

## **МЕТОДЫ ОЦЕНКИ ЭФФЕКТИВНОСТИ РЕШЕНИЯ ЗАДАЧ РАНЖИРОВАНИЯ**

Задачи ранжирования, то есть упорядочивания некоторого перечня данных по заданным критериям, являются одними из самых распространенных задач в современном обществе. С точки зрения систематики экономических задач, ранжирование можно рассматривать как разновидность задач классификации, где объекты должны располагаться в порядке уменьшения или увеличения рангового признака, который может как явно присутствовать в описании объекта, так и являться внешней по отношению к объекту характеристикой. В простейшем виде, при единственном и явном критерии, задачи ранжирования встречались и успешно решались еще в античности.

С развитием экономических отношений и появлением компьютерных информационных систем, позволяющих обрабатывать большие массивы данных, стали усложняться и задачи ранжирования. В настоящее время актуальные задачи ранжирования могут являться многофакторными и многокритериальными, а также включать показатели различной природы и значимости.

Сложная природа современных задач ранжирования, в которых ранговый признак является внешней характеристикой, не имеющей явной связи с известными параметрами ранжируемых объектов, в ряде случаев делает невозможным существование единственного достоверного метода их решения.

Анализ отечественной и зарубежной научной литературы показал, что само понятие эффективности ранжирования в настоящее время рассматривается лишь применительно к задачам обработки поисковых запросов. Цель ранжирующей модели в этом случае заключается в том, чтобы наилучшим образом обобщить способ ранжирования в обучающей выборке на новые данные [1]. Эффективность ранжирования определяется способностью модели расположить в верхних строках результатов поиска ссылки на документы, наилучшим образом соответствующие запросу пользователя.

Несмотря на мощное аппаратное и программное обеспечение современных поисковых систем, процедуру оценки ранжирующих алгоритмов до сих пор не удается сделать полностью автоматизированной. Так, еще в 2008 году представители компании Google утверждали, что их поисковая система еще не готова окончательно доверить ранжирование алгоритмам машинного обучения, поскольку автоматически созданные модели могут работать непредсказуемо на новых классах запросов, не похожих на запросы из обучающей выборки, в отличие от моделей, созданных людьми-экспертами [2].

В остальных направлениях экономики, где решаются задачи ранжирования, анализ литературных источников и материалов Internet показал, что *задача оценки качества ранжирования* в общем виде до сих пор даже не ставилась, что обуславливает научную новизну исследования.

Актуальность исследования методов оценки качества ранжирования обусловлена широким кругом экономических задач, эффективность решения которых в отсутствие такого метода является недостаточной. Вот только некоторые примеры задач ранжирования:

- ранжирование результатов обработки поисковых запросов по релевантности [1];
- ранжирование клиентов по вероятности отклика, в том числе формирование списка рассылки рекламы [3], формирование списка должников для работ по взысканию долга, формирование кредитных скоринговых карт [4];
- ранжирование мест для открытия филиалов и представительств торговых и финансовых учреждений по критерию наибольшей посещаемости клиентами целевых групп;
- ранжирование методов обучения нейросетевых моделей по критерию наибольшей эффективности работы обученной модели;
- ранжирование методов технического анализа валютных и фондовых рынков по критерию эффективности для построения интегральных оценок;
- ранжирование переменных в порядке их значимости для использования в интегральных моделях (в частности — нейросетевых).

Технология решения таких задач в общем случае должна включать построение нескольких моделей распределения ранжируемых показателей с последующим выбором той из них, которая показала лучшие результаты на тестовом наборе данных, то есть данных для которых известен их истинный ранг (рис. 1).

