**Московский авиационный институт**

**(национальный исследовательский университет)**

Факультет прикладной математики и физики

Кафедра вычислительной математики и программирования

РЕФЕРАТ

по дисциплине «Информационный поиск»

«Методы оценки качества поиска, их достоинства и недостатки»

|  |  |
| --- | --- |
|  | Выполнил  студент группы М8О-108М-18  Рожлейс И.А.  Проверил  ст. преп. каф. 806  Калинин А.Л. |

Москва

2019

**СОДЕРЖАНИЕ**

|  |  |  |
| --- | --- | --- |
| 1. | Введение | 3 |
| 2. | Глава 1. Метрики эффективности поиска | 3 |
| 3. | Глава 2. Метрики качества ранжирования | 6 |
| 4. | Заключение | 8 |
| 5. | Источники | 8 |

**Введение**

Информационный поиск — это процесс поиска некоторого набора документов, который удовлетворял бы информационной потребности пользователя. Чтобы установить насколько найденный набор документов соответствует потребностям пользователя вводят различные метрики оценки качества поиска.

По типу входных данных для расчёта метрики можно разделить на внешние и внутренние. Внешние метрики полагаются на наличие некоторого предопределённого значения с которым их можно было бы сравнить. Так, например, метрики оценки эффективности поиска полагаются на заранее предопределённую оценку документа, называемую его релевантностью. Оценку релевантности даёт человек, называемый асессором (в будущем вполне возможно специально обученная нейронная сеть). Примером внешней метрики может служить точность поисковой выдачи (**глава 1**) или качество ранжирования (**глава 2**).

Внутренние метрики самодостаточны и не требуют сравнения с некими внешними показаниями. По предмету оценки метрики можно разделить на метрики оценки качества документов, метрики ранжирования, поведенческие метрики и т.п.

**Глава 1. Метрики эффективности поиска.**

Один из способов оценить качество поисковой системы — это оценить насколько хорошо документы из поисковой выдачи соответствуют запросу. Все документы корпуса поделены асессором на релевантные и нерелевантные относительно определённого запроса. Поисковая система в свою очередь на каждый определённый запрос возвращает некоторый набор документов, который может содержать ошибки. Таким образом естественно возникают четыре класса документов и каждый документ может быть причислен к одному из этих классов как изображено на **табл. 1**.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Асессор** | |
| **Положительно** | **Отрицательно** |
| **Поисковая система** | **Положительно** | TP | FP |
| **Отрицательно** | FN | TN |

**Табл. 1.** Вероятные классы документов.

Если одновременно асессор и поисковая система «сошлись» в оценке документа, то такой документ является либо истинно-релевантным (true-positive TP), либо истинно-нерелевантным (true-negative TN). Однако поисковая система также может допустить ошибку классифицируя документ как релевантный в то время как асессор классифицировал его как нерелевантный (false-positive FP) и наоборот (false-negative FN).

Исходя из выше перечисленных классов документов можно ввести несколько метрик качества поиска. Так, например, метрика аккуратность показывает процент верных решений принятых поисковой программой и есть отношение количества верных принятых решений к общему количеству документов корпуса:

Метрика точность показывает процент найденных релевантных документов в поисковой выдаче:

Метрика полноты описывает процент релевантных документов к общему количеству релевантных документов в корпусе:

Выпадение описывает процент нерелевантных документов к общему количеству нерелевантных документов в корпусе. По смыслу это противоположная полноте величина:

Бывают случаи, когда возможных классов больше двух (не только классы «релевантный» и «нерелевантный»), тогда применяют матрицы несоответствия (confusion matrix) и метрики можно представить в общем случае. Столбцы и строки матрицы определяют конкретный класс или номер класса, а элементы матрицы количество попаданий.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Асессор** | | | | | | |
| **1** | **2** | **3** | **…** | **n-2** | **n-1** | **n** |
| **Поисковая система** | **1** | 91 |  |  |  | 2 |  | 6 |
| **2** |  | 93 |  |  |  |  |  |
| **3** |  |  | 97 |  |  | 1 | 3 |
| **…** |  | 2 |  |  |  |  |  |
| **n-2** |  |  |  |  | 98 |  |  |
| **n-1** |  | 5 |  |  |  | 94 |  |
| **n** | 1 |  |  |  |  |  | 90 |

**Табл. 2.** Матрица несоответствий.

Если диагональные элементы матрицы несоответствий явно выражены как на **табл. 2**, то в таком случае поисковая система классифицирует большинство документов правильно, однако отклонение от диагонали рассматривается как ошибка классификации.

С учётом вышесказанного метрики точности и полноты в общем случае могут быть представлены как:

Чем выше точность и полнота, тем лучше. Однако в реальности невозможно достигнуть максимума одновременно для точности и полноты. Поэтому применяется метрика Ван Ризбергена или иначе F-мера, которая объединяет эти две метрики в одну величину:

Частный случай F-меры, когда β равно 1 называется сбалансированной F-мерой и превращается в выражение вида:

**Глава 2. Метрики качества ранжирования.**

Ранжирование – это задача сортировки документов поисковой выдачи по убыванию некоторого значения называемого рангом документа. Чем больше ранг, тем ближе документ находится к началу списка. Соответственно метрика качества ранжирования основываясь на тех или иных предположениях позволяет оценить насколько хорошо был отсортирован список поисковой выдачи.

