УДК 004.032.26

Е.В. Бодянский[1], Е.А. Винокурова[1], Д.Д. Пелешко[2], И.О. Кобылин[1], О.А. Кобылин[1]

[1] *Харьковский национальный университет радиоэлектроники, Харьков*
[2] *ПВНЗ «Компьютерная академия ШАГ», Львов*

НЕЧЁТКАЯ КЛАСТЕРИЗАЦИЯ ВРЕМЕННЫХ РЯДОВ С НЕРАВНОМЕРНЫМИ И АСИНХРОННЫМИ ТАКТАМИ КВАНТОВАНИЯ

В статье предложен алгоритм нечеткой кластеризации временных рядов с неравномерным асинхронными тактами квантования (биомедицинские массивы наблюдений, сигналы цифрового видео, формирующие дискретные двумерные поля и т.д.). Этот алгоритм характеризуется простотой вычислительной реализации, высокими аппроксимируемыми свойствами, быстродействием процесса обучения и предназначен для решения широкого класса задач, в том числе, когда исходные данные имеют высокую размерность.

***Ключевые слова**: нечеткая кластеризация временных рядов, асинхронность, квантование, data mining, адаптивные процедуры обучения. online процедура нечеткой кластеризации.*

Введение

Задача кластеризации временных рядов занимает достаточно важное место в теории и практике Data Mining [1–3], а для ее решения к настоящему времени предложено множество алгоритмов. Вместе с тем, во множестве реальных приложений возникают ситуации, когда стандартные методы оказываются либо неэффективными, либо вообще неработоспособными. Здесь, прежде всего, следует отметить задачи, когда формируемые кластеры взаимно перекрываются, что требует применения аппарата нечеткой кластеризации [4–5]. При этом следует помнить, что применяемость стандартных алгоритмов нечеткой кластеризации ограничивается эффектом «концентрации норм» [6], что позволяет обрабатывать лишь достаточно короткие выборки. Ситуация усложняется, если наблюдения поступают на не равностоящие моменты времени, что не позволяет использовать метрики, применяемые в стандартном нечетком кластерном анализе. В связи с этим в [7] был предложен метод нечеткой кластеризации коротких временных рядов с неравномерным тактом квантования, в основе которого лежат так называемое, PS-расстояние (Piecewise slope distance=PS distance=STS distance=short time series distance) и стандартный метод нечетких С-средних (FCM) Дж. Бездека [5]. В [8] на основе PS-расстояния и процедуры, введенной в [9], был предложен алгоритм нечеткой кластеризации коротких временных рядов с произвольными наблюдениями, а в [10] на основе PS-расстояния и рекуррентных алгоритмов нечеткой кластеризации [11–12] – online процедура нечеткой кластеризации коротких временных рядов, являющийся по сути нечеткой модификацией WTM-правила самообучения Т. Кохонена [13].

Ситуация существенно усложняется, если квантование производится не только неравномерно, но и отличается для каждой из реализаций наблюдаемого сигнала. Именно в этом случае на первый план выходит «концентрация норм», игнорирования которой не позволит получить приемлемое решение. В связи с этим представляется целесообразным распространение подхода, введенного в [7], на ситуацию, когда информация на обработку подается в форме выборок с неравномерными и асинхронными тактами квантования.

Оценка расстояния между реализациями временного ряда

Пусть исходная информация задана в форме набора выборок $x_{i(k)}(k)$ (здесь $i(k)=1,2,\ldots,n(k)-$ номер определенного наблюдения в k-й реализации, при этом каждая новая реализация может содержать разное число наблюдений $n(k), k=1,2,\ldots,N$), содержащего N последовательностей с неравномерным тактом квантования, подлежащих нечеткой кластеризации, при этом каждая выборка может быть представлена в форме $(n(k)\times 1)-$ вектора $x(k)=\left(x_1(k),\ x_2(k),\ \ ...,x_{n(k)}(k)\right)^T$. Заметим также, что каждое наблюдение $x_{i(k)}(k)$ наблюдается в момент времени $0\le t_{i(k)}(k)\le T$. Понятно, что два вектора–выборки $x_{i(k)}\in R^{n(k)}$ и $x(l)\in R^{n(l)}$ при

$n(k) \neq n(l)$ в принципе несравнимы. Неравномерность же квантования означает то, что $\Delta t_{i(k)} = t_{i(k)} - t_{i-1(k)} \neq \Delta t_{i+1(k)} = t_{i+1(k)} - t_{i(k)}$, т.е. $\Delta t_{i(k)} \neq \text{const}$. Кроме того, в общем случае $t_{i(k)} \neq \Delta t_{i(l)}$. Рассматриваемая ситуация иллюстрируется рис. 1, где проведены реализации

$$x(1) = \left(x_1(1), x_2(1), ..., x_{n(1)}(1)\right)^T,$$

$$x(k) = x_1(k), x_2(k), ..., x_{n(k)}(k))^T$$

и $$x(N) = x_1(N), x_2(N), ..., x_{n(N)}(N))^T,$$

при этом в общем случае $t_1(1) \neq t_1(k) \neq t_1(N)$ и $t_{n(1)}(1) \neq t_{n(k)}(k) \neq t_{n(N)}(N)$.

