

АППРОКСИМАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ КРИВЫМИ БЕЗЬЕ

Виталий Вишнеvский, Владимир Калмыков, Татьяна Романенко

Abstract: Very often the experimental data are the realization of the process, fully determined by some unknown function, being distorted by hindrances. Treatment and experimental data analysis are substantially facilitated, if these data to represent as analytical expression. The experimental data processing algorithm and the example of using this algorithm for spectrographic analysis of oncologic preparations of blood is represented in this article.

Ключевые слова: график, экспериментальные данные, кривые Безье

Введение

Очень часто экспериментальные данные являются искаженной помехами реализацией определенного процесса, вполне определяемого некоторой неизвестной функцией вида $y = f(x)$. В большинстве случаев экспериментальные данные такого рода представляют в виде графиков. Графики являются, видимо, самым простым и давно используемым в самых различных областях человеческой деятельности средством когнитивного представления экспериментальных данных, которые позволяют наглядно оценить качественные особенности процесса, несмотря на помехи, погрешности измерений. Графики, отображающие один и тот же процесс, характеристику некоего объекта могут существенно отличаться друг от друга масштабами, количеством использованных измерений, уровнем помех, и пр. В то же время особенности формы кривой графика характеризуют параметры отображаемого объекта или процесса.

Автоматическая (автоматизированная) обработка таких графиков предполагает сопоставление их форм с целью идентификации, чтобы установить, характеризуют ли различные графики одни и те же или различные процессы или объекты. Решение этой задачи средствами нейросетевого или статистического распознавания зрительных образов осложняется тем, что каждый график может отличаться от других масштабами, уровнем помех, количеством измерений. В то же время известен метод [1] нейросетевого распознавания контуров изображений, представленных кривыми Безье. Именно использование кривых Безье позволило обеспечить инвариантность описаний контуров изображений относительно положения изображений в поле зрения, масштаба, поворота. В таком случае сами исходные экспериментальные данные – графики могут быть заменены другими, аналитическими кривыми, хорошо совпадающими с формой кривой исходного графика, и, в то же время, инвариантными относительно изменения масштаба, количества измерений, уровня помех. Такие кривые могут быть заданы функциями $x(t)$, $y(t)$, которые представляют собой полиномы n -ной степени от параметра t :

$$\begin{aligned}x(t) &= a_0t^n + a_1t^{n-1} + \dots + a_{n-1}t + a_n, \\y(t) &= b_0t^n + b_1t^{n-1} + \dots + b_{n-1}t + b_n.\end{aligned}\tag{1}$$

Выбор n – степени полиномов зависит от сложности аппроксимируемых кривых.

Реализация такого подхода позволила бы отобразить все существенные особенности формы графика, исключив влияние помех. Кроме того, вместо описаний графиков в пространстве сигналов можно рассматривать аппроксимирующие их полиномы в пространстве их коэффициентов. Существенным преимуществом такого описания является его инвариантность относительно количества измеренных сигналов в каждом графике.

Однако, если прямая задача получения графика по заданным коэффициентам полиномов является тривиальной, то обратная задача – определение коэффициентов аппроксимирующих полиномов по точкам графика – не имеет простого решения [2]. Одно из решений данной задачи рассматривается в настоящей работе.

Формальная постановка задачи

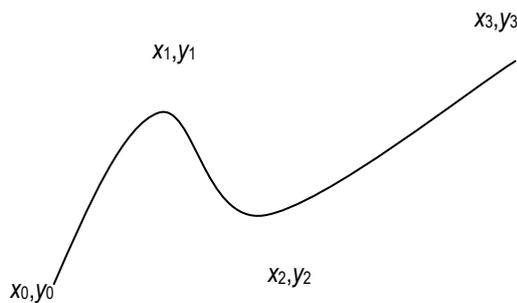


Рис. 1. Кривая Безье

График может быть представлен ломаной линией, точки излома которой образованы последовательностью N полученных экспериментально значений некоторой функции $y(x): y_{э}(x) = \{y(x_0), y(x_1), y(x_2), \dots, y(x_N)\}$, при заданных значениях $x_0, x_1, x_2, \dots, x_N$. Рассмотрим кривую Безье в качестве аппроксимирующей кривой как наиболее часто используемый в практических приложениях, параметрически заданный полином третьей степени.