Существует два основных подхода оценки качества ранжирования. Классический офлайн метод по-прежнему полагается на асессоров. Онлайн метод полагается на реальных пользователей поисковой системы и применяет кликовую модель для оценки качества ранжирования.

В классическом подходе есть асессор, который выбирает первые N документов из поисковой выдачи и по порядку присваивает каждому документу некую оценку, свидетельствующую о степени релевантности данного документа к теме исходного запроса.

Первая метрика CG (cumulative gain), представляет собой просто сумму рангов первых N документов поисковой выдачи:

где r – ранг документа.

Эта метрика не учитывает перестановку документов в поисковом запросе и имеет одинаковое значение даже если более релевантный документ находится в конце списка поискового запроса. Для того чтобы учесть положение документа в списке поискового запроса применяется метрика DCG (discontinued cumulative gain):

теперь релевантность документа в начале списка даёт больший вклад в значение метрики, чем релевантность того же документа в конце списка.

Метрика NDCG (normalized discontinued cumulative gain) нормализует метрику DCG таким образом, чтобы разные значения DCG можно было сравнивать между собой, тем самым можно сравнить между собой различные поисковые системы:

где IDCG есть некоторое идеальное значение DCG, которое вычисляется, как если бы релевантные документы находились в начале списка, а нерелевантный в конце. Например, первые пять документов поисковой выдачи асессор оценил по пятибалльной шкале как 4 5 4 2 3. В идеале поисковая система должна была отсортировать документы в порядке 5 4 4 3 2. Так что IDCG это тот же DCG только от идеальной последовательности. Таким образом IDCG всегда больше либо равен значению DCG.

Метрика ERR (expected reciprocal rank) учитывает вероятность того, насколько найденный документ релевантный и соответствует потребностям пользователя:

где N количество документов в списке, а R есть преобразование оценки релевантности документа в вероятность релевантности и задаётся формулой:

Кликовая модель данных в свою очередь предполагает сбор данных о взаимодействии пользователя с поисковой системой. Такими данными могут являться: запрос, время отклика, траектория курсора, клики и переходы. Основываясь на этих данных, поисковая система может учиться автоматически ранжировать документы под потребности пользователя. Кликовая модель – это вероятностная модель поведения пользователя при работе с поисковой системой.

Одной из распространённых кликовых моделей является модель динамической сети Байеса (dynamic Bayesian network DBN). Кликовая модель включает в себя несколько вероятностных событий:

* *Ek* (examining) пользователь изучает k-ый сниппет документа;
* *Ak* (attracted) пользователь привлечёт k-ым документом;
* *Ck* (clicking) пользователь кликает на документ;
* *Sk* (satisfied) пользователь удовлетворён результатом.

Теперь можно ввести метрики. Так, например, можно ввести понятие ожидаемой полезности:

где P это вероятность клика i-го документа, а R его релевантность.

Следующая метрика показывает вероятность удовлетворённости пользователя поисковой выдаче:

**Заключение**

В первой части были рассмотрены метрики эффективности поиска, описывающие аккуратность, точность и полноту поисковой выдачи. Были рассмотрены классы документов и представление некоторых метрик в общем виде.

Во второй части были рассмотрены различные метрики качества ранжированного поиска. Также была рассмотрена простая кликовая модель. При помощи кликовой модели можно обучить поисковую систему ранжированию.

**Источники**

[1] Баженов Д. Оценка классификатора (точность, полнота, F-мера) [Электронный ресурс] // URL: [http://‌bazhenov.me/‌blog/‌2012/‌07/‌21/‌classification-‌performance-‌evaluation.html](http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html)

[2] Информационный поиск [Электронный ресурс] // URL: [https://‌ru.wikipedia.org/‌wiki/‌Информационный\_поиск](https://ru.wikipedia.org/wiki/Информационный_поиск)

[3] Блог компании E-Contenta. Метрики качества ранжирования [Электронный ресурс] // URL: [https://‌habr.com/‌ru/‌company/‌econtenta/‌blog/‌303458/](https://habr.com/ru/company/econtenta/blog/303458/)

[4] Jacob Joseph. The Best Metric to Measure Accuracy of Classification Models // URL: <https://‌clevertap.com/‌blog/‌the-‌best-‌metric-‌to-‌measure-‌accuracy-‌of-‌classification-‌models/>

[5] Чуклин А., Сердюков П., Рийке М. Метрики информационного поиска на основе кликовых моделей [Электронный ресурс] // URL: [https://‌wseob.ru/‌seo/‌ir-click-metrics/](https://wseob.ru/seo/ir-click-metrics/)