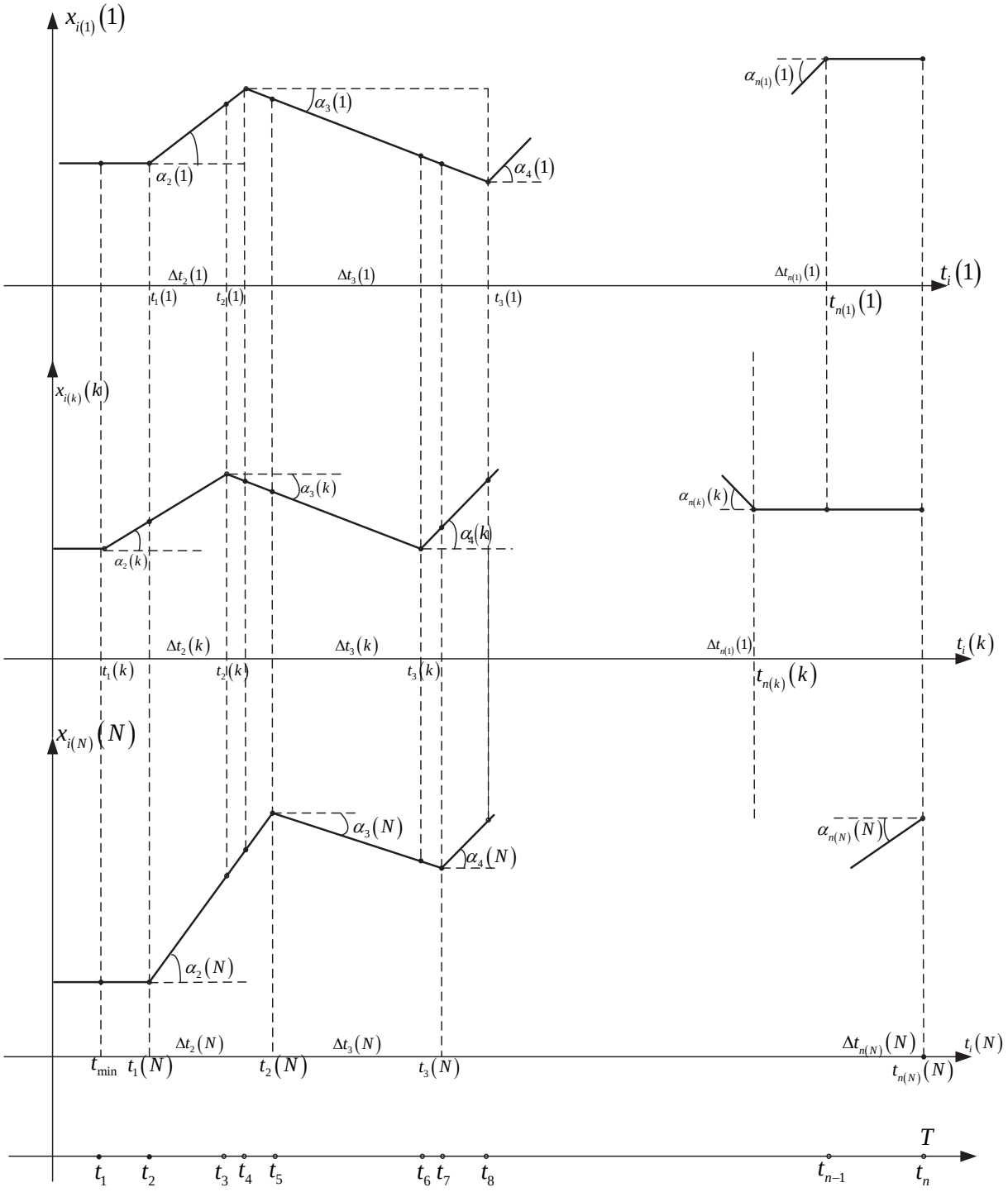

Рис. 1. Временные ряды с неравномерными асинхронными тактами квантования

Интервал наблюдения всего набора данных может быть задан в виде $\left[t_1 = t_{1\min} = \min\left\{t_1(k)\right\} - t_n = T = \max\left\{t_{n(k)}(k)\right\}\right]$.

Для оценки расстояния между выборками используем модифицированное PS-расстояние, основанное на представлении исследуемых рядов в виде кусочка линейных функций:

$$x_t(k) = a_t(k) + b_t(k)t, \tag{1}$$

где $t_i(k) \le t \le t_{i+1}(k)$,

$$\begin{cases} a_t(k) = \dfrac{t_{i+1}(k)x_{i(k)}(k) - t_i(k)x_{i+1(k)}(k)}{t_{i+1}(k) - t_i(k)}, \\ b_t(k) = \dfrac{x_{i+1}(k) - x_i(k)}{t_{i+1}(k) - t_i(k)} \end{cases} \tag{2}$$

и фактически оценивающее отличие форм анализируемых выборок. Соотношения (1–2) были использованы в [7] для оценки расстояния между выборками с неравномерными тактами квантования, однако синхронизированными во времени.

