

ОБРАБОТКА ИЗОБРАЖЕНИЙ, СИГНАЛОВ И РЕЧИ

УДК 004.9

В.А. Ковалев, А.А. Дмитрук

**ВЛИЯНИЕ МЕР БЛИЗОСТИ В ПРОСТРАНСТВЕ ПРИЗНАКОВ
НА КАЧЕСТВО ПОИСКА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ
ПО СОДЕРЖАНИЮ**

Проводится экспериментальное исследование влияния используемых мер близости в пространстве признаков на качество поиска медицинских изображений. Приводятся результаты сравнения 16 наиболее распространенных метрик при решении задачи поиска рентгеновских изображений по образцу на основе тестовой базы данных, состоящей из 3000 изображений грудной клетки испытуемых обоего пола в возрасте 20, 40 и 60 лет. Показано, что выбор метрики существенным образом влияет на качество результатов поиска, а наилучшие результаты достигаются при вычислении степени близости с использованием расстояния Kullback – Leibler.

Введение

Известно, что, несмотря на появление и развитие разнообразных методов и средств количественного анализа изображений для поддержки процессов медицинской диагностики, одним из распространенных диагностических приемов по-прежнему остается сравнение полученного изображения с предыдущими и/или с похожими снимками и случаями из клинической практики [1, 2]. Технической основой для повышения эффективности диагностических процедур, основанных на оперативном привлечении визуальной информации о похожих случаях, является развитие так называемых систем поиска изображений по содержанию. Как правило, существующие системы поддержки принятия решений данного направления обеспечивают лишь традиционные средства поиска по тексту и/или по некоторым другим метаданным, ассоциированным с изображениями. Однако кроме указанных метаданных при постановке диагноза врач часто учитывает неформализованные визуальные признаки, которые чрезвычайно трудно либо вовсе невозможно представить в виде эквивалентных словесных описаний. Таким образом, существует определенный семантический разрыв между реальными потребностями клиницистов и используемыми ими диагностическими изображениями, с одной стороны, а также данными, представляемыми существующими медицинскими информационными системами, – с другой. По указанным причинам в последнее время в мире существенно вырос интерес к проблеме поиска изображений по содержанию, обычно обозначаемой в англоязычной литературе термином Content-Based Image Retrieval [3].

По мере развития технических средств регистрации медицинских изображений объем хранимых видеоданных растет экспоненциально. Так, например, средний размер одного компьютерного томографического изображения грудной клетки пациента составляет порядка 30–40 МБ. Промышленные системы хранения и передачи изображений среднего уровня, ориентированные на организацию среднесрочного хранения цифровых радиологических изображений, рассчитаны на объемы видеоданных порядка 10–20 ТБ. По имеющимся оценкам, полный переход на цифровые диагностические изображения приведет к тому, что суммарный объем изображений, производимых крупными больницами, может достигнуть петабайт, т. е. величины порядка 10^{15} . Следует заметить, что значительную долю в этом объеме составляют современные цифровые рентгеновские изображения, получаемые в результате скрининга населения. Процесс массового внедрения используемых при этом цифровых рентгеновских сканеров с низкой лучевой нагрузкой идет в настоящее время во многих странах мира, включая Беларусь. Так, например, на момент написания статьи общий объем накопленных в базах данных (БД) Минска цифровых изображений грудной клетки составляет порядка полумиллиона снимков. Естественно, что при подобных сверхбольших объемах хранимых изображений никакой речи о «ручном», т. е. визуальном, поиске в БД изображений «похожих» снимков и подборе аналогичных диагностических ситуаций идти не может. Таким образом, разработка средств автоматического индексирования содержания и

поиска изображений БД по заданному образцу является одной из актуальных проблем в области обработки, анализа и распознавания медицинских изображений.