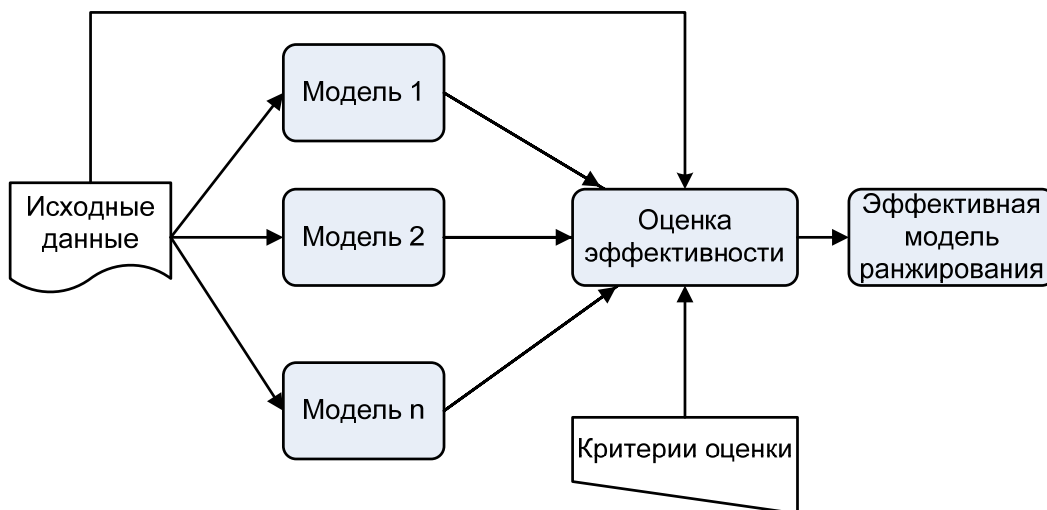


Рис. 1. Схема решения задач ранжирования со сложной природой критериев

Не останавливаясь на инструментальных средствах построения моделей ранжирования, которые могут иметь различную природу, в том числе описательную, рассмотрим понятие эффективности ранжирования и методы, которые могут быть использованы для её оценки.

Введем следующие обозначения:

- $A_i$  — вектор параметров, характеризующих  $i$ -й объект, подлежащий ранжированию. При этом набор параметров одинаков для всех таких объектов;
- $\{A\} = \{A_1, A_2, \dots, A_n\}$  — множество входных данных;
- $rp_i$  — ранговый признак  $i$ -го объекта — величина, определенная при помощи модели ранжирования на основании  $A_i$  и обуславливающая позицию объекта в ранжированном списке;
- $RP = \{rp_1, rp_2, \dots, rp_n\}$  — вектор ранговых признаков, определенных моделью:

$$RP = M_r(A), \quad (1)$$

где  $M_r$  — модель ранжирования;  $trp_i$  — истинный ранговый признак — внешняя характеристика, которая в общем случае не может быть найдена из  $A_i$ , но определяется *post factum* на основании дополнительных данных;  $TRP = \{trp_1, trp_2, \dots, trp_n\}$  — вектор истинных ранговых признаков;  $r_i$  — ранг (место в рейтинге)  $i$ -го объекта, определенный на основании  $rp_i$ .  $r_i \in \mathbb{N}$ ;  $R = \{r_1, r_2, \dots, r_n\}$  — вектор рангов, определенных моделью:

$$R = rang(RP); \quad (2)$$

$tr_i$  — истинный ранг  $i$ -го объекта, определенное на основании  $trp_i$ .  $tr_i \in \mathbb{N}$ ;  $TR = \{tr_1, tr_2, \dots, tr_n\}$  — вектор истинных рангов:

$$TR = rang(TRP). \quad (3)$$

Таким образом, определение эффективности ранжирования может быть сделано исходя из вектора расстояний между истинными рангами объектов и рангами, определенными моделью  $\Delta(R, RT) = \{d_1, d_2, \dots, d_n\}$ .

Погрешность ранжирования модели  $M_r$ , то есть величина, обратно пропорциональная ее эффективности, в этом случае можно определить так:

$$eMr = \sum_{i=1}^n |d_i|. \quad (4)$$

Для определения расстояния между истинными рангами и рангами, определенными моделью ( $d_i$ ), в простейшем случае можно воспользоваться выражением:

$$d_i = r_i - tr_i. \quad (5)$$

Однако на практике при вычислении  $d_i$  и  $eMr$  необходимо учитывать некоторые факторы, определяемые экономической сущностью решаемой задачи:

1. Задача может допускать существование объектов с одинаковыми рангами, то есть  $tr_i = tr_j$  для некоторых  $i, j = 1..n$ . Это означает, что объект может занять любое из некоторого количества мест и это не повлияет на эффективность решения. Вычисление  $d_i$  в такой задаче с помощью формулы (5) приведет к ошибочному завышению погрешности ранжирования, следовательно, при разработке алгоритма вычисления  $d_i$  необходимо обеспечить учет данного фактора.