В случае же асинхронных выборок в рассмотрение можно ввести квазинаблюдения, полученные с помощью выражений (1–2). Так, возвращаясь к рис. 1, можно записать квазинаблюдения последовательностью $x(1)$ и $x(N)$ в момент времени $t_2(k)$ в виде

$$\begin{cases} \hat{x}_{t_2(k)}(1) = a_{t_2(k)}(1) + b_{t_2(k)}(1)t_2(k), \\ a_{t_2(k)}(1) = \dfrac{t_2(1)x_2(1) - t_1(1)x_1(1)}{t_2(1) - t_1(1)}, \\ b_{t_2(k)}(1) = \dfrac{x_2(1) - x_1(1)}{t_2(1) - t_1(1)} \end{cases} \tag{3}$$

и

$$\begin{cases} \hat{x}_{t_2(k)}(N) = a_{t_2(k)}(N) + b_{t_2(k)}(N)t_2(k), \\ a_{t_2(k)}(N) = \dfrac{t_2(N)x_2(N) - t_1(N)x_1(N)}{t_2(N) - t_1(N)}, \\ b_{t_2(k)}(N) = \dfrac{x_2(N) - x_1(N)}{t_2(N) - t_1(N)}. \end{cases} \tag{4}$$

Совершенно аналогично можно ввести квазинаблюдения $\hat{x}(k)$ на основании реальных наблюдений рядов $x(1),...,x(k-1),x(k+1),...\ ...,x(N)$.

В результате формируется общая временная шкала (нижняя ось рис. 1), содержащая $n(1) = ... = n(k) = ... = n(N) \ne n$ моментов в случае полностью синхронизированных выборок и $n = \sum_{k=1}^{N} n(k)$ точек, если моменты фиксации данных во всех рядах полностью не совпадают.

На основе этой оси времени может быть сформирован набор из N векторов-выборок-

$$\hat{x}(1) = \left(\hat{x}_1(1), \hat{x}_2(1), ..., \hat{x}_i(1), ..., \hat{x}_n(1)\right)^T, ...,$$

$$\hat{x}(k) = \left(\hat{x}_1(k), ..., \hat{x}_i(k), ..., \hat{x}_n(k)\right)^T, ...,$$

$$\hat{x}(N) = \left(\hat{x}_1(N), ..., \hat{x}_i(N), ..., \hat{x}_n(N)\right)^T$$

имеющих одинаковую размерность $(n \times 1)$, при этом в качестве компонентов этих векторов $\hat{x}_i(k)$ могут быть как реальные наблюдения, так и квазинаблюдения типа (3) и (4).

Далее несложно ввести расстояние между любыми двумя векторами $\hat{x}(k)$ и $\hat{x}(l)$ в виде [7]:

$$\begin{aligned} &d^2_{STS}\left(\hat{x}(k), \hat{x}(l)\right) = \\ &= \sum_{i=1}^{n-1}\left(\frac{\hat{x}_{i+1}(k) - \hat{x}_i(k)}{t_{i+1} - t_i} - \frac{\hat{x}_{i+1}(l) - \hat{x}_i(l)}{t_{i+1} - t_i}\right)^2 = \\ &= \sum_{i=1}^{n-1}\left(\frac{\hat{x}_{i+1}(k) - \hat{x}_i(k)}{\Delta t_{i+1}} - \frac{\hat{x}_{i+1}(l) - \hat{x}_i(l)}{\Delta t_{i+1}}\right)^2, \end{aligned} \tag{5}$$

являющихся по сути стандартной эвклидовой метрикой.

На основании метрики (5) авторами [7] была введена процедура нечетной кластеризации, являющаяся несколько громоздкой модификацией алгоритма нечетных C-средних (FCM) для обработки временных рядов с неравноотстоящими наблюдениями, а в [10] – её online версия.

Как видно, компоненты выражения (5) являются по сути первыми разностями сигналов $\hat{x}_i(k)$ и $\hat{x}_i(l)$ или тангенсами углов наклона линейных функций типа (1), т.е.

$$\Delta\hat{x}_{i+1}(k) = \frac{\hat{x}_{i+1}(k) - \hat{x}_i(k)}{\Delta t_{i+1}} = \mathrm{tg}\alpha_{i+1}(k),$$

$$\Delta\hat{x}_{i+1}(l) = \frac{\hat{x}_{i+1}(l) - \hat{x}_i(l)}{\Delta t_{i+1}} = \mathrm{tg}\alpha_{i+1}(l),$$

при этом ряд, образованный первыми разностями, содержит не n, а $n-1$ наблюдение (квазинаблюдение) $\Delta\hat{x}_2(k), \Delta\hat{x}_3(k), ..., \Delta\hat{x}_n(k)$, или $\mathrm{tg}\alpha_2(k), \mathrm{tg}\alpha_2(k), ..., \mathrm{tg}\alpha_n(k)$. Поскольку в результате операции взятия разностей из каждого ряда удаляется его среднее значение, для восстановления исходных рядов по их разностям необходимо дополнить разности любым из наблюдений исходных последовательностей, например, $\hat{x}_i(k)$, $\hat{x}_i(l)$ и т.д.

Вводя далее в рассмотрение $(n+1)-$ векторы

$$\tilde{x}(k)=\left(\Delta\hat{x}_2(k),\Delta\hat{x}_3(k),...,\Delta\hat{x}_n(k),\hat{x}_n(k)\right)^T,$$

$$\tilde{x}(l)=\left(\Delta\hat{x}_2(l),\Delta\hat{x}_3(l),...,\Delta\hat{x}_n(l),\Delta\hat{x}_n(l),\hat{x}_n(l)\right)^T,$$

можно переписать метрику (5) в традиционной форме.

$$d^2_{STS}\left(\hat{x}(k),\hat{x}(l)\right)=\left\|\tilde{x}(k)-\tilde{x}(l)\right\|^2, \qquad (6)$$

после чего использовать стандартную FCM-процедуру.

