользователей, информация, размещаемая в сети, становилось все более разнообразной. На сегодняшний день это разнообразие огромно: в интернете можно найти музыку, фильмы, новости, тексты книг, научные статьи, программные приложения, кулинарные рецепты, технические характеристики товаров и отзывы о них и т. д. – все это различные типы информации. В связи с этим получили развитие системы, предназначенные для агрегации и поиска информации определенного типа. К таковым относятся, например, системы поиска изображений, видео-записей, ново-

¹1 зеттабайт (ЗБ) = 1 триллион гигабайт

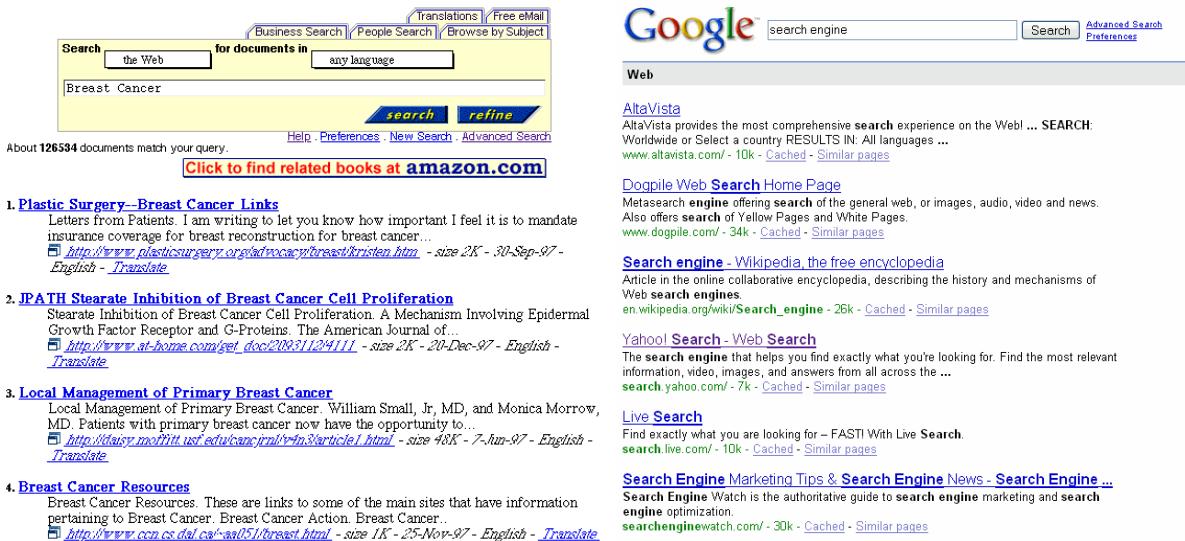


Рис. 1: Страницы результатов поиска системы AltaVista и одной из первых версий системы Google.

стей, товаров, музыки. Ясно, что такая специализированная система может быть более удобной и полезной для решения поисковой задачи из соответствующей области, чем система общего назначения. Действительно, если пользователь, к примеру, ищет фотографии Дворцовой площади, то, очевидно, ему будет гораздо удобнее, если в качестве результатов поиска он будет видеть именно фотографии, а не ссылки на веб-страницы – ему не нужно будет переходить по этим ссылкам и самостоятельно исследовать страницы в поисках фотографий. Но выбирать каждый раз, к какой из многочисленных специализированных систем обратиться, неудобно. К тому же пользователь может не знать о существовании той или иной специализированной системы, а для каких-то поисковых задач такой системы может и не быть. Поэтому возникает естественное желание, чтобы система веб-поиска сама “понимала” запрос пользователя, и выдавала в ответ информацию нужного типа. Однако ограничивать ответ на запрос каким-то одним типом информации также неоправданно – для решения своей поисковой задачи пользователю может быть полезна информация сразу нескольких типов. Так, например, при поиске информации о музыкальном исполнителе может быть полезна и его биография, и фотографии с ним, и аудио-записи исполняемой им музыки, и видео-сюжеты о нем. Или же поисковый запрос может быть многозначным – например, по запросу “политика” пользователя может интересовать как информация, касающаяся самого термина, так и политические новости. Стандартным на сегодняшний день решением является смешивание результатов поиска от разных специализированных систем и представление

их вместе с традиционными результатами веб-поиска – списком ссылок на интернет-страницы. Такое смешивание мы можем наблюдать, пользуясь современными поисковыми системами. Например, в результатах поиска систем Яндекс и Google можно увидеть разнообразные специализированные результаты: на поисковый запрос о картинах – результаты поиска по изображениям, на запрос об адресе в городе – интерактивную карту с отмеченным адресом, на запрос о новостях – результаты поиска по новостям, а на запрос о кафе – специализированный ответ с найденными заведениями и информацией о них (рис. 2). Таким образом, результаты поиска могут быть *разнородными*, поскольку могут содержать информацию разных типов.