2. Значимость правильного ранжирования объектов в начале и в конце списка может быть различной. Так, во многих задачах (поиска информации, ранжирования объектов в порядке убывания предпочтений и тому подобные) наибольшую важность имеет правильное расположение объектов в верхней части списка, тогда как за ее пределами допускаются достаточно сильные отклонения найденного рейтинга от истинного. Для учета фактора значимости, выражение (4) можно дополнить:

$$eMr = \sum_{i=1}^n f(i) |d_i|, \quad (6)$$

где  $f(i)$  — функция значимости  $i$ -го места в рейтинге.

Кроме чисто расчетных для анализа эффективности ранжирования могут быть применены и расчетно-графические методы, в частности, Lift-кривые и их разновидности (Profit-кривые, Gain-диаграммы и тому подобные).

Lift-кривая формируется на основе лифт-фактора, который был впервые определен при решении задачи оптимизации массовой рассылки как показатель, отражающий увеличение числа откликов относительно числа действий (почтовых отправок). Lift-кривая строится следующим образом [3]: по горизонтальной оси откладывается размер выборки упорядоченной по убыванию показателя  $rp_i$ , который отражает вероятность наступления положительного исхода, согласно анализируемой модели. По вертикальной оси фиксируется кумулятивное число положительных исходов в каждой подвыборке

(лифт). Поскольку истинный ранговый признак  $trp_i$  в рассматриваемом случае является бинарной величиной, которая принимает значение 1 в случае положительного исхода и значение 0 в случае отрицательного, выражение для расчета лифта можно записать следующим образом:

$$lft(i) = \sum_{x=1}^i trp_x . \quad (7)$$

Классическая Lift-кривая может быть использована для оценки эффективности моделей ранжирования только при бинарной природе истинных ранговых признаков  $trp_i$ .

В качестве примера использования Lift-кривой рассмотрим задачу ранжирования просроченных кредитных дел, в которых объекты необходимо расположить по степени убывания вероятности возобновления заемщиком платежа по кредиту. Для анализа значимости используется выборка тестовых данных, в которых исход по каждому кредитному делу, соответствующий его истинному рангу, уже известен. Так, анализируемая выборка содержит данные по 500 заемщикам, из которых в результате проведенных банком мероприятий 83 человека возобновили внесение кредитных платежей. Вектор входных данных имеет следующую структуру: {пол заемщика; возраст; сумма просрочки, сумма платежа; отношение просрочка/платеж; период просрочки; сумма кредита}.

Первая модель является эмпирической и используется во многих кредитных подразделениях для ранжирования данных о заемщиках. В качестве рангового признака  $rp_i$  в ней выбирается количество дней, прошедшее с момента последнего платежа. Действительность модели подтверждают статистические исследования, в соответствии с которыми действительно, чем больше просрочка, тем меньше вероятность возобновления оплаты.

Вторая модель основана на определении связи между входными и результирующим показателями при помощи логистической регрессии. Эта модель позволяет обеспечить учет большего количества характеристик объекта, а также определить их значимость и влияние на вероятность возобновления платежа.

Третья модель всегда присутствует на диаграмме с Lift-кривыми и соответствует «бесполезной модели» случайного выбора. Она служит визуальным ориентиром для анализа и сопоставления эффективности остальных моделей.

Графики Lift-кривых перечисленных моделей показаны на рис. 2.

