

А.Д. ВАРЛАМОВ

**Основные метрики, оценивающие
качество работы систем поиска
изображений**

УДК 025.4.03

Муромский институт
(филиал) ФГБОУ ВПО
«Владимирский
государственный
университет имени
А.Г. и Н.Г. Столетовых»
г. Муром

Приведены и описаны основные метрики, которые используются для оценки качества алгоритмов поиска изображений, описаны их свойства и особенности применения в разных видах поиска изображений. Приведен алгоритм построения графика полнота-точность для оценки качества поиска изображений.

В области информационного поиска в настоящее время интенсивно развивается направление поиска изображений, создаются и исследуются новые методы и алгоритмы. В связи с этим актуальна задача оценки их качества и сравнения между собой [4,5]. Чтобы оценить качество работы алгоритма поиска изображений должны быть сопоставлены результаты его выполнения с оценками экспертов. При полном совпадении выводов алгоритм считается идеальным, а информационно-поисковая система, основанная на нем, идеальной. На практике различные алгоритмы в разной степени близки к идеальным. Мерами этого приближения являются метрики – численные характеристики, полученные анализом результатов работы системы и выводов экспертов на одной и той же коллекции изображений [1,2].

Так как все изображения можно разделить на релевантные и нерелевантные по критерию релевантности и на выданные и не выданные по критерию выдачи, можно составить матрицу классификации:

	Релевантные	Нерелевантные
Выданные	<i>RF</i>	<i>IF</i>
Не выданные	<i>RN</i>	<i>IN</i>

Нужно отметить, что по одному критерию изображения делит человек, а по другому – система. Если применяется поиск по визуальному образцу (когда запросом является изображение), алгоритм которого сводится к классификации пар изображений по степени тематической близости [3], в данной матрице используются другие понятия и классифицируются уже не изображения, а пары изображений.

	Подобные для человека	Неподобные для человека
Подобные для системы	RF	IF
Неподобные для системы	RN	IN

По значениям элементов матрицы вычисляется большинство метрик, которые рассмотрим ниже.

1. Полнота

Полнота (recall) определяется как отношение выданных релевантных изображений к общему количеству релевантных изображений:

$$r = \frac{RF}{RF + RN}.$$

Она характеризует способность отбирать изображения, соответствующие запросу, не пропуская нужных изображений. Чем выше полнота, тем меньше релевантных изображений будет упущено при поиске. Полнота никак не учитывает количество выданных нерелевантных изображений (то есть шум), поэтому не может применяться в одиночку как мера качества.

2. Точность

Точность (precision) вычисляется как отношение выданных релевантных изображений к общему количеству выданных изображений:

$$p = \frac{RF}{RF + IF}.$$

Точность характеризует способность системы выдавать в списке результатов только релевантные изображения, однако, не учитывает количество релевантных изображений, которые выдала система. Точность, как и полнота, не может применяться в качестве единственного критерия качества при поиске. Обычно полноту и точность используют совместно.

3 F_1 -мера

F_1 -мера (сбалансированная F-мера, мера Ван Ризбергена, F_1 -measure) используется как единая метрика, объединяющая полноту и точность. Она вычисляется как гармоническое среднее полноты и точности:

$$F_1 = \frac{2}{\frac{1}{r} + \frac{1}{p}} = \frac{2rp}{r+p}.$$

4. R-точность

Для описания этой метрики сначала дадим понятие другой характеристики. Точность на уровне n документов определяется как количество релевантных документов среди первых n выданных документов, деленное на n .

R-точность (R-precision) равна точности на уровне R изображений для R , равного количеству релевантных документов для данного запроса.

$$R = RF + RN$$

Эта метрика малочувствительна к доле релевантных изображений во всей коллекции, поэтому ее часто применяют в практических задачах.