Независимо от используемой схемы поиска по содержанию одним из важнейших этапов поиска является вычисление меры сходства сравниваемых изображений. Технически данная задача представляет собой сравнение векторов признаков изображений между собой с целью получения некоторого числового значения, характеризующего их «похожесть» в смысле выбранной метрики. Поэтому очевидно, что выбор подходящей метрики оказывает существенное влияние на качество функционирования разрабатываемой системы поиска [4, 5].

Целью настоящей работы являются экспериментальное исследование влияния используемых мер близости в пространстве признаков на качество поиска медицинских изображений и их сравнительная оценка. Приводятся результаты сравнения 16 наиболее распространенных метрик при решении задачи поиска рентгеновских изображений по образцу на основе тестовой базы данных, состоящей из 3000 изображений грудной клетки.

1. Тестовая БД изображений

Экспериментальное исследование влияния мер близости в пространстве признаков изображений на качество поиска проводилось с использованием тестовой БД цифровых изображений грудной клетки 3000 здоровых взрослых испытуемых различного возраста и пола. В соответствии с существующей медицинской практикой возраст выражался в количестве полных лет, т. е. измерялся с точностью до одного года. При этом были выделены три возрастные группы испытуемых в возрасте 20, 40 и 60 лет соответственно. Каждая группа включала по 1000 человек, в том числе 500 женщин и 500 мужчин, представляющих собой естественные подгруппы соответствующих возрастных групп. Таким образом, при тестировании средств поиска рассматривалось шесть формальных классов по 500 изображений в каждом, которые были образованы путем деления исходных 3000 изображений на три категории в соответствии с возрастом испытуемых, а затем еще на две категории в соответствии с их полом (рис. 1).

Изображения грудной клетки снимались с помощью цифровых сканирующих рентгеновских аппаратов серии «Пульмоскан» отечественного производства [2]. Размер изображений составлял 1000×1500 пикселей. Яростное разрешение исходных изображений соответствовало 12 битам на пиксел. До начала экспериментов все изображения были сконвертированы в стандартное восьмибитное представление путем отсечения (сатурации) 1,5 % крайних (как правило, шумовых) значений гистограммы яркостей и масштабирования полученного диапазона яркостей в интервал 0–255.

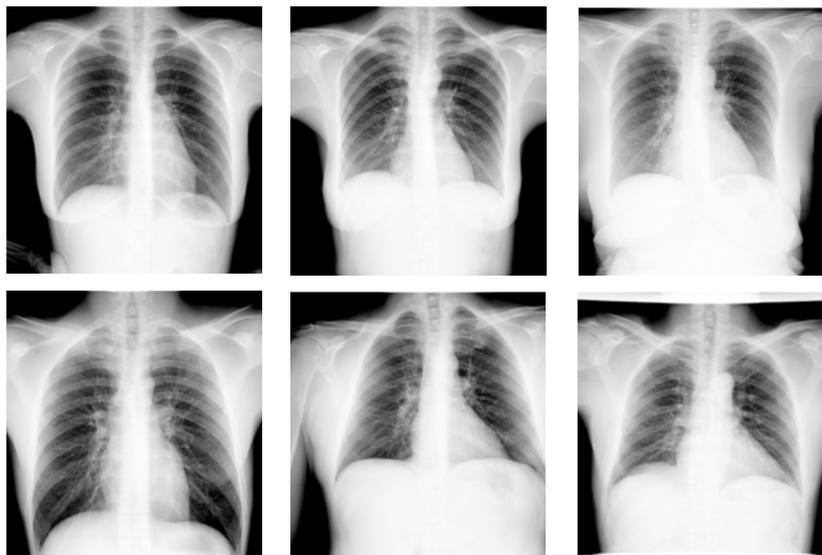


Рис. 1. Примеры классов рентгеновских изображений тестовой БД. В строках представлены изображения, различающиеся по полу испытуемых (верхняя строка – женщины, нижняя – мужчины), а в столбцах – по возрасту (слева направо: 20, 40 и 60 лет)

Таким образом, тестовая БД содержала изображения испытуемых женского и мужского пола, принадлежащие трем различным возрастным группам. Естественно, для каждой возрастной группы и для каждого пола характерны свои специфические, объективно существующие визуальные признаки, которые как-то проявляются на изображениях. Тем не менее выделить данные признаки и отнести каждое изображение к соответствующей категории далеко не просто как для разрабатываемых методов и программных средств, так и для опытного врача-рентгенолога. Поэтому использование БД рентгеновских изображений грудной клетки для решения модельной задачи автоматического поиска изображений людей, принадлежащих к заданной возрастной и/или половой группе, может служить хорошей основой для проведения объективного количественного тестирования.

