

АДАПТИВНОЕ F-ПРЕОБРАЗОВАНИЕ НА ОСНОВЕ ОБОБЩЕННОЙ РЕГРЕССИОННОЙ НЕЙРОННОЙ СЕТИ

1 Введение

В задачах обработки информации, связанных с построением аппроксимирующих моделей, наряду с традиционными методами регрессионного анализа и аппаратом различных преобразований (Фурье, Лапласа и т.п.) в настоящее время широкое распространение получили подходы, основанные на вычислительном интеллекте и, прежде всего, нейро-фаззи системы. Появившийся в последние годы аппарат F-преобразования [1] перебрасывает мостик между традиционными преобразованиями и системами нечеткого вывода, а в перспективе и с более мощными средствами вычислительного интеллекта.

В общем случае F-преобразование является средством представления бесконечномерного функционального пространства с помощью конечномерного векторного пространства путем установления определенного соответствия между множеством непрерывных вещественных функций, заданных на ограниченном интервале, и множеством векторов. Естественно, что обратное F-преобразование переводит полученные векторы в другие непрерывные функции, аппроксимирующие исходные.

Ключевая идея этого преобразования состоит в нечетком (фаззи-) разбиении исходного пространства на множество пересекающихся нечетких подпространств (кластеров) и локальной аппроксимации исходной функции в этих подпространствах с помощью средств нечеткого вывода.

Пусть преобразуемая функция $y = f(x)$ задана на ограниченном интервале $[a, b]$, на котором также задан набор фиксированных узлов $a \leq c_1 < c_2 < \dots < c_j < \dots < c_h \leq b$. С каждым из узлов связана функция принадлежности $\mu_j(x)$ так, что их полный набор $\mu_1(x), \dots, \mu_j(x), \dots, \mu_h(x)$, который также может быть задан в векторной форме $\mu^h = (\mu_1(x), \dots, \mu_j(x), \dots, \mu_h(x))^T$, формирует нечеткое разбиение интервала $[a, b]$. В теории F-преобразования функции принадлежности $\mu_j(x)$ удовлетворяющие требованиям:

$$\mu_j(x) : [a, b] \rightarrow [0, 1], \mu_j(c_j) = 1;$$

$\mu_j(x) = 0$, если $x \notin (c_{j-1}, c_j)$, т.е. функции принадлежности имеют компактное основание;

$$\mu_j(x) \text{ непрерывны};$$

$$\mu_j(x) \text{ возрастают на интервале } [c_{j-1}, c_j] \text{ и убывают на } [c_j, c_{j+1}];$$

и, наконец,

© Е.В. Бодянский, Н.А. Тесленко, 2007

$$\sum_{j=1}^h \mu_j(x) = 1, \tag{1}$$

называются базисными.

Тогда набор вещественных чисел $F^h = (F_1, \dots, F_j, \dots, F_h)^T$ таких, что

$$F_j = \frac{\int_a^b f(x)\mu_j(x)dx}{\int_a^b \mu_j(x)dx}, j = 1, 2, \dots, h, \tag{2}$$

называется интегральным F-преобразованием функции $f(x)$ относительно $\mu_1(x), \dots, \mu_j(x), \dots, \mu_h(x)$.

Поскольку в задачах, связанных с обработкой информации, функция $f(x)$ задается, как правило, в отдельных точках $x(1), y(1), x(2), y(2), \dots, x(k), y(k), \dots, x(l), y(l), h \leq l$, более “практичным” представляется использование дискретного F-преобразования в форме

$$F_j(l) = \frac{\sum_{k=1}^l y(k)\mu_j(x(k))}{\sum_{k=1}^l \mu_j(x(k))} = \frac{\sum_{k=1}^l f(x(k))\mu_j(x(k))}{\sum_{k=1}^l \mu_j(x(k))}, j = 1, 2, \dots, h, \tag{3}$$

при этом можно заметить, что $F_j(l)$ доставляет минимум критерию

$$E_j^l = \sum_{k=1}^l (f(x(k)) - F_j)^2 \mu_j(x(k)). \tag{4}$$

В простейшем случае при неравномерном расположении узлов-центров c_j в качестве функций принадлежности можно использовать треугольные конструкции

$$\begin{cases} \mu_j^R(x) = \frac{c_{j+1}-x}{c_{j+1}-c_j}, \\ \mu_{j+1}^L(x) = \frac{x-c_j}{c_{j+1}-c_j}, \\ x \in [c_j, c_{j+1}], j = 1, 2, \dots, h-1, \end{cases} \tag{5}$$

вид которых приведен на рис. 1.