5. Средняя точность

Средняя точность определяется следующим образом: пусть для данного запроса имеется R релевантных изображений (или R подобных пар изображений в выборке для поиска по визуальному образцу). Точность на уровне i -го релевантного изображения $p(i)$ равна точности на уровне $\text{pos}(i)$ изображений, если i -е релевантное изображение находится в результатах запроса на позиции $\text{pos}(i)$. Если i -е релевантное изображение не найдено, то $p(i)=0$. Средняя точность для данного запроса равна среднему значению величины $p(i)$ по всем R релевантным изображениям:

$$p_{\text{aver}} = \frac{1}{R} \sum_{i=1}^R p(i), R = RF + RN$$

6. Аккуратность

Аккуратность (ассигасу) - это отношение правильно принятых системой решений к общему числу решений. Формально вычисляется по формуле:

$$a = \frac{RF + IN}{RF + IN + IF + RN}.$$

7. Ошибка

Ошибка (error) - это отношение неправильно принятых системой решений к общему числу решений. Формально вычисляется по формуле:

$$e = \frac{IF + RN}{RF + IN + IF + RN}.$$

8. Информационный шум

Информационный шум вычисляется как отношение числа выданных нерелевантных изображений к общему количеству выданных изображений. Формально вычисляется по формуле:

$$n = \frac{IF}{RF + IF}.$$

9 Потери информации

Потери информации - отношение числа не выданных релевантных изображений к общему количеству релевантных изображений

$$l = \frac{RN}{RF + RN}$$

10 Специфичность

Специфичность (specificity) - отношение числа не выданных нерелевантных изображений к общему количеству нерелевантных изображений.

$$sp = \frac{IN}{IF + IN}$$

11 Избирательность

Избирательность (selectivity) - это отношение принятых системой решений выдать изображения к общему числу решений.

$$s = \frac{RF + IF}{RF + IN + IF + RN}$$

Избирательность характеризует способность отбирать изображения, соответствующие запросу, не включая лишних изображений. Специфичность и избирательность практически применяются при оценке эффективности поиска только в случаях особой необходимости.

Свойства метрик.

Как можно было наблюдать выше, каждая метрика определяет некоторую особенность алгоритма поиска, которая является, или не

является показателем качества. Отметим основные свойства метрик.

- сумма значений аккуратности и ошибки равна 1:

$$a + e = 1;$$

- сумма значений точности и информационного шума равна 1:

$$p + n = 1;$$

- сумма значений полноты и потери равна 1:

$$r + l = 1$$

- средняя точность не превышает значение полноты:

$$p_{aver} \leq r;$$

- если все релевантные изображения находятся только в начале списка результатов, то

$$p = 1, \text{ либо } r = 1;$$

$$p_{aver} = r;$$

В данном случае также возможно подобрать критерий выдачи, при котором

$$p = 1, r = 1, p_{aver} = 1;$$

- если релевантные изображения равномерно распределены по списку результатов, то

$$p_{aver} = p - r$$

- количество изображений, ранжированных ниже последнего релевантного, не влияет на значение p_{aver} (отсекается «хвост»);

- значение F_1 -меры не может быть ниже полноты и точности и превышать их среднеарифметическое:

$$\min(r, p) \leq F_1 \leq \frac{r + p}{2}.$$

Другие важные свойства сведены в таблицу 1. Возможность использования в качестве единственного критерия качества показывает, можно ли только по одному данному критерию судить о качестве алгоритма, либо, данную метрику нужно дополнить другой. Инвариантность к соотношению количества релевантных и нерелевантных изображений показывает, изменится ли данный показатель, при изменении данного соотношения. Это учитывается при оценке алгоритмов поиска и формировании требований к тестовым данным.

Свойства метрик

Метрика	Значение для идеального алгоритма	Возможность использования в качестве единственного критерия качества	Инвариантность к соотношению релевантных/нерелевантных изображений
Полнота	1	Нет	Да
Точность	1	Нет	Нет
F ₁ -мера	1	Да	Нет
R-точность	1	Да	Нет
Средняя точность	1	Да	Нет
Аккуратность	1	Да	Нет
Ошибка	0	Да	Нет
Информационный шум	0	Нет	Нет
Потери информации	0	Нет	Да
Специфичность	1	Нет	Да
Избирательность	от 0 до 1	Нет	Нет

Отношения между метриками

Можно выделить некоторое количество пар метрик с прямой зависимостью между собой, которая может быть выражена формулами. Например, аккуратность и ошибка – величины прямо противоположные, сумма их значений всегда равна единице. В данном случае значение одной метрики можно получить, зная значение другого показателя, и имеет место обыкновенная избыточность при выборе метрик как критериев качества. Поэтому больший интерес представляют метрики с непрямой зависимостью. Чаще всего в совокупности исследуется пара значений: полнота и точность. Пример зависимости этих величин показан на рисунке 1.