2. Дескрипторы изображений

Поскольку рентгеновские изображения грудной клетки достаточно хорошо описываются пространственным распределением яркостей пикселей и, что в данном случае более важно, целью данного исследования является сравнение мер близости, а не распознавание изображений как таковое, в качестве дескрипторов использовались редуцированные версии обобщенных матриц совместной встречаемости типа *IID* [6, 7]. Матрицы типа *IID* (латинские буквы аббревиатуры соответствуют типам каждого из четырех измерений, т. е. в данном случае «интенсивность 1 – интенсивность 2 – интенсивность 3 – расстояние») описывают частоты совместной встречаемости троек пикселей, расположенных в углах всех возможных равносторонних треугольников на изображении с заданными расстояниями между пикселями (сторонами треугольников). Яркости (интенсивности «*I*») пикселей обычно квантуются путем разбиения значений 0–255 на достаточно небольшое количество уровней, а расстояния *D* измеряются в единицах раstra, обычно локально, в пределах 1–7 пикселей. На уровне программной реализации матрица совместной встречаемости типа *IID* представляет собой четырехмерную гистограмму (массив частот встречаемости) с нулевыми значениями над главной диагональю каждого из двухмерных сечений по однотипным измерениям (в данном случае «интенсивность – интенсивность»). Соответственно дескриптор изображения – это не что иное, как четырехмерная матрица *IID*, которая представляется в виде последовательности элементов, выписанных друг за другом в вектор подходящей длины, и интерпретируется как вектор признаков изображения. Все необходимые технические подробности могут быть найдены в работах [6–8].

При проведении экспериментов весь диапазон яркостей разбивался на 16 уровней и рассматривались всевозможные элементарные пространственные структуры (равносторонние треугольники) с размерами сторон 1, 3 и 5 пикселей. Опыт показывает, что дальнейшее увеличение данного параметра для используемого типа изображений нецелесообразно, поскольку взаимозависимость (например, значение автокорреляционной функции) между яркостями троек пикселей быстро падает с увеличением межпиксельного расстояния [9]. Для заданного параметра квантования яркости и трех рассматриваемых межпиксельных расстояний, а также с учетом нижней треугольности двухмерных сечений «интенсивность – интенсивность» матриц *IID* размер вектора признаков изображения составляет 2448 элементов.

3. Исследуемые меры близости в пространстве признаков изображений

В настоящей работе исследуются 16 наиболее распространенных метрик, используемых для вычисления степени близости сравниваемых изображений в контексте задачи поиска рентгеновских изображений по образцу, т. е. отбора из БД изображений, «наиболее похожих» на заданный образец поиска (так называемая технология Query-by-Example). Краткое название и формальное определение каждой метрики представлены ниже. Элементы сравниваемых векторов *X* и *Y* обозначены через x_i и y_i соответственно. Во избежание ненужной путаницы, связанной с известными различиями в терминологии, в дальнейшем приводятся оригинальные (англоязычные) версии названий метрик, используемые в международной практике:

Manhattan:	$L1 = \sum_i x_i - y_i $.
Euclidean:	$L2 = \sqrt{\sum_i (x_i - y_i)^2}$.
Chebyshev:	$L3 = \max_i x_i - y_i $.
Chord:	$L4 = \sqrt{2 \left(1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}} \right)}$.
Correlation:	$L5 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$.
Cosine:	$L6 = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}}$.
Curtis:	$L7 = \frac{\sum_i x_i - y_i }{\sum_i (x_i + y_i)}$.
Divergence:	$L8 = \sum_i \frac{(x_i - y_i)^2}{(x_i + y_i)^2}$.
Bhattacharyya:	$L9 = \sqrt{\sum_i (\sqrt{x_i} - \sqrt{y_i})^2}$.
Extended Jaccard:	$L10 = \frac{\sum_i x_i y_i}{\sum_i (x_i^2 + y_i^2 - x_i y_i)}$.
Geodesic:	$L11 = \arccos \left(\frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}} \right)$.
Hellinger:	$L12 = \sqrt{\sum_i \left(\sqrt{x_i / \sum_i x} - \sqrt{y_i / \sum_i y} \right)^2}$.
Kullback – Leibler:	$L13 = \sum_i \left[x_i \log \frac{\sum_j x_j / \sum_j y_j}{\sum_j x_j} \right]$.
Canberra:	$L14 = \sum_i \frac{ x_i - y_i }{ x_i + y_i }$.
Soergel:	$L15 = \frac{\sum_i x_i - y_i }{\sum_i \max\{x_i, y_i\}}$.
Whittaker:	$L16 = \frac{1}{2} \sum_i \left \frac{x_i}{\sum_i x_i} - \frac{y_i}{\sum_i y_i} \right $.

Программная реализация средств оценки влияния используемых расстояний (мер близости) на результаты поиска выполнена с использованием библиотеки программ «ргоху» [8] системы программирования и статистического анализа R.

4. Методика оценки качества поиска изображений

Качество поиска изображений оценивалось путем оценки способности системы находить изображения «правильного» класса в тестовой БД. Тестирование выполнялось по схеме поиска изображений по образцу, т. е. отбора из базы 20 изображений, наиболее похожих на заданный образец с сортировкой полученных результатов по степени их близости (в смысле используе-

мой метрики) к образцу. Методика тестирования включала подачу на вход системы каждого изображения БД и оценку правильности получаемых результатов. Изображение-результат поиска, стоящее на k -м месте ($k = 1, \dots, 20$), считалось правильным, если оно принадлежало той же возрастной группе и тому же полу, что и изображение-образец (т. е. запрос на поиск). После выполнения всех 3000 запросов подсчитывался процент правильных ответов по каждой k -й позиции результата отдельно. Данная серия экспериментов повторялась 16 раз, по одному полному циклу из 3000 запросов для каждой из используемых мер близости.

5. Результаты

Результаты сравнения 16 наиболее распространенных метрик при решении задачи поиска рентгеновских изображений по образцу представлены на рис. 2 и 3 в виде графиков, которые выражают процент правильных ответов, стоящих на первом ($k = 1$, изображение, наиболее похожее на образец), втором и так далее, вплоть до двадцатого, месте в рейтинге близости. При этом кривые на рис. 2 выражают суммарные результаты поиска по трем возрастным группам, включающим изображения грудной клетки испытуемых в возрасте 20, 40 и 60 лет при использовании различных мер близости. На рис. 3 показаны результаты поиска в условиях, когда учитывается только пол испытуемых, т. е. получаемые в результате поиска изображения испытуемых относятся к тому же самому полу, что и изображение-запрос, вне зависимости от его возрастной группы.