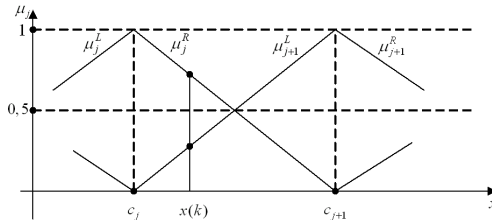


Рис. 1 – Базисные функции принадлежности.

Обратное F-преобразование определяется выражением

$$\widehat{y}_F(k) = f_{F,h}(x(k)) = \sum_{j=1}^h F_j(l)\mu_j(x(k)) = (F^h(l))^T \mu^h(x(k)), \quad (6)$$

при этом существует такое число узлов h_ε , что

$$|f(x(k)) - f_{F,h_\varepsilon}(x(k))| < \varepsilon, \quad (7)$$

(здесь ε – произвольное малое положительное число), т.е. обратное F-преобразование сколь угодно точно аппроксимирует исходную функцию.

Заметим также, что в F-преобразовании расположение узлов c_j никак не связано с расположением наблюдений $x(k)$, а структура самого преобразования, если рассматривать её с позиции обучения нейронных сетей, может быть отнесена к системам, основанным на оптимизации.

Альтернативой системам, обучение которых основано на оптимизации, являются системы, основанные на памяти, обучаемые за один проход и реализующие концепцию “нейроны в точках данных”. Наиболее типичным представителем таких систем является обобщенная регрессионная нейронная сеть (GRNN) [2], широко применяемая в задаче идентификации нелинейных объектов, представляющей собой по сути ту же задачу аппроксимации функций по “зашумленным” данным.

В основе обобщенных регрессионных нейронных сетей лежит идея непараметрического оценивания с помощью окон Парзена [3], при этом для обучающей выборки $\{x(k), y(k)\}$, $k = 1, 2, \dots, l$ и наиболее популярных ядер Гаусса сеть реализует отображение

$$\widehat{y}_G(x) = \frac{\sum_{k=1}^l y(k) \exp\left(-\frac{D^2(k)}{2\sigma^2}\right)}{\sum_{k=1}^l \exp\left(-\frac{D^2(k)}{2\sigma^2}\right)}, \quad (8)$$

где $D(k) = |x - x(k)|$, σ – параметр ширины, выбираемый из эмпирических соображений (чаще всего методом проб и ошибок).

Поскольку вместо гауссианов сети могут быть использованы ядерные функции достаточно общего вида $\varphi(D(k), \sigma(k))$, отображение (8) можно переписать в обобщенной форме

$$\widehat{y}_G(x) = \frac{\sum_{k=1}^l y(k)\varphi(D(k), \sigma(k))}{\sum_{k=1}^l \varphi(D(k), \sigma(k))}, \quad (9)$$

где функции $\varphi(\cdot)$ могут иметь разные параметры ширины. Обучение GRNN реализуется крайне просто и состоит в однократной установке центров ядерных функций $\varphi(\cdot)$ в точках наблюдений $x(k)$ (нейроны в точках данных). Как отмечено в [2], основными преимуществами GRNN

являются быстрее обучение и сходимость к оптимальной нелинейной кривой регрессии.

В случае, когда объем обучающей выборки l велик, размещение нейронов в точках данных чрезмерно усложняет сеть и может вести к проклятию размерности. В этом случае, центры ядерных функций $\varphi(\cdot)$ предлагается размещать в прототипах кластеров, каждый из которых характеризуется некоторым радиусом влияния r . При этом число кластеров h может быть много меньше l .