Нечёткая кластеризация временных рядов

Задача кластеризации в рамках стандартного метода нечетных $C-$ средних сводится к минимизации целевой функции:

$$E^{FC}\left(u_j(k),\tilde{c}_j\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}u_j^{\beta}(k)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2 \qquad (7)$$

при наличии ограничений

$$\sum_{j=1}^{m}u_j(k)=1,\quad 0<\sum_{k=1}^{N}u_j(k)\le N, \qquad (8)$$

где $u_j(k)$ – уровень принадлежности вектора $\tilde{x}(k)$ j-му кластеру с прототипом-центроидом $\tilde{c}_j, j=1,2,...,m;$ $m-$ число кластеров, задаваемое априорно, $\beta>1$ параметр фаззификации (фаззификатор), определяющий «размытость» границ между кластерами, при этом $\beta=2$ ведет к наиболее популярному алгоритму нечетной кластеризации Дж. Бездека $[5]$.

Эффективность подхода к кластеризации коротких временных рядов на основе критерия (7) была подтверждена в [7], однако в рассматриваемой здесь задаче, когда общая длина обрабатываемых выборок n (в пределе $n=\sum_{k=1}^{N}n(k)$) может быть достаточно велика, начинает проявляться «концентрация норм», приводящая к неэффективности использования этого критерия. Для преодоления этого негативного эффекта был предложен целый ряд альтернативных (7) критериев, среди которых следует, прежде всего, отметить целевую функцию [14]:

$$E^{KH}\left(u_j(k),\tilde{c}_j\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}\left(\alpha u_j^2(k)+(1-\alpha)u_j(k)\right)\left\|\tilde{x}(k)-\tilde{c}_j^2\right\|^2 \qquad (9)$$

с ограничениями (8).

Здесь $0<\alpha\le 1$ – настраиваемый параметр, устанавливающий компромисс между FCM и четким методом $K-$ средних. Авторами [14] был введен пакетный вариант минимизации (9), а в [15] его online версия.

В [16] была введена процедура нечеткой кластеризации со взвешиванием компонентов обрабатываемых данных. При этом используется целевая функция

$$E^{KK}\left(u_j(k),\tilde{c}_j\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}u_j^{\beta}(k)\times\times\sum_{i=1}^{n}\gamma_{ji}^{t}\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2 \qquad (10)$$

с дополнительным ограничением (кроме (8)):

$$\sum_{i=1}^{n}\gamma_{ji}=1, \qquad (11)$$

где $\gamma_{ji}-$ параметр взвешивания для каждого компонента вектора $\tilde{x}(k)$, $t>0-$ параметр, имеющий смысл аналогичный фаззификатору и выбираемый эмпирически.

В 2013 г. Ф. Клавонн предложил авторам [17] в качестве целевой функции использовать «гибрид» (9) и (10) вида

$$E^{FK}\left(u_j(k),\tilde{c}_j\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}u_j^{\beta}(k)\times\times\left(\alpha\sum_{i=1}^{n}\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2+(1-\alpha)\times\times\sum_{i=1}^{n}\gamma_{ji}^{t}\left(\tilde{x}(k)-\tilde{c}_{ji}\right)^2\right), \qquad (12)$$

обеспечивающий компромисс между стандартным FCM и процедурой кластеризации со взвешиванием данных.

В рассматриваемой здесь задаче мы предлагаем использовать целевую функцию:

$$E^{KB}\left(u_j(k),\tilde{c}_j\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}\left(\alpha u_j^2(k)\times\times\sum_{i=1}^{n}\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2+(1-\alpha)u_j(k)\times\times\sum_{i=1}^{n}\gamma_{ji}^{t}\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}\left(\alpha u_j^2(k)\times\times\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2+(1-\alpha)u_j(k)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}\right), \qquad (13)$$

где

$$\Gamma_j=\mathrm{diag}\left(\gamma_{j1},\gamma_{j2},\ldots,\gamma_{jn}\right),\ T_r\Gamma_j=1. \qquad (14)$$

Для минимизации (13) с учетом (14) введем в рассмотрение функцию Лагранжа

$$L\left(u_j(k),\tilde{c}_j,\rho_j\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}\left(\alpha u_j^2(k)\sum_{i=1}^{n}\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2+\right.$$
$$\left.+(1-\alpha)u_j(k)\sum_{i=1}^{n}\gamma_{ji}^t\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2\right)-$$
$$-\sum_{j=1}^{m}\rho_j\left(\sum_{i=1}^{n}\gamma_{ji}-1\right),$$

(здесь ρ_j – неопределенные множители Лагранжа) и систему уравнений Каруша-Куна-Таккера

$$\begin{cases}\dfrac{\partial L\left(u_j(k),\tilde{c}_j,\rho_j\right)}{\partial\gamma_{ji}}=\sum_{k=1}^{N}(1-\alpha)u_j(k)t\gamma_{ji}^{t-1}\times\\ \times\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2-\rho_j=0,\\ \dfrac{\partial L\left(u_j(k),\tilde{c}_j,\rho_j\right)}{\partial\rho_j}=\sum_{k=1}^{N}\gamma_{ji}-1=0.\end{cases}\qquad(15)$$