Под кривой Безье понимаем дугу плоской кривой третьего порядка, заданную в параметрическом виде таким образом [3]:

$$Bz(t) = \sum_{j=0}^3 Br_j(t) \cdot Q_j \quad (2)$$

где $Br_j(t)$ – базовые полиномы Бернштейна третьей степени

Q_j – коэффициенты кривой

t – действительный параметр, который изменяется в интервале $[0, 1]$.

Полиномы Бернштейна определяются следующими формулами:

$$Br_j(t) = C_3^j \cdot (1-t)^{3-j} t^j, \quad 0 \leq j \leq 3 \quad (3)$$

где C_3^j , $0 \leq j \leq 3$ – биномиальные коэффициенты

$$C_3^j = \frac{3!}{j!(3-j)!} \quad (4)$$

Таким образом, кривая Безье (рис. 1) может быть задана двумя полиномами третьей степени, зависящими от параметра t :

$$\begin{aligned} X(t) &= x_0 \cdot (1-t)^3 + x_1 \cdot (1-t)^2 \cdot t + x_2 \cdot (1-t) \cdot t^2 + x_3 \cdot t^3 \\ Y(t) &= y_0 \cdot (1-t)^3 + y_1 \cdot (1-t)^2 \cdot t + y_2 \cdot (1-t) \cdot t^2 + y_3 \cdot t^3 \end{aligned} \quad (5)$$

Представление экспериментальных данных в виде кривой Безье имеет ряд преимуществ. Описывая достаточно широкий класс кривых, кривые Безье используют относительно небольшое количество коэффициентов – 8. Важным преимуществом с точки зрения пользователя является наглядность. Коэффициенты (x_0, y_0) , (x_3, y_3) являются координатами исходной и конечной точек дуги аппроксимирующей кривой, соответственно, а коэффициенты (x_1, y_1) , (x_2, y_2) – координатами точек, которые управляют ее формой (так называемых "усов"). Изменяя коэффициенты или перемещая соответствующие им точки на экране монитора (как начальную и конечную, так и управляющие), пользователь может в интерактивном

режиме редактировать форму аппроксимирующей кривой, улучшая результаты автоматической аппроксимации.

Таким образом, каждый график может быть представлен вектором $\mathbf{v} = \{x_0, y_0, x_1, y_1, x_2, y_2, x_3, y_3\}$, компонентами которого являются коэффициенты уравнений кривой Безье. Описание графика в виде вектора \mathbf{v} , всегда имеет одно и то же количество компонент независимо от числа измеренных экспериментально точек, что упрощает решение различных задач, в частности, задач распознавания графиков.

Как ранее было отмечено, определение коэффициентов кривой Безье, наиболее точно воспроизводящей тот или иной график, не является тривиальной задачей. В работе [2] для определения коэффициентов кривой Безье, аппроксимирующей последовательность экспериментальных точек, использован итерационный способ минимизации суммы квадратов расстояний от каждой точки графика до аппроксимирующей кривой. Такая постановка обусловлена тем, что рассматриваются произвольные графики, а не только те, которые отображают некоторую неизвестную функцию. Допускается, что график представляет собой совокупность несвязанных, взаимно независимых точек. Для вычисления расстояния от каждой точки графика до аппроксимирующей кривой должна быть найдена соответствующая ближайшая точка на аппроксимирующей кривой, что само по себе при параметрическом задании кривой является итерационной задачей. При значительном количестве точек на графике и заранее не определенном количестве итераций такой способ решения задачи привел к существенным вычислительным трудностям.