Процесс обучения GRNN в этом случае включает в себя и уточнение параметров кластеров в форме

$$\begin{cases} a_j(k) = a_j(k-1) + y(k), j = 1, 2, \dots, h; k = 1, 2, \dots, l, \\ b_j(k) = b_j(k-1) + 1, \\ c_j(k) = \frac{k-1}{k}c_j(k-1) + \frac{1}{k}x(k), \text{ если } |x(k) - c_j(k-1)| \leq r, \end{cases} \quad (10)$$

$$\begin{cases} a_j(k) = a_j(k-1), \\ b_j(k) = b_j(k-1), \\ c_j(k) = c_j(k-1), \text{ если } |x(k) - c_j(k-1)| > r, \end{cases} \quad (11)$$

Если же характеристики аппроксимируемой функции меняются со временем, то в процедуру (10) можно ввести экспоненциальное забывание в виде

$$\begin{cases} a_j(k) = \frac{\tau-1}{\tau}a_j(k-1) + \frac{1}{\tau}y(k), j = 1, 2, \dots, h; k = 1, 2, \dots, l, \\ b_j(k) = \frac{\tau-1}{\tau}b_j(k-1) + \frac{1}{\tau}, \\ c_j(k) = \frac{\tau-1}{\tau}c_j(k-1) + \frac{1}{\tau}x(k), \text{ если } |x(k) - c_j(k-1)| \leq r, \end{cases} \quad (12)$$

где τ – постоянная времени экспоненциально затухающей функции. Процедура (11) остается без изменений.

При этом обобщенная регрессионная нейронная сеть в результате обучения реализует отображение

$$\widehat{y}_G(x) = \frac{\sum_{j=1}^h a_j(l)\varphi_j(D_j, \sigma_j)}{\sum_{j=1}^h b_j(l)\varphi_j(D_j, \sigma_j)} = \frac{\sum_{j=1}^h a_j\varphi_j(D_j, \sigma_j)}{\sum_{j=1}^h b_j\varphi_j(D_j, \sigma_j)} \quad (13)$$

(здесь $D_j = |x_i - c_j|$, σ_j – параметр ширины ядерной функции $\varphi(\cdot)$ j -го кластера), а её архитектура для скалярных входного и выходного сигналов приведена на рис. 2.

При подаче на вход сети сигнала $x(p)$, не принадлежащего к обучающей выборке, на её выходе появляется отклик

$$\widehat{y}_G(x(p)) = \widehat{y}_G(p) = \frac{\sum_{j=1}^h a_j\varphi_j(D_j(p), \sigma_j)}{\sum_{j=1}^h b_j\varphi_j(D_j(p), \sigma_j)}, \quad (14)$$

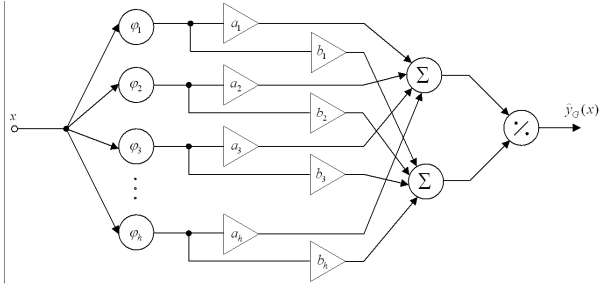


Рис. 2 – Архитектура обобщенной регрессионной нейронной сети с одним входом и одним выходом

(здесь $D_j(p) = |x(p) - c_j|$), принадлежащий оптимальной регрессионной кривой, аппроксимирующей исходную функцию.

Немаловажным также является то, что GRNN может обучаться в реальном времени, последовательно обрабатывая поступающие данные $x(k), y(k), k = 1, 2, \dots, l, \dots$

2 Адаптивное F-преобразование

Введем в рассмотрение архитектуру, реализующую обратное F-преобразование (6), как показано на рис. 3.

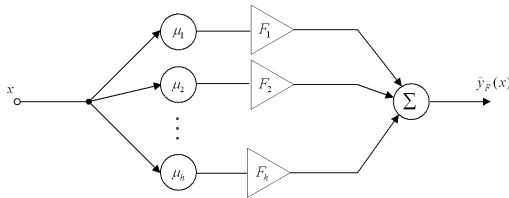


Рис. 3 – Архитектура обратного F-преобразования.