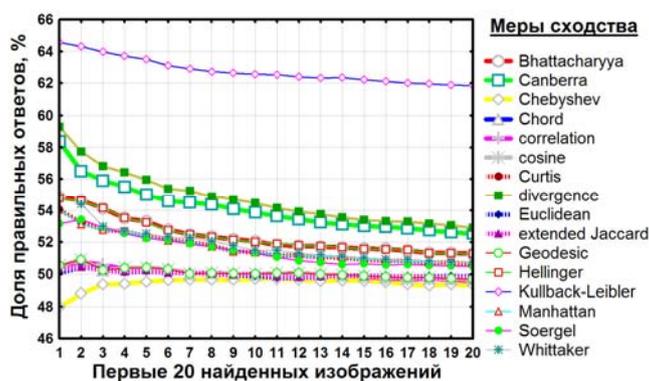


Рис. 2. Качество поиска изображений по возрастной группе

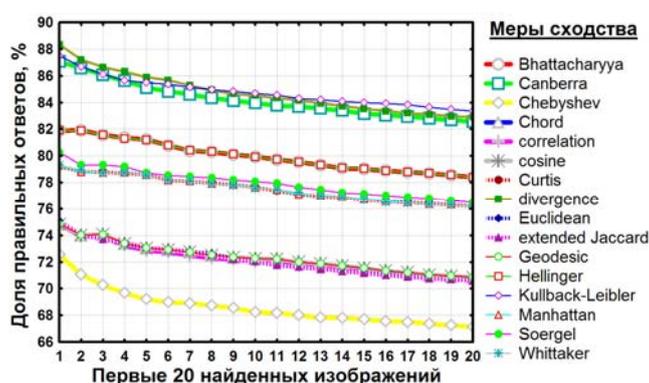


Рис. 3. Качество поиска изображений по полу

Как видно из кривых на рис. 2, в зависимости от типа метрики качество поиска для первого места изменяется в широких пределах – от 48 % при использовании расстояния Chebyshev до 64,6 % при использовании расстояния Kullback – Leibler. Кроме того, легко видеть, что результаты поиска при оценке степени близости в пространстве признаков на основе метрики Kullback – Leibler превосходят аналогичные показатели, полученные при использовании других метрик, примерно на 8–10 %. Ко второму эшелону лидеров, следующих со значительным отрывом от

метрики Kullback – Leibler, относятся метрики Canberra и Divergence, демонстрирующие очень похожие результаты. Самые низкие результаты показывает группа метрик, включающая расстояния Geodesic, Extended Jaccard, Cosine и Chebyshev. Остальные девять метрик демонстрируют некоторые промежуточные результаты в качестве поиска изображений по возрастной группе испытуемых. Следует обратить внимание читателя на то, что рассматриваемая тестовая задача является весьма трудной даже для специалиста-рентгенолога высшей квалификации. В частности, многочисленные эксперименты с отнесением изображений испытуемых используемой тестовой БД к указанным возрастным группам показали, что качество решения задачи экспертом составляет 63,4 %. Это формально даже несколько уступает автоматическому поиску с использованием метрики Kullback – Leibler на величину порядка 1 %.

Три верхних графика (см. рис. 3) показывают, что в случае решения задачи поиска по полу наилучшие результаты достигаются при использовании трех следующих метрик: Kullback – Leibler, Divergence и Canberra. В количественном выражении их применение обеспечивает близкое качество поиска – порядка 85 % с незначительными колебаниями процента правильных ответов в зависимости от места в рейтинге лучших 20 результатов. Следующие две метрики (Hellinger и Bhattacharyya) практически полностью совпадают и имеют качество поиска примерно 80 %. Немного худшие результаты в 78 % показали метрики Whittaker, Soegel, Curtis и Manhattan, которые опять же практически совпадают друг с другом, вследствие чего их трудно разделить визуально на основе графика, представленного на рис. 3. За указанной группой следует плотное семейство графиков, соответствующих мерам сходства Geodesic, Correlation, Chord, Cosine, Euclidean и Extended Jaccard с качеством поиска на уровне 72 %. Как и в предыдущем случае, замыкает список мер сходства расстояние Chebyshev. Таким образом, наилучшие результаты для данного теста снова демонстрирует расстояние Kullback – Leibler с 87,5 % правильных ответов. Данный результат всего на 1,5 % хуже результата, показанного при решении той же самой задачи экспертом.

Следует отметить, что проведенный в данной работе количественный и качественный анализ зависимости результатов от типа используемой меры близости в пространстве признаков относится лишь к задаче поиска изображений по образцу. Задача теоретического сравнения мер близости как таковых здесь не рассматривалась. Результаты исследований данного типа могут быть найдены в соответствующей литературе (см., например, [10]).