В данной работе предлагается принять в качестве меры сходства данного экспериментального графика с некоторой аппроксимирующей кривой площадь $S(\mathbf{v}, y(x))$ фигуры, ограниченной контуром, образованным упомянутыми кривой, определяемой коэффициентами \mathbf{v} , и экспериментальной кривой, образованной точками графика, отображающего некоторую неизвестную функцию $y(x)$. Тогда поиск кривой Безье с коэффициентами \mathbf{v}_{opt} , которая наилучшим образом аппроксимирует заданный экспериментальный график, можно рассматривать как определение коэффициентов кривой Безье, которые минимизируют значение $S(\mathbf{v}, y_3(x))$:

$$\mathbf{v}_{opt} = \underset{\mathbf{v}}{\operatorname{argmin}} S(\mathbf{v}, y_3(x)) \quad (6)$$

Алгоритм

Как видно из формулы 6, основой алгоритма поиска оптимального значения коэффициентов \mathbf{v}_{opt} является вычисление площади, ограниченной контуром $S(\mathbf{v}, y_3(x))$. Непосредственное вычисление $S(\mathbf{v}, y_3(x))$ не представляется возможным из-за того, что данный контур не является односвязным.

График $y_3(x)$ и кривая Безье с коэффициентами \mathbf{v} неоднократно пересекаются. Вычисление координат точек пересечения затруднено параметрическим заданием кривой Безье. Наиболее простым решением представляется замена кривой Безье ломаной линией V_t , точки излома которой (x_t, y_t) принадлежат кривой Безье и вычислены для значений $t = 0, \delta, 2\delta, 3\delta, \dots, 1$. Значение δ выбрано достаточно малым, чтобы обеспечить достаточно точное для практических вычислений отображение ломаной линией V_t исходной кривой Безье. Результатом принятого упрощения является возможность вычислить для каждого значения $y(x)$ соответствующее значение $V_t(n)$. Тогда площадь $S(\mathbf{v}, y_3(x))$ может быть вычислена как

$$S(\mathbf{v}, y_3(x)) = \sum_{n=0}^N |V_t(x_n) - y_3(x_n)| \quad (7)$$

Значение v_{opt} , соответствующее минимальному значению $S(v, D(n))$, может быть получено путем использования метода градиентного спуска.

Экспериментальная проверка алгоритма

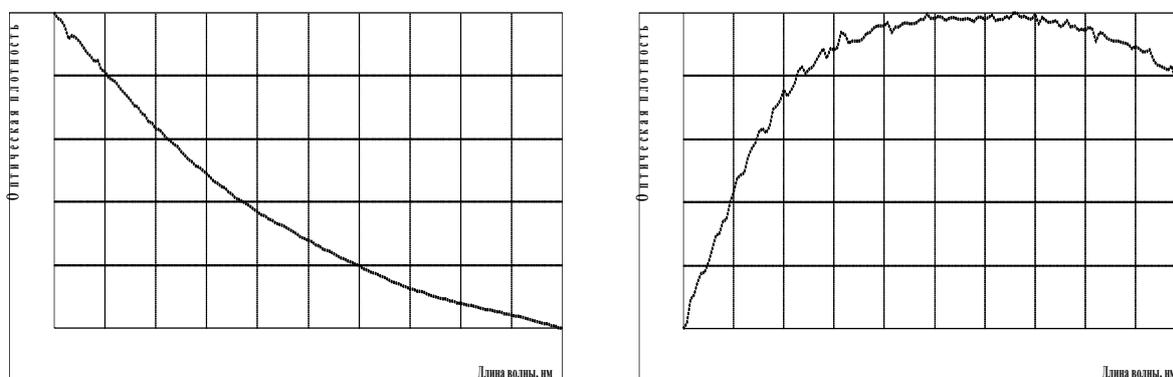


Рис.2. Примеры спектрограмм препарата сыворотки крови:
слева – спектрограмма здорового, справа – больного человека.