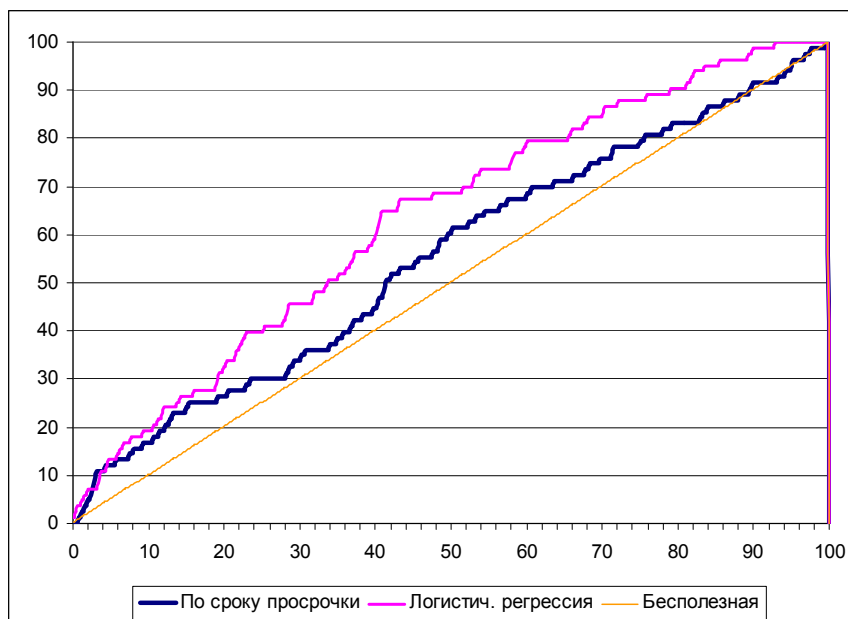


Рис. 2. Lift-кривые различных моделей ранжирования заемщиков в коллекторском скоринге

Практический смысл сопоставления эффективности при помощи Lift-кривых состоит в определении модели, позволяющей совершить наименьшее количество действий, необходимых для достижения определенного результата.

В качестве общего критерия оценки модели используется площадь под кривой, выраженная в процентах. Для бесполезной модели этот показатель всегда равен 50 %. Для модели логистической регрессии, показанной на рис. 2, площадь под кривой равна 63,3 %. Для приведенной там же эмпирической модели ранжирования — 55,1 %. Очевидно, что по общей эффективности наилучшей оказалась модель логистической регрессии.

Другим критерием оценки является доля откликов, получаемых при совершении определенного количества действий. Так, чтобы получить 50 % положительных откликов в «бесполезной» модели необходимо обработать 50 % кредитных дел, с использованием модели логистической регрессии — 31 % кредитных дел, а с использованием модели оценки вероятности погашения по сроку просрочки — 41 % дел. Очевидно, что в данном случае также целесообразно выбрать именно модель логистической регрессии. При этом следует отметить, что эмпирическая модель демонстрирует хорошие результаты при обработке малого количества заявок. Возможность исследования не только общей эффективности моделей, но и эффективности на отдельных участках диапазона ранжирования является безусловным преимуществом расчетно-графических методов оценки.

С позиций данного исследования недостатком классических Lift-кривых является невозможность их использования для анализа эффективности моделей ранжирования в том случае, если элементы множества истинных ранговых признаков  $trp_i$  имеют не бинарную природу. Однако принципы, положенные в основу этого метода, могут быть расширены для анализа эффективности при произвольной природе истинных ранговых признаков.

Выражение для расчета лифта при этом можно представить так:

$$lft(i) = \sum_{x=1}^i \{r \leq tr_x\}, \quad (8)$$

где  $\{r \leq tr_x\} = 1$  если истинно и 0, если ложно.

Метод построения кривой эффективности ранжирования, основанный на функции (8), позволяет обеспечить приоритетную значимость правильного ранжирования объектов в начале списка, что, как уже отмечалось, соответствует условиям большого количества экономических задач. Действительно, если за основной критерий эффективности модели ранжирования принять площадь под кривой (в долях от общей площади графика):

$$S_{lft} = \frac{\sum_{i=1}^n lft(i)}{n^2}, \quad (9)$$

то правильное расположение объекта на первом месте в ранге увеличит общую площадь под кривой на величину  $S_{lft} = \frac{n}{n^2}$ , тогда как правильное расположение объекта на последнем месте в ранге только величину  $S_{lft} = \frac{1}{n^2}$ . Таким образом, значимость мест в ранге в предложенном методе оценки эффективности (8) убывает в арифметической прогрессии.