Проводя цепочку очевидных преобразований:

$$\rho_j=(1-\alpha)t\gamma_{ji}^{t-1}\sum_{k=1}^{N}u_j(k)\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2,$$

$$\gamma_{ji}^{t-1}=\left(\frac{\rho_j}{(1-\alpha)t\sum_{k=1}^{N}u_j(k)\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2}\right)^{\frac{1}{t-1}},$$

$$1=\left(\frac{\rho_j}{(1-\alpha)t}\right)^{\frac{1}{t-1}}\left(\frac{1}{\sum_{k=1}^{N}u_j(k)\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2}\right)^{\frac{1}{t-1}},$$

$$\rho_j=\frac{(1-\alpha)t}{\left(\sum_{i=1}^{n}\left(\frac{1}{\sum_{k=1}^{N}u_j(k)\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2}\right)^{\frac{1}{t-1}}\right)^{t-1}},$$

$$\gamma_{ji}^{t-1}=\frac{\rho_j}{(1-\alpha)t\sum_{k=1}^{N}u_j(k)\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2},$$

в итоге получим

$$\gamma_{ji}=\frac{1}{\sum_{l=1}^{n}\left(\frac{\sum_{k=1}^{N}u_j(k)\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2}{\sum_{k=1}^{N}u_j(k)\left(\tilde{x}_l(k)-\tilde{c}_{jl}\right)^2}\right)^{\frac{1}{t-1}}},\qquad(16)$$

откуда следует, что параметры γ_{ji} не зависят от α, т.е. взвешивание отдельных компонентов кластеризуемых векторов проводится по типу [16].

Для нахождения центроидов кластеров запишем очевидное уравнение –

$$\frac{\partial E^{KB}\left(u_j(k),\tilde{c}_j\right)}{\partial\tilde{c}_{ji}}=-2\sum_{k=1}^{N}\left(\alpha u_j^2(k)\times\right.$$
$$\times\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)+(1-\alpha)u_j(k)\times\qquad(17)$$
$$\left.\times\gamma_{ji}^t\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)\right)=0,$$

откуда следует

$$\tilde{c}_{ji}=\frac{\sum_{k=1}^{N}\left(\alpha u_j^2(k)+(1-\alpha)u_j(k)\gamma_{ji}^t\right)\tilde{x}_i(k)}{\sum_{k=1}^{N}\left(\alpha u_j^2(k)+(1-\alpha)u_j(k)\gamma_{ji}^t\right)}.\qquad(18)$$

Несложно заметить, что при $\alpha=1$ приходим к стандартному FCM-алгоритму, а при $\gamma_{ji}=1-$ к алгоритму , предложенному в [14].

Для учета ограничений (8) введем еще одну функцию Лагранжа

$$L\left(u_j(k),\tilde{c}_j,\lambda(k)\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}\left(\alpha u_j^2(k)\sum_{i=1}^{n}\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2+\right.$$
$$\left.+(1-\alpha)u_j(k)\sum_{i=1}^{n}\gamma_{ji}^t\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2\right)-$$
$$-\sum_{k=1}^{N}\lambda(k)\left(\sum_{j=1}^{m}u_j(k)-1\right)=\sum_{k=1}^{N}\sum_{j=1}^{m}\left(\alpha u_j^2(k)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2+\right.\qquad(19)$$
$$\left.+(1-\alpha)u_j(k)\left\|\tilde{x}(k)-\tilde{c}_j\right\|_{\Gamma_j^t}^2\right)-\sum_{j=1}^{m}u_j(k)-1$$

(здесь $\lambda(k)-N$ неопределенных множителей Лагранжа) и систему Каруша-Куна-Таккера:

$$\begin{cases}\dfrac{\partial L\left(u_j(k),\tilde{c}_j,\lambda(k)\right)}{\partial u_j(k)}=2\alpha u_j(k)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2\\ +(1-\alpha)\left\|\tilde{x}(k)-\tilde{c}_j\right\|_{\Gamma_j^t}^2-\lambda(k)=0,\\ \dfrac{\partial L\left(u_j(k),\tilde{c}_j,\lambda(k)\right)}{\partial\lambda(k)}=\sum_{j=1}^{m}u_j(k)-1=0.\end{cases}\qquad(20)$$

Проводя аналогично предыдущему цепочку очевидных преобразований:

$$2\alpha u_j(k)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2=\lambda(k)-(1-\alpha)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t},$$

$$u_j(k)=\frac{\lambda(k)-(1-\alpha)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}}{2\alpha\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2},$$

$$\sum_{j=1}^{m}\frac{\lambda(k)-(1-\alpha)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}}{2\alpha\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2}=1,$$

$$\lambda(k)\frac{1}{2\alpha}\sum_{j=1}^{m}\left\|\tilde{x}(k)-\tilde{c}_j\right\|^{-2}=$$

$$=1+\frac{(1-\alpha)}{2\alpha}\sum_{j=1}^{m}\frac{\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}}{\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2},$$

$$\lambda(k)=\frac{1+\frac{(1-\alpha)}{2\alpha}\sum_{j=1}^{m}\frac{\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}}{\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2}}{\frac{1}{2\alpha}\sum_{j=1}^{m}\left\|\tilde{x}(k)-\tilde{c}_j\right\|^{-2}},$$

$$2\alpha u_j(k)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2+(1-\alpha)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}-$$

$$-\frac{1+\frac{(1-\alpha)}{2\alpha}\sum_{j=1}^{m}\frac{\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}}{\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2}}{\frac{1}{2\alpha}\sum_{j=1}^{m}\left\|\tilde{x}(k)-\tilde{c}_j\right\|^{-2}}=0,$$