При подаче на вход этой архитектуры сигнала $x(p)$, на её выходе появляется отклик

$$\widehat{y}_F(x(p)) = \widehat{y}_F(p) = \sum_{j=1}^h F_j(l) \mu_j(x(p)), \quad (15)$$

аппроксимирующий исходную функцию, но, естественно, отличный от (14). Вместе с тем можно заметить, что если вместо выражения (10) для уточнения параметров кластеров использовать соотношения

$$\begin{cases} a_j(k) = \frac{k-1}{k} a_j(k-1) + \frac{1}{k} y(k), & j = 1, 2, \dots, h; k = 1, 2, \dots, l, \\ b_j(k) = \frac{k-1}{k} b_j(k-1) + \frac{1}{k}, \\ c_j(k) = \frac{k-1}{k} c_j(k-1) + \frac{1}{k} x(k), & \text{если } |x(k) - c_j(k-1)| \leq r, \end{cases} \quad (16)$$

а вместо ядерных функций $\varphi(\cdot)$ – базисные функции принадлежности (5), с учетом (1) преобразования (14) и (15) становятся структурно эквивалентными, хотя первое из них получено в результате применения процедуры, основанной на памяти, а второе – в результате решения задачи оптимизации.

Данное подобие позволяет построить адаптивную процедуру вычисления дискретного F-преобразования, в которой компоненты F_j уточняются в реальном времени по мере поступления данных с учетом возможной нестационарности функций $f(x)$. Исходными данными при этом являются требуемая точность аппроксимации ε и радиус влияния r .

После поступления первых двух точек обучающей выборки $x(1), y(1), x(2), y(2)$ (без потери общности положим $x(1) < x(2)$) полагаются, что $x(1) = c_1, x(2) = c_2$ и формируются первые две базисные функции $\mu_1^R(x)$ и $\mu_2^L(x)$. Следующая точка $x(3)$ проверяется на принадлежность зоне влияния c_1 и c_2 . Если эта точка находится вне зон влияния c_1 и c_2 ($|x(3) - c_1| > r, |x(3) - c_2| > r$), формируется третья базисная функция $\mu_3(x)$, если же $x(3)$ попадает в зону влияния одного из центров $c_j, (j = 1, 2)$, происходит уточнение его положения согласно выражению

$$c_j(3) = (1 - \alpha)c_j(2) + \alpha x(3), \quad (17)$$

или для k -й итерации

$$c_j(k) = (1 - \alpha)c_j(k - 1) + \alpha x(k), \quad (18)$$

где $0 \leq \alpha \leq 1$ – параметр сглаживания.

Этот процесс продолжается либо до исчерпания обучающей выборки, либо происходит непрерывно в реальном времени. В результате такого обучения будет сформировано

$$\frac{b - a}{2r} \leq h \leq \frac{b - a}{2r} + 1 \quad (19)$$

прототипов-центров кластеров c_j , относительно которых согласно (5) строятся базисные функции μ_j .

По исчерпании обучающей выборки производится проверка условия (7) и, если неравенство удовлетворяется, полагается, что r , а следовательно и h выбраны верно. Если же требуемая точность не обеспечивается, уменьшается радиус влияния r и настройка производится при большем числе кластеров h . Заметим также, что при подобном подходе не обязательно задавать интервал $[a, b]$, а вместо него использовать минимальное и максимальное значение из обучающей выборки $x(k)$.

Подобный подход позволяет связать расположение узлов F-преобразования c_j с характером аппроксимируемой функции и изменять их количество h в зависимости от требуемой точности ε .

3 Выводы

В статье установлено структурное подобие между дискретным F-преобразованием и обобщенной регрессионной нейронной сетью. Предло-

жена адаптивная модификация F-преобразования, позволяющая последовательно обрабатывать данные и изменять количество узлов по ходу процесса обучения.

Литература

1. I. Perfilieva. Fuzzy transforms: Theory and Applications// Fuzzy Sets and Systems -2006.-157.- P993-1023.
2. Specht D.E. A general regression neural network//IEEE Trans. on Neural Networks.-1991.-2.-P568-576.
3. Parzen E. On the estimation of a probability density function and the mode//Ann. Math. Stat.-1962.-38.-P1065-1076.