На рис. 3 изображены два крайних случая кривых эффективности ранжирования.

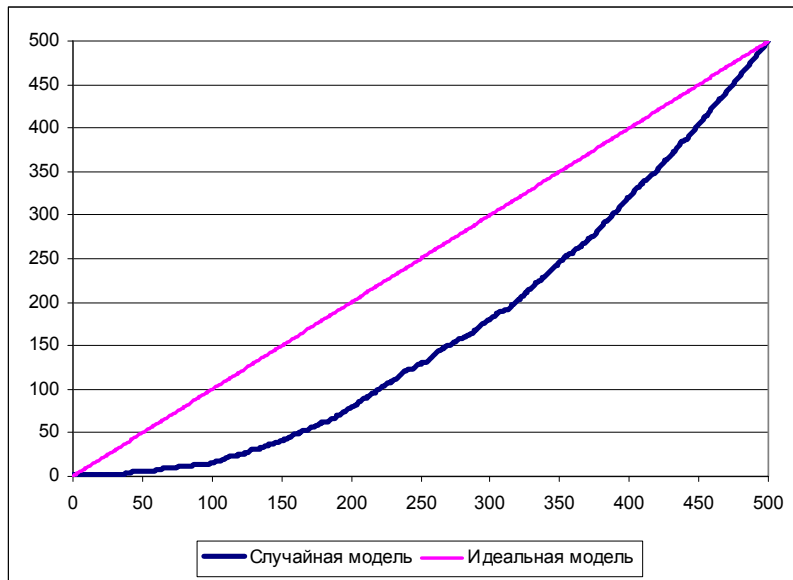


Рис. 3. Кривые эффективности ранжирования идеальной и бесполезной модели

Диагональная прямая линия ( $S_{ff} = 0.5$ ) на рис. 3 соответствует идеальной модели ранжирования, которая расположила все объекты на правильных местах. Вогнутая линия под диагональю ( $S_{ff} = 0.33$ ) соответствует бесполезной модели, в которой выбор мест в ранге осуществляется случайно. Сравнительно небольшой зазор между линиями идеальной и бесполезной моделей несколько ухудшает возможности визуального анализа, поэтому среди возможных направлений развития предлагаемого метода прежде всего нужно отметить совершенствование его графической интерпретации.

Визуальный анализ рис. 3 показывает, что и внешний вид, и интерпретация построенных кривых эффективности существенно отличаются от рассмотренных выше Lift-кривых, что дает основание считать их новым методом анализа эффективности решения задач ранжирования.

**Выводы:** В статье рассмотрена актуальная проблема оценки эффективности решения задач ранжирования. Предложены расчетный и расчетно-графические методы решения этой проблемы. Преимуществом расчетно-графических методов является большая наглядность получаемых результатов. Недостатком — несколько худшие возможности формализации выбора оптимальных моделей.

Среди основных научных результатов работы следует отметить метод построения кривых эффективности ранжирования, который позволяет проводить анализ эффективности при произвольной природе истинных ранговых признаков. Возможными направлениями развития предложенного метода является совершенствование его графической интерпретации.

### Литература

1. Обучение ранжированию: [Электронный ресурс]. — Режим доступа : [http://491.ru/a/obuchenie\\_ranjirovaniyu](http://491.ru/a/obuchenie_ranjirovaniyu).
2. Are Machine-Learned Models Prone to Catastrophic Errors?: [Электронный ресурс]. — Режим доступа : <http://anand.typepad.com/datawocky/2008/05/are-human-experts-less-prone-to-catastrophic-errors-than-machine-learned-models.html>.
3. Паклин Н. Б. Бизнес-аналитика. От данных к знаниям. 2-е изд. / Н. Б. Паклин, В. И. Орешков. — СПб. : Питер, 2010. — 704 с.
4. Уланов С. В. Оценка качества и сравнение скоринговых карт / С. В. Уланов // Экономические науки. — 2009. — № 9 (58). — С. 330–335.