Предложенный алгоритм был экспериментально проверен более чем на 1100 реальных графиках спектрограмм медицинских препаратов крови, полученных по методу Онкотест [4]. Метод ранней диагностики онкологических заболеваний позволяет определить наличие или отсутствие онкологического заболевания у пациента по виду спектрограмм медицинских препаратов крови (рис.2). В настоящее время проводятся исследования для определения, при условии наличия онкологического заболевания, локализации заболевания. На рис.3 показаны спектрограммы препарата крови больных онкологическим

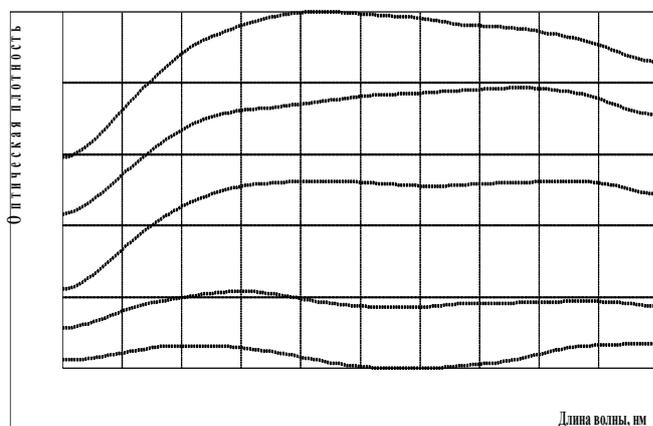


Рис.3 Примеры спектрограмм препарата крови
больных онкологическим заболеванием одинаковой
локализации (желудок).

заболеванием одинаковой локализации (желудок). Несмотря на очевидные различия спектрограмм по уровням сигналов, помех, формы кривых имеют несомненное сходство. Специфическим свойством приведенных графиков является форма соответствующих им кривых, которая может соответствовать диагнозу. Решение этой задачи предполагает дополнительные исследования, которые должны основываться на обработке большого количества экспериментальных данных. В свою очередь, обработка больших массивов данных невозможна без решения задач их компьютерной обработки, в числе которых

обязательно должны быть решены задачи автоматической или автоматизированной классификации спектрограмм по их форме с использованием методов распознавания образов.