Процесс обработки поискового запроса можно условно разделить на два этапа: поиск информации, соответствующей запросу, и представление найденной информации. На первом этапе из всего множества объектов, известных системе, выбираются те, которые по тем или иным критериям соответствуют заданному запросу. На втором этапе

The image displays two side-by-side search result pages. On the left is a Yandex search page for 'картины маринистов' (Marinist paintings), showing a search bar, a sidebar with filters like 'Поиск', 'Картинки', 'Видéo', etc., and a main content area with search results for 'Море, картины художников маринисты...' and 'Картины маринистов - корабли'. It also features a 'Смотрите картинки по запросу "картины маринист...' section with four thumbnail images of paintings. On the right is a Yandex search page for 'невский проспект 20' (Nevsky Prospekt 20), showing a search bar, a sidebar with filters, and a main content area featuring a map of the area around Nevsky Prospekt 20, with a blue marker indicating the location. Below the map is a snippet about the building 'Biblioteka Food and the City'. To the right is a Google search page for 'новости спорта' (Sports news), showing a search bar, a sidebar with filters, and a main content area with news snippets from 'Новости спорта' and 'Спортивная аналитика'. At the bottom is a Google search page for 'кафе в санкт петербурге' (Cafes in Saint Petersburg), showing a search bar, a sidebar with filters, and a main content area with a list of cafe reviews and a 'Еще результаты по запросу "кафе"' button.

Рис. 2: Специализированные ответы в результатах поиска систем Яндекс и Google.

множество найденных объектов представляется некоторым образом и выдается пользователю. Задача *ранжирования* относится ко второму из этих этапов. Ранжирование – это упорядочение результатов поиска в соответствии с некоторым принципом [19, 15]. От того, как упорядочены результаты, во многом зависит то, насколько успешно пользователь сможет решить свою поисковую задачу. Так, например, если пользователь задал запрос “скололазание википедия”, то, вероятно, он ищет статью о скалолазании из интернет-энциклопедии Википедия. Если среди найденных результатов эта статья присутствует, то разумно расположить ее первой в списке результатов поиска, чтобы пользователь смог сразу ею воспользоваться. В противном случае ему будет сложнее найти этот результат среди остальных, а если расположить его за пределами первых десяти результатов, то он и вовсе может решить, что эта статья не была найдена. Каждый из специализированных ответов, встраиваемых в результаты поиска, также может быть более или менее полезен пользователю в сравнении с другими результатами при решении своей поисковой задачи. Кроме этого, какие-то из специализированных ответов могут быть в принципе не уместны для заданного поискового запроса. Таким образом, возникает задача выбора специализированных результатов для показа по запросу и ранжирования их в поисковой выдаче.

Задача ранжирования, возникающая при смешивании специализированных ответов с традиционными результатами веб-поиска, существенно отличается от классической задачи ранжирования [15]. Во-первых, классическая задача формулируется для однотипных объектов, в то время как в рассматриваемой задаче ранжируемые результаты разнородны. Во-вторых, специализированные ответы встраиваются только на первую страницу результатов поиска, поэтому количество ранжируемых объектов при этом невелико. Эти отличия, в особенности первое из них, не позволяют применить для решения рассматриваемой задачи существующие методы решения классической задачи. Соответственно, требуется либо их обобщение, либо разработка принципиально новых подходов.

Также следует отметить, что список, – не единственный способ представления результатов поиска. Модели поисковой выдачи могут быть различными. Например, поисковая выдача системы Google для настольных компьютеров имеет две колонки, в левой из которых располагается список результатов, а в правой могут располагаться специализированные ответы (рис. 3). А выдача мобильного приложения поисковой системы

Рис. 3: Поисковая выдача системы Google со специализированным результатом в отдельной колонке.