Заключение

Результаты проведенных экспериментальных исследований позволяют сделать следующие выводы:

1. Выбор метрики, используемой для сравнения векторов признаков при решении задачи поиска рентгеновских изображений по образцу, существенным образом влияет на качество результатов поиска.

2. Наилучшие результаты как при поиске по полу, так и при поиске по возрастной группе показала мера близости, вычисляемая с использованием расстояния Kullback – Leibler, которая трактует сравниваемые векторы признаков как экспериментальные версии функций плотностей вероятностей, т. е. как гистограммы частот.

Опыт решения задач поиска по содержанию биомедицинских изображений различных типов (см., например, работы [1, 2, 6, 9]) с использованием в качестве дескрипторов изображений обобщенных матриц совместной встречаемости показывает, что выводы относительно эффективности исследуемых мер близости, сделанные по результатам данного исследования, могут быть применимы ко всем задачам поиска, решаемым на основе предложенного авторами подхода.

Данная работа частично финансировалась из средств проекта МНТЦ В-1682.

Список литературы

1. Ковалев, В.А. Анализ структуры трехмерных медицинских изображений / В.А. Ковалев. – Минск : Беларуская навука, 2008. – 263 с.
2. Применение цифровых сканирующих аппаратов и передовых телемедицинских технологий в диагностике заболеваний легких / В.В. Анищенко [и др.]. – Минск : ОИПИ НАН Беларуси, 2010. – 136 с.

3. Müller, H. Content-Based Medical Image Retrieval / H. Müller, T.M. Deserno // *Biomedical Image Processing*. – Heidelberg : Springer, 2011. – P. 471–494.
4. A comparison of similarity measures for use in 2-D-3-D medical image registration / G.P. Penney [et al.] // *IEEE Transactions on Medical Imaging*. – 1998. – Vol. 17. – P. 586–595.
5. Learning similarity measure for natural image retrieval with relevance feedback / G.D. Guo [et al.] // *IEEE Transactions on Neural Networks*. – 2002. – Vol. 13. – P. 811–820.
6. Ковалев, В.А. Поиск медицинских изображений по содержанию в условиях шумов / В.А. Ковалев, А.А. Дмитрук // *Информатика*. – 2009. – № 2 (22). – С. 39–48.
7. Three-dimensional texture analysis of MRI brain datasets / V.A Kovalev [et al.] // *IEEE Transactions on Medical Imaging*. – 2001. – Vol. 20, № 5. – P. 424–433.
8. Proxy: Distance and Similarity Measures [Electronic resource]. – 2010. – Mode of access : <http://cran.r-project.org/web/packages/proxy/index.html>. – Date of access : 10.12.2010.
9. Дмитрук, А.А. Сравнительный анализ эффективности поиска изображений мозга различных модальностей / А.А. Дмитрук, В.А. Ковалев // АИТТН'2008 : материалы II Междунар. конф. «Современные информационные и телемедицинские технологии для здравоохранения». – Минск, 2008. – С. 174–178.
10. Старовойтов, В.В. Локальные геометрические методы цифровой обработки и анализа изображений / В.В. Старовойтов. – Минск : Ин-т техн. кибернетики НАН Беларуси, 1997. – 284 с.

Поступила 14.12.10

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: vassili.kovalev@gmail.com,
dmitruk@newman.bas-net.by*

V.A. Kovalev, A.A. Dmitruk

INFLUENCE OF SIMILARITY MEASURES ON THE QUALITY OF MEDICAL IMAGE RETRIEVAL

This paper presents results of comparative study of 16 most commonly used similarity measures and their influence on the results of medical image retrieval. The test database consists of X-ray chest images of 3000 healthy subjects aged 20, 40 and 60 years including 1500 males and 1500 females. The testing was performed using query-by-example paradigm. Results of the study suggest that selection of specific similarity measures for comparing images in feature space substantially influences the final retrieval results. It was found that the most prominent similarity measure is the Kullback – Leibler divergence.