

УДК 514.18

Д.А. Ницын

Харьковский государственный университет питания и торговли, Харьков

ПРИЛОЖЕНИЕ ПРАВИЛА БАЙЕСА К КЛАССИФИКАЦИИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Предлагается методика классификации рентгеновских изображений по их фрактальной размерности. Методика основывается на оценке вероятности того, что при измеренной фрактальной размерности изображения данная рентгенограмма не содержит патогенных областей.

Ключевые слова: правило Байеса, фрактальная размерность.

Введение

В настоящее время в связи с ростом числа заболеваний туберкулезом развитие автоматизированных диагностических систем является актуальной задачей [1 – 3]. Одним из методов раннего обнаружения туберкулеза является фрактальный анализ рентгеновских снимков.

Это обусловлено тем, что рентгенограмма легких человека содержит фрагменты изображения, имеющие древовидную структуру [4, 5].

Таким образом, для эффективного диагностирования туберкулеза необходимо разработать методику, которая позволяет по измеренному значению фрактальной размерности классифицировать рентгенограммы на снимки, содержащие патогенные области, и на снимки, не имеющие признаков заболевания.

Анализ литературы. Применению методов фрактальной геометрии к классификации медицинских изображений посвящены работы многих авторов [4 – 6].

Однако, как правило, в этих работах применяется детерминированный подход, в то время как, по нашему мнению, достоверность диагноза повышается, если результат анализа медицинского изображения представить в виде вероятности того, что данная рентгенограмма не содержит признаков заболевания. Поэтому разработка вероятностных методов классификации рентгеновских снимков является актуальной задачей медицинской диагностики.

Цель статьи. Таким образом, целью статьи является разработка метода классификации рентгеновских изображений по их фрактальной размерности. При этом метод основывается на оценке вероятности того, что при данном спектре фрактальных размерностей для фрактальной меры на множестве пикселей данной яркости предьявленная рентгенограмма не содержит патогенных областей изображения.

Методика определения вероятности диагноза

Поскольку рентгенограмма представляет собой однотонное изображение, под диагностическим признаком будем понимать фрактальную размерность множества пикселей, имеющих одинаковую яркость. Дадим способ определения меры множества точек, принадлежащих изображению. Выберем квадрат, площадь которого равна δ^d . Покроем данным обобщенным квадратом пространство, заполненное точками изображения. Получим меру множества точек в данном пространстве $M_d = \sum \delta^d$. Заметим, что в общем случае при стремлении размера квадрата к нулю $\delta \rightarrow 0$ мера множества точек M_d равна нулю или бесконечности в зависимости от выбора величины d , которую будем называть размерностью меры. При этом размерность Хаусдорфа – Безиковича D множества точек пространства представляет собой критическую размерность, при которой мера M_d изменяет свое значение с нуля на бесконечность [4, 5]:

$$M_d = \sum \delta^d = N(\delta) \delta^d \xrightarrow{\delta \rightarrow 0} \begin{cases} 0, d > D, \\ \infty, d < D. \end{cases} \quad (1)$$

Определение меры множества точек M_d , выраженное формулой (1), можно применить для практического вычисления его фрактальной размерности D . Действительно, из формулы (1) следует, что при стремлении размера квадрата к нулю $\delta \rightarrow 0$ число квадратов $N(\delta)$, покрывающих выделенную область изображения, обратно пропорционально величине δ^D :

$$N(\delta) \sim \frac{1}{\delta^D}.$$

Рассмотрим фрактальную размерность множе-

ства точек изображения как угловой коэффициент графика, выражающего зависимость числа квадратов $N(\delta)$ от размера δ квадрата, построенного в два раза логарифмическом масштабе

$$\ln N(\delta) \sim -D \ln \delta.$$

Отсюда следует формула, по которой можно вычислить фрактальную размерность множества

$$D \sim -\frac{\ln N(\delta)}{\ln \delta}.$$

Если изображение является полутоновым, в качестве диагностических признаков выберем спектры фрактальных размерностей для фрактальной меры на множестве пикселей данной яркости. При этом под спектром фрактальных размерностей будем понимать совокупность фрактальных размерностей множества пикселей данной яркости, которые также можно вычислить с помощью подсчета числа клеток, содержащих пиксели данной яркости.