Яндекс состоит из страниц, каждая из которых может содержать список результатов или специализированные ответы. Набор этих страниц и их порядок зависит от запроса (рис. 4). В таких случаях задача ранжирования усложняется и превращается в задачу расположения результатов поиска в соответствии с заданной моделью поисковой выдачи. Это также требует обобщения существующих методов ранжирования.

Рис. 4: Выдача мобильного приложения поисковой системы Яндекс с результатами поиска на отдельных страницах.

В данной работе рассматривается задача ранжирования разнородных результатов поиска и предлагается универсальный метод ее решения. Этот метод может быть применен для ранжирования результатов произвольного вида и для разных моделей поисковой выдачи. Также он не требует экспертных оценок, поскольку основывается на поведении пользователей. (TODO: Описать, что в каком разделе).

Глава 1

Обзор литературы

Задача ранжирования занимает важное место в информационном поиске, и на эту тему имеется достаточно большое количество публикаций. Однако подавляющее большинство из них относится к задаче ранжирования однотипных объектов (классической задаче ранжирования) [15]. При встраивании же специализированных ответов в результаты поиска требуется ранжировать разнородные объекты. Это не дает возможности применить имеющиеся методы решения классической задачи – главным образом потому, что разнородные объекты нельзя описать одним набором признаков. В связи с этим рассматриваемая задача требует отдельных исследований, которых на данный момент пока не так много.

В этой главе сначала даются краткие сведения о классической задаче ранжирования и основных методах ее решения. Затем проводится обзор имеющихся на данный момент публикаций, посвященных ранжированию разнородных результатов.

1.1 Классическая задача ранжирования

1.1.1 Формулировка задачи

(TODO:)

1.1.2 Обзор методов решения

(TODO:)

1.2 Ранжирование специализированных ответов

1.2.1 Формулировка задачи

(TODO:)

1.2.2 Обзор методов решения

Наиболее ранние исследования в области смещивания разнородных результатов поиска касаются встраивания одного конкретного специализированного ответа на первое место в поисковой выдаче [4, 12, 14] и встраивания одного из нескольких специализированных ответов так же на первое место [1, 2, 5]. Однако подобные методы подходят лишь в тех случаях, когда для заданного запроса уместен только один специализированный результат, и этот результат наиболее подходящий из всех остальных. В более поздних исследованиях рассматривается возможность встраивания специализированных ответов на различные позиции в поисковой выдаче [13, 16, 3, 8]. Однако методы, предлагаемые в работах [13, 16, 3] не достаточно универсальны – они предназначены для встраивания какого-то конкретного специализированного ответа, или существенно зависят от специфики используемых специализированных ответов. Также методы [16, 3] используют экспертные оценки. Это делает их применение дорогостоящим и затрудняет регулярное обновление используемых машинно-обученных моделей, необходимое для учета изменений в потоке поисковых запросов.

(TODO: Расширить обзор подробнее – сказать подробнее о методах из публикаций)

Глава 2

Описание метода

2.1 Неформальная постановка задачи

Решаемую задачу можно неформально сформулировать следующим образом:

Дано: поисковый запрос и результаты поиска по нему – упорядоченный список ссылок на веб-страницы (веб-результаты) и набор специализированных ответов.

Требуется: выбрать специализированные ответы и встроить их в список веб-результатов таким образом, чтобы максимизировать полезность получаемой поисковой выдачи в контексте заданного поискового запроса.

(TODO: что-то добавить по задаче?)

Имея постановку задачи в таком неформальном виде, можно обозначить основные идеи ее решения. Затем требуется определить формальный критерий полезности поисковой выдачи, чтобы перейти к решению задачи в формальном виде.

2.2 Основные идеи

В основе предлагаемого метода для ранжирования разнородных результатов поиска лежат следующие идеи:

- рассматривать при ранжировании получаемую поисковую выдачу в целом, а не результаты по-отдельности;
- для оценки полезности поисковой выдачи использовать действия, совершаемые пользователями.

В общих чертах метод можно описать следующим образом:

1. Определить набор действий-индикаторов, совершаемых пользователем на поисковой выдаче, которые свидетельствуют о том, что выданные результаты были полезны, и поиск завершился успешно.
2. Построить модель, которая по поисковой выдаче будет предсказывать, какова вероятность того, что пользователь совершил на ней какое-то из действий-индикаторов.
3. Разработать алгоритм, который будет составлять разные варианты расположения результатов поиска, оценивать для каждого из них с помощью построенной модели вероятность совершения какого-либо из действий-индикаторов, и в качестве результата возвращать тот вариант, для которого оценка этой вероятности наибольшая.