окончательно получаем процедуру нечетной классификации на основе критерия (13):

$$\begin{cases} u_j(k)=\dfrac{1+\frac{(1-\alpha)}{2\alpha}\sum_{j=1}^{m}\frac{\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}}{\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2}-(1-\alpha)\left\|\tilde{x}(k)-\tilde{c}_j\right\|^2_{\Gamma_j^t}}{\frac{1}{2\alpha}\sum_{k=1}^{m}\left\|\tilde{x}(k)-\tilde{c}_j\right\|^{-2}}}{2\alpha\left\|\tilde{x}(k)-\tilde{c}_j\right\|^{-2}}, \\ \gamma_{ji}=\left(\sum_{l=1}^{n}\left(\dfrac{\sum_{k=1}^{N}u_j(k)\left(\tilde{x}_i(k)-\tilde{c}_{ji}\right)^2}{\sum_{l=1}^{N}u_j(k)\left(\tilde{x}_l(k)-\tilde{c}_{jl}\right)^2}\right)^{\frac{1}{t-1}}\right)^{-1}, \\ \tilde{c}_{ji}=\dfrac{\sum_{k=1}^{N}\left(\alpha u_j^2(k)+(1-\alpha)u_j(k)\gamma_{ji}^t\right)\tilde{x}_i(k)}{\sum_{k=1}^{N}\left(\alpha u_j^2(k)+(1-\alpha)u_j(k)\gamma_{ji}^t\right)}. \end{cases}$$

Несложно заметить, что соотношения (20) являются обобщением алгоритмов нечеткой кластеризации, основанных на целевых функциях (7; 9–10; 12), т.е. при соответствующем выборе свободных параметров превращаются в известные процедуры.

Заключение

Рассмотрена задача нечетной кластеризации временных рядов с неравномерными и асинхронными тактами квантования. Предложена процедура фаззи-кластеризация, не подверженная «концентрации норм» и являющаяся обобщением ряда известных алгоритмов вероятностной нечетной кластеризации. Введенная процедура может быть полезна при решении задач, возникающих в рамках Data Mining, когда исходные данные имеют высокую размерность.

Список литературы

1. Liao T.W. Clustering of time series data. – A survey // Pattern Recognition. – 2005. – 38. – №11 – P. 1857-1874.

2. Aggarwal C.C. Data Clustering. Algorithms and applications / C.C. Aggarwal, C.K. Reddy. – Boca Raton: CRC Press, 2014. – 620 p.

3. Aggarwal C.C. Data Mining / C.C. Aggarwal. – N.Y.: Springer, 2015. – 734 p.

4. Höppner F. Fuzzy Clustering Analysis: Methods of Classifications, Data Analysis, and image Recognition / F. Höppner, F. Klawonn, R. Kruse, T. Runkler. – Chichester: John Wiley & Sons, 1999. – 289 p.

5. Bezdek J.C. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing / J.C. Bezdek, J. Keller, R. Krisnapuram, N.R. Pal. – N.Y.: Springer Science + Business Media, 2005. – 776 p.

6. Klawonn F. What can Fuzzy cluster analysis contribute to clustering of high-dimensional data? / F. Klawonn // Lecture Notes in Artificial intelligence. – Switzerland: Springer Publ., 2013. – V.8256. – P. 1-14.

7. Möller-Levet C.S. Fuzzy clustering of short time series and unevenly distributed sampling points / C.S. Möller-Levet, F. Klawonn, K.-H. Cho, O. Wolkenhauer // Lecture Notes in Computers Science. – Heidelberg: Springer, 2003. – Vol. /2810. – P. 330-340.

8. Cruz C.P. Fuzzy clustering for incomplete short time series data / C.P. Cruz, S.M. Vilira, S. Vinga // Lecture Notes in Artificial intelligence. – Heidelberg: Splinger, 2015. – 9273. – P. 353-359.

9. Hathaway R.J. Fuzzy c-means clustering of incomplete data / R.J. Hathaway, J.C. Bezdek // IEEE Trans. on systems, Man, and Cybernetics. – Part B. – 2001. – 31. – №5. – P. 735-744.

10. Бодянский Е.В. Адаптивная нечетная кластеризация коротких временных рядов в интеллектуальном анализе потоков данных / Е.В. Бодянский, А.А. Дейнеко, И.О. Кобылин, И.П. Плисс // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту: Матеріали міжнар. наук. конф. – Херсон: Вид-во ПП Вишемирський В.С., 2016. – С. 255-258.

11. Bodyanskiy Ye. Recursive fuzzy clustering, algorithms / Ye. Bodyanskiy, V. Kolodyazhniy, A. Stephan // Proc. East West Fuzzy Colloquium. – Zittau/Görlitz: HS, 2002. – P. 164-172.