Из графика видна тенденция: стремление улучшить один показатель приводит к ухудшению другого. Данный график является параметрическим, где свободным параметром является критерий выдачи изображений. Ужесточение критерия приводит к сокращению количества выданных изображений и, как следствие, увеличивает

точность и сокращает полноту; ослабление критерия наоборот увеличивает полноту и уменьшает точность.

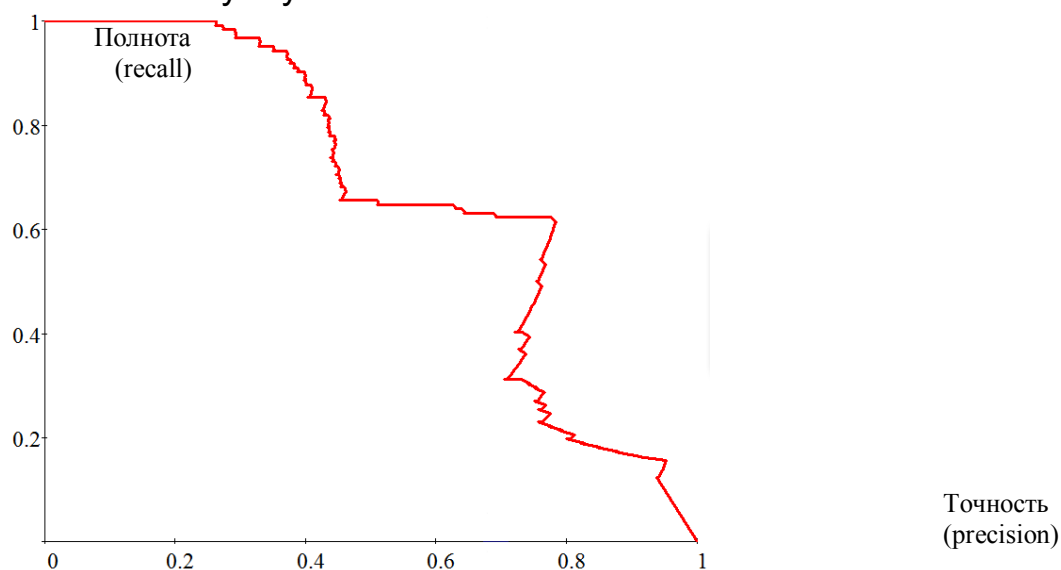


Рис. 1. График полнота-точность

Исследования совокупности этих метрик выполняют по следующим причинам:

1. По отдельности полнота и точность не могут использоваться в качестве критериев качества. Однако по паре этих значений можно сделать соответствующий вывод об алгоритме поиска.

2. Проследив за зависимостью между этими величинами можно определить оптимальное значение критерия выдачи. Например, наилучшим может быть принят такой критерий выдачи, при котором значение F_1 -меры (определяемой как раз по полноте и точности) будет наибольшим. Оптимальные значения легко визуализируется точкой на графике.

3. Можно провести анализ чувствительности, оценив, как изменятся значения метрик при незначительном изменении критерия выдачи.

4. Можно ответить, например, на один из следующих вопросов:

- Сколько система выдаст полезных изображений при заданной точности?
- Какова будет точность выдачи при требуемой полноте?
- И другие.

Данная кривая может быть построена по следующему алгоритму:

1. Берется множество пар изображений, для каждой из которых определяется степень близости между изображениями в паре (для поиска по визуальному образцу); или берется множество изображений и для каждого определяется степень близости запросу (для текстового поиска). Эти величины должны быть числовыми, с возможностью принятия одного из множества значений на некоторой шкале.

2. Все результаты ранжируются по убыванию.

3. Устанавливается критерий выдачи таким образом, чтобы в результатах выдачи было только одно изображение.

4. Определяются полнота и точность. Их значения отмечаются на графике.

5. Если в соответствии с текущим критерием были выданы все изображения, алгоритм заканчивается.