В последующих подразделах дается более подробное описание каждого из намеченных шагов.

2.3 Критерий полезности поисковой выдачи

Как уже отмечалось выше, для оценки полезности поисковой выдачи предлагаются ввести набор действий-индикаторов. И затем считать, что выдача была полезна, если пользователь совершил какое-то из этих действий, и была не полезна в противном случае. Выбор набора действий-индикаторов играет основополагающую роль в предлагаемом методе. Состав этого набора в первую очередь зависит от того, какое взаимодействие с тем или иным результатом ожидается от пользователя. Так, например, результат, содержащий видео-запись, предполагает ее просмотр, результат с аудио-записью предполагает ее прослушивание и т.д. Также этот набор зависит от технических возможностей системы логирования пользовательских действий. Выбор тех или иных действий в качестве индикаторов должен следовать из продуктовых особенностей ранжируемых результатов или статистических обоснований.

Основной тип действий, которые пользователь совершает на поисковой выдаче – это клики по результатам поиска. В работах [9, 6, 18, 10] исследуется связь между характером поведения пользователя и его удовлетворенностью предоставленными результатами поиска. В этих работах рассматривается следующий эвристический критерий

для оценки полезности неспециализированного результата поиска: результат считается полезным пользователю, если он перешел по ссылке и не возвращался на поисковую выдачу в течение T и более секунд, где T – некоторый порог. В работе [6] приводится статистическое обоснование для значения $T = 30$. Этот эвристический критерий получил широкое распространение [18, 10] и называется критерием “длинного клика” (long dwell time click). (TODO: подробнее по используемые результаты из статей)

Большинство специализированные ответов, встраиваемых в поисковую выдачу, для взаимодействия также предполагают клик. Кликнув на специализированный ответ, пользователь переходит в соответствующую специализированную поисковую систему или веб-сервис. Это во многом схоже с переходом на веб-страницу. Таким образом, критерий “длинного клика” может быть также применен и для оценки полезности специализированных результатов.

На основании этих соображений был выбран набор действий-индикаторов, состоящий из одного элемента – “длинного клика”. При таком наборе формальный критерий полезности поисковой выдачи – это вероятность того, что пользователь совершил “длинный клик” по какому-то из результатов. Будем считать, что чем больше эта вероятность, тем полезнее поисковая выдача будет для пользователя.

2.4 Формальная постановка задачи

Определив формальный критерий полезности поисковой выдачи, рассматриваемую задачу можно сформулировать следующим образом:

Дано:

1. Поисковый запрос q .
2. Упорядоченное множество веб-результатов $W = \{w_1, \dots, w_k\}$, $\forall w_i, w_j, 1 \leq i < j \leq k : w_i \prec w_j$.
3. Специализированные ответы K типов:

$$S_1 = \{s_1^1, \dots, s_{k_1}^1\},$$

...

$$S_K = \{s_1^K, \dots, s_{k_K}^K\}.$$

Требуется: построить упорядоченное множество результатов поиска $R = \{r_1, \dots, r_n\}, \forall r_i, r_j, 1 \leq i < j \leq n : r_i \prec r_j$, такое, что:

1. $R \subseteq W \cup \bigcup_{i=1}^K S_i - R$ состоит из веб-результатов и специализированных ответов.
2. $\forall w_i, w_j \in W \exists r_l, r_m \in R$, т. ч. $r_l = w_i, r_m = w_j \ \& \ w_i \prec w_j \Leftrightarrow r_l \prec r_m$ – все веб-результаты из W присутствуют в R , и их порядок сохраняется.
3. $R = \arg \max_R P(LDTClick | (R, q))$ – R максимизирует вероятность “длинного клика”.

(TODO: что-то пояснить про постановку?)

2.5 Построение модели оценки полезности поисковой выдачи

Для решения задачи, сформулированной в конце предыдущего подраздела, требуется построить модель, которая позволит оценивать вероятность “длинного клика” для различных вариантов расположений результатов поиска. Эта модель должна принимать на вход поисковый запрос и вариант расположения результатов, и выдавать оценку вероятности “длинного клика”.