Определим вероятность того, что рентгенограмма не содержит патогенных зон, при данном спектре фрактальных размерностей для фрактальной меры на множестве пикселей данной яркости. Произведем случайную выборку из рентгенограмм, содержащих симптомы болезни, и случайную выборку из рентгенограмм, не содержащих симптомов болезни.

Вычислим спектры фрактальных размерностей для фрактальной меры на множестве пикселей данной яркости каждого рентгеновского изображения. Чтобы найти законы и параметры распределения вероятности диагноза в зависимости от данного спектра фрактальных размерностей, произведем статистическую обработку числовых значений фрактальных размерностей рентгеновских изображений, как содержащих, так и не содержащих симптомы болезни.

При этом статистическая обработка производится в следующей последовательности:

выберем длину и количество интервалов яркости, по которым распределяются пиксели данного изображения. Представим совокупность интервалов, по которым группируются пиксели изображения, в виде множества L_i , $i = 1, \dots, m$;

выберем длину и количество интервалов, которым могут принадлежать значения фрактальных размерностей областей изображения, содержащих пиксели данной яркости. Представим совокупность интервалов, по которым группируются значения фрактальных размерностей, в виде множества D_j , $j = 1, \dots, n$;

выполним группировку значений фрактальной размерности по выбранным интервалам ее измене-

ния и выбранным интервалам яркости, по которым распределяются пиксели данного изображения. Представим множество значений фрактальных размерностей в виде матрицы, которая состоит из столбцов, соответствующих данному интервалу изменения яркости Φ_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$. Тогда под спектром фрактальных размерностей будем понимать сочетание S_k значений фрактальных размерностей, взятых из каждого столбца матрицы Φ_{ij} , соответствующего данному интервалу L_i изменения яркости, и данного интервала D_j изменения фрактальных размерностей;

найдем распределение статистической вероятности данного спектра фрактальных размерностей для фрактальной меры на множестве пикселей данной яркости для рентгеновских изображений, содержащих симптомы заболевания;

найдем распределение статистической вероятности данного спектра фрактальных размерностей для фрактальной меры на множестве пикселей данной яркости для рентгеновских изображений, содержащих симптомы заболевания.

Пусть фрактальный анализ рентгенограмм позволяет выявить два состояния, а именно: состояние H_1 , к которому относятся рентгеновские снимки, не содержащие патогенных зон, и состояние H_2 , к которому относятся рентгеновские снимки, содержащие патогенные зоны. Воспользуемся формулой Байеса [7, 8] для оценки апостериорной вероятности того, что данное рентгеновское изображение не имеет признаков заболевания, при условии, что фрактальные размерности множества пикселей данной яркости имеют данное сочетание

$$P(H_1/S_k) = \frac{P(H_1)P(S_k/H_1)}{P(H_1)P(S_k/H_1) + P(H_2)P(S_k/H_2)}, \quad (2)$$

где $P(H_1)$ – априорная вероятность того, что рентгенограмма не содержит патогенных зон; $P(H_2)$ – априорная вероятность того, что рентгенограмма содержит патогенные зоны; $P(S_k/H_1)$ – условная вероятность данного сочетания фрактальных размерностей множества пикселей данной яркости при условии, что рентгеновское изображение не содержит симптомов туберкулеза; $P(S_k/H_2)$ – условная вероятность данного сочетания фрактальных размерностей множества пикселей данной яркости при условии, что рентгеновское изображение содержит симптомы туберкулеза.

Определим пороговое значение вероятности то-

го, что данное рентгеновское изображение не содержит симптомов болезни. Найдем законы и параметры распределения вероятности диагноза, указывающего на здоровье пациента, для рентгеновских изображений, как содержащих, так и не содержащих патологию.