12. Gorshkov Ye. New recursive learning algorithms for fuzzy Kohonen clustering network / Ye. Gorshkov, Ye.V. Bodyanskiy, V. Kolodyazhniy // Proc. 17th Workshop on Nonlinear-Dynamics of Electronic Systems. – Rapperswil, Switzerland, 2009. – P. 58-61.

13. Kohonen T. Self-Organizing Maps / T. Kohonen. – Berlin: Springer–Verlag, 1995. – 362 p.

14. Klawonn F. What is fuzzy about clustering? Understanding and improving the concept of the fuzzifier / F. Klawonn, F. Höppner // Lecture Notes in Computer Science. – Berlin Heidelberg. – Springer, 2002. – V. 2811. – P. 276-283.

15. Kolchygin B.V. Adaptive fuzzy clustering with variable fuzzifier / B.V. Kolchygin, Ye.V. Bodyanskiy // Cybernetics and System Analysis. – 2013. – 49. – №3. – P. 176-181.

16. Keller A. Fuzzy clustering with weighting of data weight variables / A. Keller, F. Klawonn // Uncertainty, Fuzzinees, and Knowledge –Based Systems. – 2000. – 8. – №6. – P. 235-246.

17. Бодянский Е.В. Об одном алгоритме нечетной кластеризации данных высокой размерности / Е.В. Бодянский, И.П. Плисс, А.Ю. Шафроненко // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту: Матеріали Міжнар. наук. конф. – Херсон: ХНТУ, 2014. – С. 249-251.

References

1. Liao, T.W. (2005), Clustering of time series data. – A survey, *Pattern Recognition*, 38, No. 11, pp. 1857-1874.

2. Aggarwal, C.C. and Reddy, C.K. (2014), *Data Clustering. Algorithms and applications*, CRC Press, Boca Raton, 620 p.

3. Aggarwal, C.C. (2015), *Data Mining*, Springer, N.Y., 734 p.

4. Höppner, F., Klawonn, F., Kruse, R. and Runkler, T. (1999), *Fuzzy Clustering Analysis: Methods of Classifications, Data Analysis, and image Recognition*, John Wiley & Sons, Chichester, 289 p.

5. Bezdek, J.C., Keller, J., Krisnapuram, R and Pal, N.R. (2005), *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Springer Science + Business Media, N.Y., 776 p.

6. Klawonn, F. (2013), What can Fuzzy cluster analysis contribute to clustering of high-dimensional data? *Lecture Notes in Artificial intelligence*, Vol. 8256, Springer Publ., Switzerland, pp. 1-14.

7. Möller-Levet, C.S., Klawonn, F., Cho, K.-H. and Wolkenhauer, O. (2003), Fuzzy clustering of short time series and unevenly distributed sampling points, *Lecture Notes in Computers Science*, Vol. /2810, Springer, Heidelberg, pp. 330-340.

8. Cruz, C.P., Vilira, S.M. and Vinga, S. (2015), Fuzzy clustering for incomplete short time series data, *Lecture Notes in Artificial intelligence*, 9273, Splinger, Heidelberg, pp. 353-359.

9. Hathaway, R.J. and Bezdek, J.C. (2001), Fuzzy c-means clustering of incomplete data, *IEEE Trans. on systems, Man, and Cybernetics*, Part B, 3, No.5, pp. 735-744.

10. Bodyanskiy, E.V., Dejneko, A.A., Kobylin, Y.O. and Plyss, Y.P. (2016), "Adaptyvnaja nechetnaja klasteryzacyja korotkykh vremennih rjadov v yntellektualjnom analyze potokov dannikh" [Adaptive Odd Clustering of Short Time Series in Intellectual Analysis of Data Flows], *Intelektualjni systemy pryjnjattja rishenj i problemy obchysljuvaljnogho intelektu: Materialy mizhnar. nauk. konf.*, Vyd-vo PP Vyshemyrsjkyj V.S., Kherson, pp. 255-258.

11. Bodyanskiy, Ye., Kolodyazhniy, V. and Stephan, A. (2002), Recursive fuzzy clustering, algorithms, *Proc. East West Fuzzy Colloquium*,HS, Zittau/Görlitz, pp. 164-172.

12. Gorshkov, Ye., Bodyanskiy, Ye.V. and Kolodyazhniy, V. (2009), New recursive learning algorithms for fuzzy Kohonen clustering network, *Proc. 17th Workshop on Nonlinear-Dynamics of Electronic Systems*, Rapperswil, Switzerland, pp. 58-61.

13. Kohonen, T. (1995), *Self-Organizing Maps*, Springer – Verlag, Berlin, 362 p.

14. Klawonn, F., and Höppner, F. (2002), What is fuzzy about clustering? Understanding and improving the concept of the fuzzifier, *Lecture Notes in Computer Science*, Vol. 2811, Springer, Berlin Heidelberg, pp. 276-283.

15. Kolchygin, B.V. and Bodyanskiy, Ye.V. (2013), Adaptive fuzzy clustering with variable fuzzifier, *Cybernetics and System Analysis*, 49, No. 3, pp. 176-181.

16. Keller, A. and Klawonn, F. (2000), Fuzzy clustering with weighting of data weight variables, *Uncertainty, Fuzzinees, and Knowledge –Based Systems*, 8, No. 6, pp. 235-246.