Построение этой модели заключалось в обучении бинарного классификатора. Рассматривалась следующая задача:

Дано: множество классов $\mathbb{C} = \{0, 1\}$ и обучающая выборка $L = \{(x_1, y_1), \dots, (x_N, y_N)\}$, где $x_i = (R_i, q_i) \in X$, X – множество всех возможных пар (R, q) , $y_i = y(x_i) \in \mathbb{C}$. Значение $y(x_i) = 1$ означает, что пользователь, получив в ответ на запрос q_i поисковую выдачу R_i , совершил на ней “длинный клик”, $y(x_i) = 0$ означает обратное.

Требуется: построить функцию $F : X \rightarrow [0, 1]$, такую, что $\forall x \in X$ значение $F(x)$ приближает значение $P(y(x) = 1)$ (вероятность принадлежности x к классу 1).

Для обучения классификатора требуется описать пару (R, q) вектором признаков, то есть ввести отображение $f : X \rightarrow \mathbb{R}^m$. Набор признаков должен удовлетворять следующим требованиям:

- наличие признаков, характеризующих как запрос, так и поисковую выдачу;
- чувствительность к наличию или отсутствию того или иного специализированного ответа;
- чувствительность к перестановке результатов;

Использовались признаки следующих видов:

- характеристики поискового запроса (лексические, тематические, частотные и т.д.);
- универсальные характеристики содержимого поисковой выдачи (не зависящие от типов представленных результатов):
 - характеристики расположения базовых элементов, используемых для представления результатов;
 - характеристики соответствия фрагментов текста поисковому запросу;
- признаки, характеризующие содержимое используемых специализированных результатов;
- агрегированные признаки веб-результатов:
 - признаки текстовой и ссылочной релевантности;
 - тематические признаки;
 - признаки, основанные на поведении пользователя.

Для обучения бинарного классификатора использовался метод градиентного бустинга [7, 22, 21] на “забывчивых” деревьях решений [11].

(TODO: подробнее: метод обучения, данные для обучения)

2.6 Алгоритм ранжирования

На основании построенной модели был разработан алгоритм для ранжирования разнородных результатов поиска. Основная идея алгоритма состоит в составлении разных вариантов расположения результатов и выборе наилучшего из них в соответствии с используемой моделью оценки полезности поисковой выдачи. Далее приводится общий

Алгоритм 1 Общий алгоритм ранжирования разнородных результатов поиска.

Вход: $W, S = \bigcup_{i=1}^K S_i, q$

Выход: R

```
1: function RANK( $W, S, q$ )
2:    $R \leftarrow W$ 
3:    $p \leftarrow \text{estimate\_ldt\_click\_probability}(R, q)$ 
4:   for  $R' \in \text{allowed\_orderings}(W, S)$  do
5:      $p' \leftarrow \text{estimate\_ldt\_click\_probability}(R', q)$ 
6:     if  $p' > p$  then
7:        $R \leftarrow R'$ 
8:      $p \leftarrow p'$ 
return  $R$ 
```

вариант алгоритма и его вариант с “жадной” стратегией расположения специализированных результатов.

2.6.1 Базовый алгоритм

На листинге 1 представлен псевдокод общего варианта алгоритма. Здесь функция $\text{estimate_ldt_click_probability}(R, q)$ возвращает оценку вероятности “длинного клика” при помощи описанной в предыдущем подразделе модели. Функция $\text{allowed_orderings}(W, S)$ выдает возможные варианты расположения результатов поиска (с учетом возможных ограничений). Эти варианты могут генерироваться в соответствии со стратегией полного перебора или какой-либо другой стратегией. От того, как именно генерируются варианты расположения результатов, зависит количество вызовов функции $\text{estimate_ldt_click_probability}(R, q)$ и, соответственно, производительность алгоритма.

2.6.2 “Жадный” вариант алгоритма

Из соображений производительности при реализации алгоритма было невозможно использовать стратегию полного перебора. (**TODO: пояснить?**) Поэтому использовался вариант алгоритма с жадной стратегией расположения специализированных результатов (листинг 2). Этот вариант алгоритма ранжирует результаты поиска следующим образом. Сначала в множество R добавляются все элементы из множества веб-результатов W (с учетом порядка), и для него оценивается вероятность “длинного клика”. Затем

Алгоритм 2 “Жадный” алгоритм ранжирования разнородных результатов поиска.