Для этого проведем статистическую обработку значений фрактальных размерностей рентгенограмм пациентов, признанных здоровыми, и рентгенограмм пациентов, признанных больными. При этом статистическая обработка проводится в следующей последовательности:

произведем случайную выборку из рентгеновских изображений, содержащих симптомы болезни, и случайную выборку из рентгеновских изображений, не содержащих симптомов болезни;

вычислим спектры фрактальных размерностей для фрактальной меры на множестве пикселей данной яркости для рентгеновских изображений, как содержащих, так и не содержащих патологию;

определим распределение апостериорной вероятности $f(P/H_1)$ того, что рентгенограмма не содержит патогенных зон, в зависимости от данного сочетания фрактальных размерностей множества пикселей данной яркости;

определим распределение апостериорной вероятности $f(P/H_2)$ того, что рентгенограмма содержит патогенные зоны, в зависимости от данного сочетания фрактальных размерностей множества пикселей данной яркости.

Установим пороговое значение вероятности того, что данное рентгеновское изображение не имеет симптомов болезни, при данном спектре фрактальных размерностей. Сравним апостериорную вероятность, полученную с помощью формулы (2), с ее пороговым значением. Если значение вероятности диагноза выше порогового $P(H_1/S_k) > P_0$, то принимается решение о том, что пациент здоров. Если значение вероятности диагноза ниже порогового $P(H_1/S_k) < P_0$, то принимается решение о том, что пациент болен.

При этом вероятность принятия ошибочного решения состоит из вероятности того, что пациент, для которого оценка вероятности диагноза, указывающего на его здоровье, выше ее порогового значения, будет признан больным, и вероятности того, что пациент, для которого оценка вероятности диагноза, указывающего на его здоровье, ниже ее порогового значения, будет признан здоровым. Отсюда следует, что риск принятия ошибочного решения складывается из ущерба для здоровья пациента, признанного здоровым, но впоследствии заболевше-

го туберкулезом, и ущерба, связанного со стоимостью лечения пациента, признанного больным, но на самом деле являющегося здоровым. Оценим ущерб от принятия ошибочного решения по следующей формуле [9]:

$$R = C_1 P(H_1) \int_{P_0}^{\infty} f(P/H_1) dP + C_2 P(H_2) \int_{-\infty}^{P_0} f(P/H_2) dP,$$

где $f(P/H_1)$ – плотность распределения величины вероятности того, что данная рентгенограмма не содержит патогенных зон, в зависимости от данного сочетания фрактальных размерностей множества пикселей данной яркости; $f(P/H_2)$ – плотность распределения величины вероятности того, что рентгенограмма содержит патогенные зоны, в зависимости от данного сочетания фрактальных размерностей множества пикселей данной яркости; C_1 – весовой коэффициент, учитывающий риск принятия решения о том, что рентгенограмма не содержит патогенных зон, в то время как рентгенограмма принадлежит пациенту, больному туберкулезом; C_2 – весовой коэффициент, учитывающий риск принятия решения о том, что рентгенограмма содержит патогенные зоны, в то время как рентгенограмма принадлежит здоровому пациенту.

Пороговое значение вероятности определяется по методу наибольшего правдоподобия [9], являющегося частным случаем метода наименьших материальных затрат.

Это обусловлено тем, что риск принятия решения о том, что рентгенограмма не содержит патогенных зон, в то время как рентгенограмма принадлежит пациенту, больному туберкулезом, во много раз больше стоимости лечения пациента, признанного больным, но на самом деле являющегося здоровым, то есть $C_1 \gg C_2$. Поэтому пороговое значение вероятности положительного диагноза, которое обеспечивает минимальную вероятность принятия ошибочного решения о том, что данное рентгеновское изображение не имеет симптомов болезни, находится из следующего условия [9]:

$$f(P_0/H_1) = f(P_0/H_2).$$

Таким образом, если прогнозировать положительный диагноз в той области сочетаний интервала изменения яркости и фрактальной размерности множества пикселей, в которой апостериорная вероятность положительного диагноза выше порогового значения $P(H_1/S_k) > P_0$, можно уменьшить число рентгенограмм, дающих положительный диагноз, но принадлежащих лицам, больным туберкулезом. Поскольку методика диагностики туберкулеза предусматривает прогнозирование положительного

диагноза в области сочетаний интервала изменения яркости и фрактальной размерности множества пикселей, в которой апостериорная вероятность положительного диагноза выше порогового значения $P(H_1/S_k) > P_0$, апостериорная вероятность положительного диагноза возрастает после ее применения.