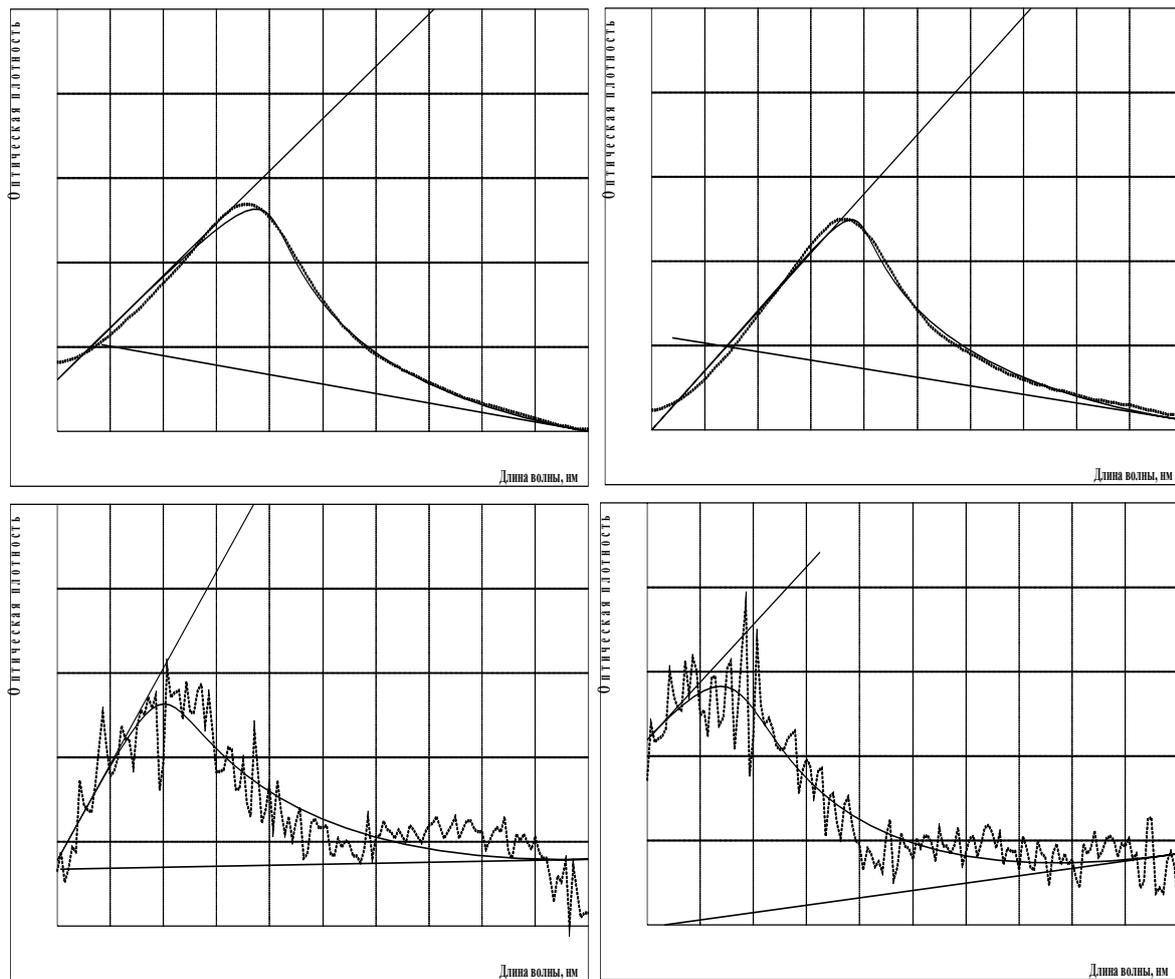


Рис.4 Спектрограммы препаратов крови и их аппроксимация кривыми Безье; локализация – печень

Непосредственное использование статистических методов распознавания для автоматизированной постановки диагноза в данном случае не представляется возможным из-за большой изменчивости значений масштаба (десятки раз), и, также, конфигурации функции реализующей один и тот же образ. Поэтому представилось целесообразным для каждого графика найти некоторую подходящую кривую Безье, полиномиальные коэффициенты которой, после приведения к одному масштабу, могут быть использованы как параметры образа, инвариантные к изменениям масштаба, количества измерений, уровня помех. Примеры спектрограмм и аппроксимирующие их кривые Безье приведены на рис. 5,6. Результат достигался в большинстве случаев за 15 - 40 итераций (в особенно сложных случаях - до 300). Как видно из рисунков, аппроксимирующие кривые Безье дают удовлетворительное приближение.