Вход: $W, S = \bigcup_{i=1}^K S_i, q$

Выход: R

```
1: function RANKGREEDY( $W, S, q$ )
2:    $R \leftarrow W$ 
3:    $p \leftarrow \text{estimate\_ldt\_click\_probability}(R, q)$ 
4:   while  $S \neq \emptyset$  do
5:      $\text{best\_}_s \leftarrow \text{nil}$ 
6:      $\text{best\_}_pos \leftarrow \text{nil}$ 
7:     for  $s \in S$  do
8:       for  $pos \in \text{allowed\_positions}(s, R)$  do
9:          $R' \leftarrow \text{insert}(R, s, pos)$ 
10:         $p' \leftarrow \text{estimate\_ldt\_click\_probability}(R', q)$ 
11:        if  $p' > p$  then
12:           $\text{best\_}_s \leftarrow s$ 
13:           $\text{best\_}_pos \leftarrow pos$ 
14:           $p \leftarrow p'$ 
15:        if  $\text{best\_}_s \neq \text{nil}$  then
16:           $R \leftarrow \text{insert}(R, \text{best\_}_s, \text{best\_}_pos)$ 
17:           $S \leftarrow \text{remove}(S, \text{best\_}_s)$ 
18:        else
19:           $S \leftarrow \emptyset$ 
return  $R$ 
```

из множества всех специализированных ответов $S = \bigcup_{i=1}^K S_i$ выбирается такой элемент best__s , и для него выбирается такая позиция best__pos , что добавление элемента best__s на позицию best__pos в список результатов поиска максимизирует вероятность совершения “длинного клика”. Если такой элемент нашелся, то он добавляется в R на найденную позицию и удаляется из множества S . Затем выбирается следующий элемент из оставшихся в S и т. д. Алгоритм заканчивает свою работу, когда найдены позиции для всех специализированных ответов, или же если добавление ни одного специализированного ответа ни на какую позицию не увеличивает вероятность “длинного клика”.

В таком варианте алгоритма число вызовов функции $\text{estimate_ldt_click_probability}(R, q)$ не превосходит $|S|^2(|W| + |S|)$. При $|W| = 10, |S| \approx 30$ время работы алгоритма не сказывалось на общем времени обработки поискового запроса.

Глава 3

Программная реализация

3.1 Схема системы ранжирования

3.2 Уменьшение числа обращений к поисковым источникам

Глава 4

Оценка качества работы метода

Качество работы метода оценивалось по метрикам, основанным на экспертных оценках и на поведении пользователей. В этой главе дается описание используемых метрик качества и приводятся результаты оценки.

(TODO: пояснить про метрики, основанным на экспертных оценках, и про поведенческие метрики?)

4.1 Методы оценки качества поиска

4.1.1 Методы, основанные на экспертных оценках

Для оценки качества работы метода по экспертным оценкам использовалась метрики точности и полноты показа специализированных ответов [17] и метрика *pfound* [20].

Метрика точности показа специализированных ответов – это среднее отношение числа уместных специализированных ответов, выдаваемых в ответ на запрос, к общему числу выдаваемых специализированных ответов:

$$P_{show} = \frac{1}{|Q|} \sum_{q \in Q} \frac{|S_{show}^{rel}(q)|}{|S_{show}(q)|}. \quad (4.1)$$

Здесь Q – выборка поисковых запросов, по которым вычисляется метрика, $S_{show}(q)$ – множество специализированных ответов, показываемых по запросу q , $S_{show}^{rel}(q)$ – множество уместных специализированных ответов, показываемых по запросу q ($S_{show}^{rel}(q) \subseteq S_{show}(q)$). Метрика полноты показа специализированных ответов – это сред-

нее отношение числа уместных специализированных ответов, выдаваемых в ответ на запрос, к общему числу уместных для запроса специализированных ответов:

$$R_{show} = \frac{1}{|Q|} \sum_{q \in Q} \frac{|S_{show}^{rel}(q)|}{|S^{rel}(q)|}. \quad (4.2)$$

Здесь также Q – выборка поисковых запросов, по которым вычисляется метрика, $S_{show}^{rel}(q)$ – множество уместных специализированных ответов, показываемых по запросу q . А $S^{rel}(q)$ – множество всех уместных специализированных ответов для запроса q ($S_{show}^{rel}(q) \subseteq S^{rel}(q)$). Уместность специализированных ответов определяется по экспертным оценкам.