Оценка эффективности применения методики по определению вероятности диагноза

Выведем формулу, по которой вычисляется априорная вероятность положительного диагноза $P(H_1)$ после применения методики классификации рентгенограмм.

Очевидно, что априорная вероятность положительного диагноза $P(H_1)$ после применения методики классификации рентгенограмм увеличивается за счет уменьшения числа рентгенограмм, принадлежащих лицам, больным туберкулезом, вследствие того, что при анализе статистических данных отбрасываются рентгенограммы с сочетанием интервала изменения яркости и фрактальной размерности множества пикселей, при котором апостериорная вероятность положительного диагноза меньше порогового значения

$$P(H_1/S_k) > P_0.$$

Отсюда следует равенство, которое связывает частоты предъявления рентгенограмм, не содержащих симптомы болезни, до и после применения методики диагностики туберкулеза

$$\frac{\bar{N}^1}{N^1 + N^2} = \frac{N^1}{N^1 + N^2} + \sum_{k=1}^l \frac{N_k^2}{N^1 + N^2},$$

где N^1 – число объектов наблюдения, для которых верна гипотеза H_1 о том, что рентгенограмма не содержит симптомы болезни; N^2 – число объектов наблюдения, для которых верна гипотеза H_2 о том, что рентгенограмма содержит симптомы болезни; l – число сочетаний интервала изменения яркости и фрактальной размерности множества пикселей, для которых апостериорная вероятность положительного диагноза меньше порогового значения $P(H_1/S_k) > P_0$.

Заметим, что второе слагаемое, стоящее в правой части данного равенства, представляет собой сумму статистических оценок полных вероятностей $P(S_k, H_2)$ данного сочетания S_k диагностических признаков при условии, что верна гипотеза H_2 о том, что рентгенограмма содержит симптомы болезни.

Поэтому априорная вероятность положительного диагноза после применения методики классификации рентгенограмм $\bar{P}(H_1)$ увеличивается на величину, равную сумме полных вероятностей $P(S_k, H_2)$, соответствующих той области сочетаний интервала изменения яркости и фрактальной размерности множества пикселей, для которых апостериорная вероятность положительного диагноза меньше порогового значения $P(H_1/S_k) > P_0$:

$$\bar{P}(H_1) = P(H_1) + \sum_{k=1}^l P(S_k, H_2).$$

Выразим априорную вероятность положительного диагноза после применения методики классификации рентгенограмм $\bar{P}(H_1)$ через условную вероятность $P(S_k/H_2)$ данного сочетания интервала изменения яркости и фрактальной размерности множества пикселей при условии, что верна гипотеза H_2 о том, что рентгенограмма содержит симптомы болезни, с помощью формулы, по которой определяется полная вероятность [9]:

$$\bar{P}(H_1) = P(H_1) + \sum_{k=1}^l P(H_2)P(S_k/H_2).$$

Вычислим меру априорной неопределенности или энтропии человека как объекта, который по результатам наблюдений может находиться в одном из двух состояний, а именно: «симптомы болезни не обнаружены» и «симптомы болезни обнаружены», – по формуле [9]:

$$H(X) = -(P(H_1) \log P(H_1) + P(H_2) \log P(H_2)).$$

Тогда количество информации о здоровье человека, приобретаемое в результате применения разработанной методики диагностики туберкулеза, определяется по формуле [9]:

$$I_X = H(X) - \bar{H}(X),$$

где $H(X)$ – энтропия пациента до применения методики диагностики туберкулеза; $\bar{H}(X)$ – энтропия пациента после применения методики диагностики туберкулеза.