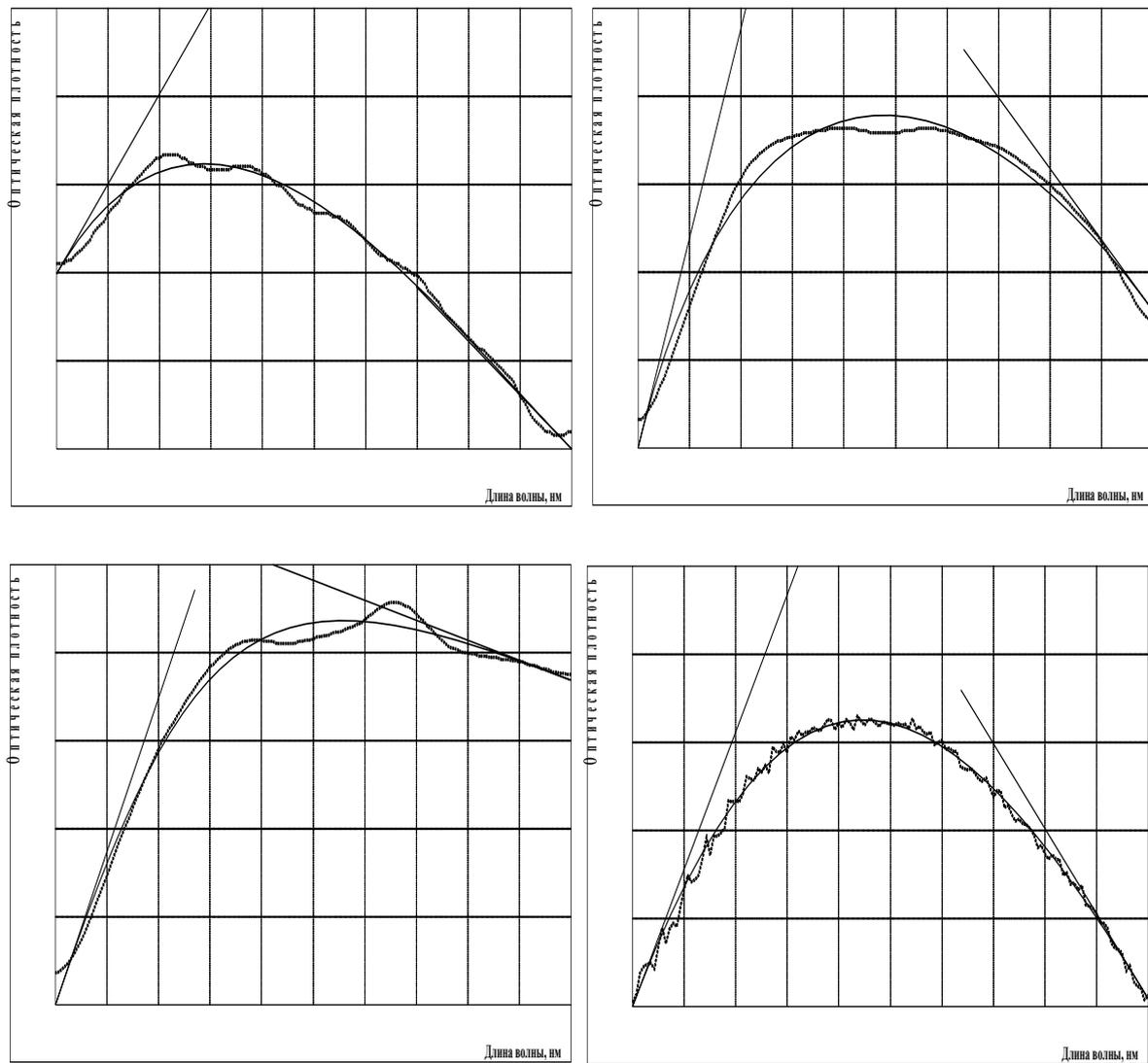


Рис.5 Спектрограммы препаратов крови и их аппроксимация кривыми Безье;
локализация – молочная железа

Заключение

Предложенный алгоритм позволяет найти параметры кривой Безье, которая аппроксимирует экспериментальный график, с целью автоматической классификации. Приведены примеры аппроксимации спектрограмм с целью выполнения автоматизированной диагностики. Это даст в будущем возможность автоматизировать процесс обследования населения с целью раннего выявления злокачественных опухолей.

Литература

- [1] Кодирование контуров, представленных кривыми Безье, в задачах нейросетевой классификации / Куссуль М.Э. // Математические машины и системы. – 2004. – № 3. – С.17-30.

-
- [2] Вишневский В.В., Рысцов I.K., Волжева М.В. Итерационный алгоритм построения кривой Безье по заданными точками //Математические машины и системы. - 2004 №4 - С. 108 – 116.
- [3] Денискин Ю.И. Особенности аппроксимации обводов параметрическими полиномами в форме Бернштейна //Прикладная геометрия. - 1999. - вып. 2 № 2
- [4] Вишневский В.В., Владимиров В.А., Романенко Т.Н. Программно-аппаратный комплекс "Онкотест-WM-01" // Тезисы доклада II Международной школы-семинара "Телемедицина: опыт и перспективы".
-

Информация об авторах

Виталий Вишневский – заведующий отделением, старший научный сотрудник, кандидат технических наук, Институт проблем математических машин и систем, просп. акад. Глушкова 42, 03680, Киев 187, Украина; e-mail: vit@immsp.kiev.ua

Владимир Калмыков – старший научный сотрудник, кандидат технических наук, Институт проблем математических машин и систем, просп. акад. Глушкова 42, 03680, Киев 187, Украина; e-mail: kvg@immsp.kiev.ua

Татьяна Романенко – младший научный сотрудник, Институт проблем математических машин и систем, просп. акад. Глушкова 42, 03680, Киев 187, Украина; e-mail: romanenko@immsp.kiev.ua