Описанные метрики точности и полноты не учитывают позиции, на которых располагаются специализированные результаты. Для оценки качества работы метода с учетом позиций дополнительно использовалась метрика $pfound$ [20]. Эта метрика является оценкой вероятности того, что пользователь найдет интересующую его информацию в результатах поиска. Она вводится в рамках следующей модели:

- пользователь просматривает результаты поиска в списке по порядку;
- просматривая очередной результат поиска пользователь с некоторой вероятностью, зависящей от релевантности этого результата запросу, может найти интересующую его информацию, и завершить поиск;
- вероятность просмотра следующего результата зависит от релевантности предыдущего результата и некоторых внешних причин (например, накапливающейся при просмотре результатов усталости).

Значение метрики вычисляется следующим образом:

$$pfound(R, q) = \sum_{i=1}^n pLook(r_i) * pRel(r_i, q), \quad (4.3)$$

где $pLook(r_i)$ – вероятность просмотра i -го результата, $pRel(r_i, q)$ – вероятность того, что пользователь найдет исходную информацию при просмотре i -го результата. Значение вероятностного веса $pRel(r_i, q)$ определяется в соответствии с оценкой релевантности, выставленной паре (r_i, q) экспертами (насколько результат r_i соответствует запросу q).

Значение $pLook(r_i)$ вычисляется по следующей формуле:

$$pLook(r_i) = pLook(r_{i-1}) * (1 - pRel(r_{i-1}, q)) * (1 - pBreak), \quad 2 \leq i \leq n \quad (4.4)$$

$$pLook(r_1) = 1, \quad (4.5)$$

где $pBreak$ – вероятность того, что пользователь завершит поиск после просмотра ($i - 1$)-го результата по каким-то внешним причинам (например, из-за утомления). Значение $pBreak$, так же как и значения $pRel(r_i, q)$, является эвристикой.

4.1.2 Методы, основанные на поведении пользователей

(TODO:)

4.2 Описание результатов

Для оценки качества работы метода по точности и полноте показа специализированных результатов была составлена выборка из 20 тыс. запросов, и для нее были собраны экспертные оценки (набор уместных специализированных результатов для каждого запроса). (TODO: подробнее)

Качество по метрике $pfound$ оценивалось на выборке запросов, используемой для приемки изменений в алгоритмах ранжирования. В сравнении с текущим используемым методом предлагаемый метод показал повышение точности показа специализированных результатов на 21.22% при снижении полноты на 29.21% и прирост качества по метрике $pfound$ на 0.27%. (TODO: + результаты online-эксперимента)

(TODO: пояснить результаты – насколько это хорошо и т.д.)

Глава 5

Вопросы охраны труда

- 5.1 Общая характеристика санитарно-гигиенических условий труда**
- 5.2 Эргономические требования**
- 5.3 Микроклиматические условия**
- 5.4 Уровень шума**
- 5.5 Системы освещения**
- 5.6 Излучения**
- 5.7 Электробезопасность**
- 5.8 Инженерно-технические мероприятия по созданию благоприятных условий труда**
- 5.9 Методика и приборы контроля параметров среды**

Заключение

В данной работе предложен новый метод ранжирования разнородных результатов поиска. Его отличительные особенности состоят в следующем:

- ранжируемые результаты рассматриваются в совокупности, а не по отдельности;
- результаты располагаются в соответствии с критерием полезности, основанным на действиях пользователей на поисковой выдаче.

Благодаря этим особенностям метод обладает рядом преимуществ. Во-первых, он универсален: он может быть применен для ранжирования результатов произвольного вида и для разных моделей поисковой выдачи. Во-вторых, он позволяет естественным образом учитывать взаимосвязи между результатами. И в-третьих, он не требует экспертных оценок для обучения.

Предложенный метод был реализован и применен для встраивания 32-х видов специализированных результатов в поисковую выдачу системы Яндекс для мобильных устройств. Встраивались результаты поиска по изображениям, видео, мобильным приложениям, товарам, новостям, результаты гео-поиска и других сервисов компании Яндекс.

Было оценено качество работы метода по поисковым метрикам, основанным на экспертных оценках и на поведении пользователей. В сравнении с текущим используемым методом было получено улучшение точности показа специализированных результатов на 21.22% при снижении полноты на 29.21% и прирост качества по метрике *pfound* на 0.27%. (TODO: уточнить результаты) (TODO: + online-метрики)

В процессе реализации метода также была решена задача нахождения заданного числа кандидатов в аргументы максимизации значения функции, представляющей собой ансамбль “забывчивых” деревьев решений (oblivious decision trees), по частично

вычисленному вектору признаков. Решение этой задачи позволяет избежать обращения к тем поисковым источникам, результаты которых заведомо нерелевантны заданному поисковому запросу. Разработанное решение имеет самостоятельную ценность и может быть применено и в других задачах.