Поскольку априорная вероятность положительного диагноза после применения методики классификации рентгенограмм больше априорной вероятности до применения методики классификации рентгенограмм $\bar{P}(H_1) > P(H_1)$, количество информации о пациенте является положительной величиной $I_X > 0$.

Следовательно, применение методики диагностики туберкулеза по рентгеновским снимкам позволяет уменьшить неопределенность состояния

здоровья человека, подвергающегося медицинскому обследованию. Это свидетельствует о том, что результаты, полученные в статье, целесообразно использовать в математическом и алгоритмическом обеспечении методики диагностики туберкулеза по рентгеновским снимкам для повышения точности определения болезни на ранней стадии ее развития. Это обеспечит снижение опасности, вызванной принятием решения о том, что рентгенограмма не содержит патогенных зон, в то время как рентгенограмма принадлежит человеку, больному туберкулезом.

Кроме того, ее применение позволит уменьшить расходы, связанные с принятием решения о том, что рентгенограмма содержит патогенные зоны, в то время как рентгенограмма принадлежит здоровому человеку.

Выводы

Предложен метод принятия решения о наличии у пациента диагноза, указывающего на его здоровье, по измеренному спектру фрактальных размерностей для фрактальной меры на множестве пикселей данной яркости. Предложенный метод позволяет повысить достоверность диагноза, основанного на анализе медицинского изображения. Направление дальнейших исследований связано с программной реализацией предложенного метода.

ЗАСТОСУВАННЯ ПРАВИЛА БАЙЕСА ДО КЛАСИФІКАЦІЇ МЕДИЧНИХ ЗОБРАЖЕНЬ

Д.О. Ніцин

Пропонується методика класифікації рентгенівських зображень за їх фрактальною розмірністю. Методика ґрунтується на оцінюванні ймовірності того, що при вимірюванні фрактальній розмірності зображення надана рентгенограма не містить патогенних ділянок.

Ключові слова: правило Байєса, фрактальна розмірність.

APPLICATION OF THE BAYESIAN RULE TO THE CLASSIFICATION OF MEDICAL IMAGES

D.A. Nitsyn

The technique of classification of the x-ray images till them fractal dimension is offered. The technique is based on an estimation of probability that at measured fractal dimension of the image given x-ray image does not contain pathological areas.

Keywords: the Bayesian rule, fractal dimension.

Список литературы

1. Постнова Т.Б. Информационно-диагностические системы в медицине / Т.Б. Постнова. – М.: Наука, 1972. – 376 с.
2. Завалишин Н.В. Модели зрительного восприятия и алгоритмы анализа изображений / Н.В. Завалишин, И.Б. Мучник. – М.: Наука, 1976. – 402 с.
3. Максимов Г.К. Статистическое моделирование многомерных систем в медицине / Г.К. Максимов, А.Н. Сеницын. – Л.: Медицина, 1983. – 144 с.
4. Федер Е. Фракталы: пер. с англ. / Е. Федер. – М.: Мир, 1991. – 254 с.
5. Кроновер Р. Фракталы и хаос в динамических системах / Р. Кроновер. – М.: Техносфера, 2006. – 488 с.
6. Бондаренко А.Н. Адаптивный двухступенчатый метод классификации изображений / А.Н. Бондаренко, А.В. Кацук // Искусственный интеллект. – Донецк: Институт проблем искусственного интеллекта МОН и НАН Украины. – 2006. – №4. – С. 676-680.
7. Венцель Е.С. Теория вероятностей / Е.С. Венцель. – М.: Наука, 1969. – 576 с.
8. Румишский Л.З. Математическая обработка результатов эксперимента / Л.З. Румишский. – М.: Наука, 1971. – 192 с.
9. Биргер И.А. Техническая диагностика / И.А. Биргер. – М.: Машиностроение, 1978. – 240 с.

Поступила в редколлегию 22.06.2011

Рецензент: д-р техн. наук, проф. П.А. Качанов, НТУ «Харьковский политехнический институт», Харьков.