17. Bodyanskiy, E.V., Plyss, Y. and Shafronenko, A. (2014), "Ob odnom alghorytme nechetnoj klasteryzacyy dannikh visokoj razmernosty" [On an algorithm for odd high-dimensional data clustering], I*ntelektualjni systemy pryjnjattja rishenj i problemy obchysljuvaljnogho intelektu: Materialy Mizhnar. nauk. konf.*, KhNTU, Kherson, pp. 249-251.

Поступила в редколлегию 8.09.2017
Одобрена к печати 16.11.2017

Відомості про авторів:

Бодянський Євгеній Володимирович
доктор технічних наук професор
професор кафедри Харківського національного університету радіоелектроніки,
Харків, Україна
https://orcid.org/0000-0001-5418-2143
e-mail: yevgeniy.bodyaskiy@nure.ua

Винокурова Олена Анатоліївна
доктор технічних наук професор
головний науковий співробітник Харківського національного університету радіоелектроніки,
Харків, Україна
https://orcid.org/0000-0002-9414-2477
e-mail: vynokurova@gmail.com

Пелешко Дмитро Дмитрович
доктор технічних наук, професор
проректор з науково-педагогічної роботи ДВНЗ «Комп'ютерна академія ШАГ»,
Львів, Україна
https://orcid.org/0000-0003-4881-6933
e-mail: dpeleshko@gmail.com

Кобилін Ілля Олегович
аспірант кафедри Харківського національного університету радіоелектроніки,
Харків, Україна
https://orcid.org/0000-0002-4552-9616
e-mail: ilya.kobylin@nure.ua

Кобилін Олег Анатолійович
кандидат технічних наук доцент
доцент кафедри Харківського національного університету радіоелектроніки,
Харків, Україна
https://orcid.org/0000-0003-0834-0475
e-mail: oleg.kobylin@nure.ua

Information about the authors:

Bodyanskiy Yevgeniy
Doctor of Technical Sciences Professor
Professor of Department at Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine
https://orcid.org/0000-0001-5418-2143
e-mail: yevgeniy.bodyaskiy@nure.ua

Vynokurova Olena
Doctor of Technical Sciences Professor
Principal Research Scientist at Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine
https://orcid.org/0000-0002-9414-2477
e-mail: vynokurova@gmail.com

Peleshko Dmytro
Doctor of Technical Sciences, Professor
Vice-rector for research at University "IT Step Academy"
Lviv, Ukraine
https://orcid.org/0000-0003-4881-6933
e-mail: dpeleshko@gmail.com

Kobylin Ilya
Postgraduate Student of Department at Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine
https://orcid.org/0000-0002-4552-9616
e-mail: ilya.kobylin@nure.ua

Kobylin Oleg
Candidate of Sciences Associate Professor
Senior Lecturer of Department at Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine
https://orcid.org/0000-0003-0834-0475
e-mail: oleg.kobylin@gmail.com

НЕЧІТКА КЛАСТЕРИЗАЦІЯ ЧАСОВИХ РЯДІВ З НЕРІВНОМІРНИМИ ТА АСИНХРОННИМИ ТАКТАМИ КВАНТУВАННЯ

Є.В. Бодянський, О.А. Винокурова, Д.Д. Пелешко, І.О. Кобилін, О.А. Кобилін

У статті запропоновано алгоритм нечіткої кластеризації часових рядів з нерівномірними асинхронними тактами квантування (біомедичні масиви спостережень, сигнали цифрового відео, формуючі дискретні двомірні поля и тощо). Цей алгоритм характеризується простотою обчислювальної реалізації, високими апроксимуючими властивостями, високою швидкістю процесами навчання та призначений для вирішення широкого класу задач, у тому числі, коли початкові дані мають високу розмірність.

Ключові слова: *нечітка кластеризація часових рядів, асинхронність, квантування, data mining, адаптивні процедури навчання. online процедура нечіткої кластеризації.*

FUZZY CLUSTERING OF TIME SERIES WITH NON-UNIFORM AND ASYNCHRONOUS QUANTIZATION

Y. Bodyanskiy, O. Vynokurova, D. Peleshko, I. Kobylin, O. Kobylin

In fuzzy clustering tasks of time series clustering, there are situations when the methods proposed earlier can not be used under conditions of processing uneven time series due to problems with the effect of "concentration of norms", as well as for the asynchrony in the realizations of the observed signal. To do this, in the article the authors proposed the algorithm for fuzzy clustering of time series with non-uniformly-asynchronous quantization tacts, which can be successfully applied in biomedical observation arrays, digital video signals and more. To overcome the negative effect, alternative criterias were proposed, some of which are a trade-off between FCM and crisp K-medium method, the basis of which is the standard C-means modified method, for using in uneven and asynchronous quantization conditions, and a modified PS-distance based on the investigations series in the form of piecewise linear functions which is used to estimate the distance. The article considers examples with rather large serious. This algorithm is characterized by the simplicity of the computational realization, high approximative properties, high speed of the learning process, and is designed to solve a wide range of problems, including situations when the initial data have a high dimension.

Keywords: *fuzzy time series clustering, asynchrony, quantization, data mining, adaptive learning procedure, online procedure for fuzzy clustering.*