Список литературы

- [1] J. Arguello, F. Diaz, J. Callan, and J.-F. Crespo. Sources of evidence for vertical selection. // SIGIR 2009, pp. 315–322. ACM, 2009.
- [2] J. Arguello, F. Diaz, and J.-F. Paiement. Vertical selection in the presence of unlabeled verticals. // SIGIR 2010, pp. 691–698. ACM, 2010.
- [3] J. Arguello, F. Diaz, and J. Callan. Learning to aggregate vertical results into web search results. // CIKM 2011, pp. 201–210. ACM, 2011.
- [4] F. Diaz. Integration of news content into web results. // WSDM 2009, pp. 182–191. ACM, 2009.
- [5] F. Diaz and J. Arguello. Adaptation of offline vertical selection predictions in the presence of user feedback. // SIGIR 2009, pp. 323–330. ACM, 2009.
- [6] S. Fox, K. Karnawat, M. Mydland, S. Dumais, T. White. Evaluating Implicit Measures to Improve Web Search. // ACM TOIS, 23(2), pp. 147-168, 2005.
- [7] T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning, 2nd edition. // Springer, 2009.
- [8] L. Jie, S. Lamkhede, R.t Sapra, E. Hsu, H. Song, Y. Chang. A Unified Search Federation System Based on Online User Feedback. // Proceedings of KDD 2013. ACM, 2013.
- [9] T. Joachims. Optimizing Search Engines using Clickthrough Data. // Proceedings of KDD 2002. ACM, 2002.
- [10] Y. Kim, A. Hassan, R. White, I. Zitouni. Modeling Dwell Time to Predict Click-level Satisfaction. // Proceedings of WSDM 2014.

- [11] R. Kohavi. Ch. Li. Oblivious Decision Trees, Graphs, and Top-Down Pruning. // IJCAI, pp. 1071-1079, 1995.
- [12] A. C. König, M. Gamon, and Q. Wu. Click-through prediction for news queries. // SIGIR 2009, pp. 347–354. ACM, 2009.
- [13] D. Lefortier, P. Serdyukov, F. Romanenko, M. de Rijke. Blending Vertical and Web results: A Case Study using Video Intent. // ECIR 2014, pp. 184-196.
- [14] X. Li, Y.-Y. Wang, and A. Acero. Learning query intent from regularized click graphs. // SIGIR 2008, pp. 339–346. ACM, 2008.
- [15] T. Liu. Learning to rank for information retrieval. // Foundations and Trends in Information Retrieval, vol. 3, no. 3, pp. 225–331, 2009.
- [16] A. K. Ponnuswami, K. Pattabiraman, Q. Wu, R. Gilad-Bachrach, and T. Kanungo. On composition of a federated web search result page: Using online users to provide pairwise preference for heterogeneous verticals. // WSDM 2011, pp. 715–724. ACM, 2011.
- [17] M. Sokolova, G. Lapalme. A systematic analysis of performance measures for classification tasks. // Information Processing and Management 45, p. 427–437. Elsevier, 2009.
- [18] R. White, J. Huang. Assessing the Scenic Route: Measuring the Value of Search Trails in Web Logs. // Proceedings of SIGIR 2010. ACM 2010.
- [19] И. Ашманов, А. Иванов. Оптимизация и продвижение сайтов в поисковых системах, 3-е издание. // Спб.: Питер, 2011.
- [20] А. Гулин, П. Карпович, Д. Расковалов, И. Сегалович. Оптимизация алгоритмов ранжирования методами машинного обучения. // РОМИП 2009.
- [21] А. Фонарев, А. Дьяконов. Обзор алгоритмов бустинга. // МГУ, 2012.
- [22] Алгоритм машинного обучения MatrixNet. // URL: <https://company.yandex.ru/technologies/matrixnet/>. Дата обращения: 3.06.2015.

- [23] International Data Corporation (IDC). The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. // EMC website, URL: <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm> (дата обращения: 7.05.2015).
- [24] A Brief History of Search Engines. // Webreference website, URL: http://www.webreference.com/authoring/search_history (дата обращения: 19.05.2015).

Словарь терминов

(TODO: дать определение терминам)

Поисковая система?

Поисковый запрос

Поисковая выдача

Поисковый источник

Специализированный ответ (специализированный результат)

Ранжирование

Тип информации