

Оценка качества моделей информационного поиска

Андрей Лебедев, группа 424

Постановка задачи

Целью данной работы является оценка качества ранжирования следующих моделей:

- TF;
- TF-IDF;
- Языковая модель с параметрами $\lambda = 0.5$ и $\lambda = 0.9$;
- FastText;
- FastText с усреднением по IDF.

Качество измеряется с использованием метрики NDCG@10.

Идеальная разметка релевантности

Для оценки качества выдачи каждой модели была составлена идеальная разметка релевантности предложений по шкале:

- 2 — предложение содержит полный факт;
- 1 — предложение содержит часть факта;
- 0 — предложение нерелевантно запросу.

Идеальная выдача для запросов:

Запрос 1: На московском острове растёт восьмиметровая облепиха.

- Ранг 1: «Широко распространена облепиха (отдельные её экземпляры достигают высоты 7—8 метров), часто встречается рейнхутрия сахалинская.» - Оценка: 1.
- Ранг 1: «На острове широко распространена облепиха.» — Оценка: 1.
- Ранг 2: «Облепиха.» — Оценка: 1.
- Ранг 3 и далее: Остальные предложения явно нерелевантны запросу. Оценка для всех: 0.

Ранжирование: [1, 1, 1, 0, 0, 0, 0, 0, 0, 0].

Запрос 2: Гватемальско-испанские отношения были разорваны после того, как посольство Испании было сожжено.

- Ранг 1: «2 февраля 1980 года Испания разорвала дипломатические отношения с Гватемалой из-за инцидента в посольстве и угроз в адрес дипломатического персонала.» — Оценка: 2.
- Ранг 2: «В результате случившегося Испания на 4 года разорвала дипломатические отношения с Гватемалой.» — Оценка: 1.
- Ранг 3: «Как только испанский посол покинул воздушное пространство страны пребывания, 2 февраля 1980 года Испания разорвала дипломатические отношения с Гватемалой из-за инцидента в посольстве и угроз в адрес дипломатов.» — Оценка: 1.
- Остальные предложения явно нерелевантны запросу. Оценка для всех: 0.

Ранжирование: [2, 1, 1, 0, 0, 0, 0, 0, 0, 0].

Запрос 3: Булгаков не смог приехать на похороны матери из-за отсутствия денег.

- Ранг 1: «Михаил Булгаков не смог приехать на похороны из Москвы в Киев из-за отсутствия денег.» — Оценка: 2.

- Ранг 2: «Если ему хотелось проехать на такси на последние деньги, он без раздумья решался на этот шаг: «Мать ругала за легкомыслие.» — Оценка: 1.
- Остальные предложения: Оценка: 0.

Ранжирование: [2, 1, 0, 0, 0, 0, 0, 0, 0, 0].

Оценки релевантности моделей по запросам

Запрос 1

Модель	Оценки
TF	[2, 1, 0, 0, 0, 0, 0, 0, 0, 0]
TF-IDF	[1, 2, 0, 0, 0, 0, 0, 0, 0, 0]
LM (=0.5)	[0, 0, 2, 1, 0, 0, 0, 0, 0, 0]
LM (=0.9)	[0, 0, 2, 1, 0, 0, 0, 0, 0, 0]
FastText	[1, 2, 0, 0, 0, 0, 0, 0, 0, 0]
FastText+IDF	[2, 1, 0, 0, 0, 0, 0, 0, 0, 0]

Таблица 1: Оценки релевантности для запроса 1

Запрос 2

Модель	Оценки
TF	[0, 0, 2, 2, 1, 0, 0, 0, 0, 0]
TF-IDF	[0, 2, 2, 1, 0, 0, 0, 0, 0, 0]
LM (=0.5)	[0, 2, 1, 2, 0, 0, 0, 0, 0, 0]
LM (=0.9)	[2, 2, 1, 0, 0, 0, 0, 0, 0, 0]
FastText	[1, 0, 2, 2, 0, 0, 0, 0, 0, 0]
FastText+IDF	[2, 1, 2, 0, 0, 0, 0, 0, 0, 0]

Таблица 2: Оценки релевантности для запроса 2

Модель	Оценки
TF	[2, 0, 0, 0, 0, 0, 0, 0, 0, 0]
TF-IDF	[2, 0, 1, 0, 0, 0, 0, 0, 0, 0]
LM ($\lambda=0.5$)	[2, 1, 0, 0, 0, 0, 0, 0, 0, 0]
LM ($\lambda=0.9$)	[2, 0, 1, 0, 0, 0, 0, 0, 0, 0]
FastText	[2, 0, 1, 0, 0, 0, 0, 0, 0, 0]
FastText+IDF	[2, 0, 1, 0, 0, 0, 0, 0, 0, 0]

Таблица 3: Оценки релевантности для запроса 3

Запрос 3

Оценки релевантности моделей

Запрос	TF	TF-IDF	LM (0.5)	LM (0.9)	FastText	FastText+IDF
Запрос 1	0.7654	0.7654	0.2893	0.2346	0.4693	0.5922
Запрос 2	0.7181	0.8600	0.8378	1.0000	0.9139	1.0000
Запрос 3	0.7602	0.7502	1.0000	0.9502	0.9502	0.9502
Среднее	0.7479	0.7985	0.7090	0.7283	0.7778	0.8475

Таблица 4: Результаты оценки NDCG@10 для различных моделей

Выводы

На основе полученных данных можно сделать следующие выводы:

- Модель TF продемонстрировала более слабые результаты, так как она не учитывает важность редких терминов.
- TF-IDF улучшает ранжирование благодаря учету инверсной частоты документа, показав хорошие результаты на всех запросах.
- Языковая модель (LM) показала неоднозначные результаты: с $\lambda = 0.9$ она работает лучше, чем с $\lambda = 0.5$, но уступает FastText.
- FastText с учетом IDF продемонстрировал лучшую производительность, достигнув максимального качества на всех запросах.

Таким образом, модель FastText+IDF является наиболее эффективной для обработки текущих запросов, так как она достигает максимальных значений NDCG@10.