6. Устанавливается критерий выдачи таким образом, чтобы в результатах выдачи было на одно изображение больше и выполняется переход на этап 4.

Результаты исследований.

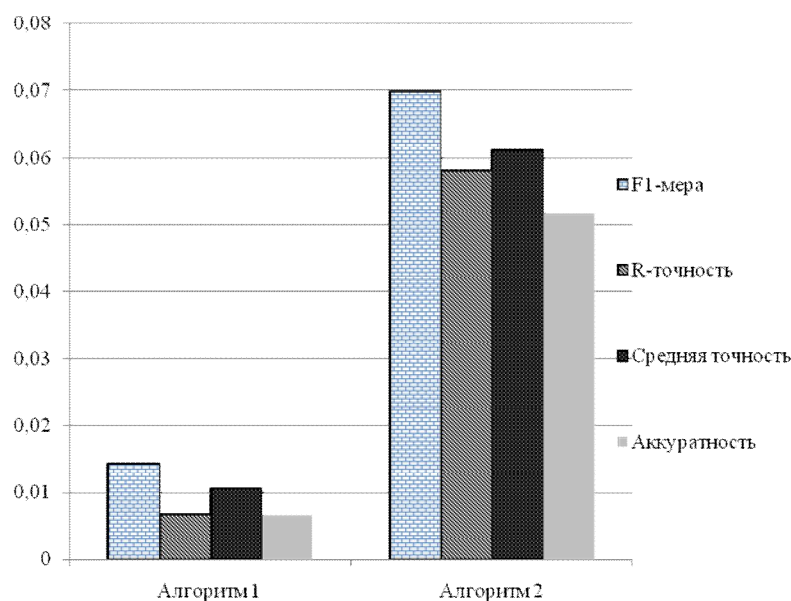


Рис. 2. Значения метрик при использовании разных алгоритмов поиска изображений

Исследовательский интерес представляют метрики, часто используемые в качестве критериев эффективности поиска документов: F_1 -мера, R-точность, средняя точность и аккуратность. Для ис-

следования были взяты два алгоритма поиска изображений: первый использовал ограниченное количество вычислимых признаков; во втором применялся полный набор признаков. Данное различие было введено намеренно с целью придать алгоритмам разные качественные характеристики: априорно первый алгоритм имеет худшие показатели по отношению ко второму. Применив алгоритмы к подготовленной тестовой коллекции изображений и, рассчитав метрики, были получены величины, отображенные на рисунке 2.

Заметим, что каждый показатель отличает оба алгоритма по качеству в пользу второго. Также можно отметить, что соотношения значений каждой метрики примерно одинаковые.

В заключении можно отметить, что выбор метрики (метрик) для машинного обучения возлагается на программиста, однако, сравнение алгоритмов поиска изображений рекомендуется проводить с использованием всех описанных в статье метрик.

Литература

1. Агеев М., Кураленок И., Некрестьянов И. Официальные метрики РОМИП 2010. "Российский семинар по Оценке Методов Информационного Поиска. Труды РОМИП 2010". Казань, 2010, с. 172-187.

2. Гулин А., Карпович П., Расковалов Д., Сегалович И. Яндекс на РОМИП'2009. Оптимизация алгоритмов ранжирования методами машинного обучения // Российский семинар по Оценке Методов Информационного Поиска. Труды РОМИП 2009, с.163-168.

3. Жизняков А.Л., Зуев В.В. Влияние изменения размерности вектор-контуров изображений на их меру близости. Вопросы радиоэлектроники. 2010. Т. 1. № 1. С. 165-170.

4. Канунова Е.Е. Система хранения, распределения и передачи изображений архивных документов. Алгоритмы, методы и системы обработки данных. 2006. № 11. С. 92-97.

5. Шаратов Р.В., Шаралова Е.В. Применение теории нечетких множеств в информационном поиске. Методы и устройства передачи и обработки информации. 2009. № 11. С. 427-432.

СЛ. ТЕЛЕФОН: (49234) 7-71-22.

Д. ТЕЛЕФОН: (83175) 5-57-56.

E-MAIL: VARLAMOV_ALEKS@MAIL.RU