

КЛАСТЕРИЗАЦИЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ВИДЕОДАНЫХ НА ОСНОВЕ ГАРМОНИЧЕСКИХ k -СРЕДНИХ

Аннотация. Работа посвящена сегментации–кластеризации видеопоследовательностей с помощью анализа многомерных временных последовательностей. Предложен подход к использованию глубокой итеративной временной деформации совместно с матричным методом гармонических k -средних. Такая процедура сегментации–кластеризации в отличие от традиционного подхода нечувствительна к начальному выбору центроидов, что особенно удобно в условиях анализа произвольных данных больших объемов.

Ключевые слова: сегментация, кластеризация, многомерные последовательности, видео, динамическая деформация.

ВВЕДЕНИЕ

Задача кластеризации и классификации массивов наблюдений занимает важное место в Data Mining, а для ее решения на сегодняшний день существует множество подходов, методов и алгоритмов, основанных на соответствующем математическом аппарате при различных априорных предположениях о природе данных [1–6]. Особое место при этом занимают методы, основанные на прототипах-центроидах благодаря ясной физической интерпретации и простой численной реализации, базирующиеся на тех или иных алгоритмах оптимизации соответствующих критериев качества. Здесь прежде всего отметим процедуры k -means, k -medoids, k -representatives, существующие как в четких (crisp), так и в нечетких (fuzzy) вариантах и породившие множество модификаций. Эти процедуры отличаются одна от другой видом используемой метрики, задающей расстояние между наблюдениями из исходного массива, и возможными прототипами-центроидами. Наиболее популярными здесь являются процедуры k -means при использовании евклидовой метрики. При этом исходную задачу удастся свести к соответствующему классу задач квадратичной оптимизации.

ПОСТАНОВКА ЗАДАЧИ

Пусть исходной информацией является массив векторных наблюдений

$$X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}, \quad x(k) = (x_1(k), \dots, x_i(k), \dots, x_N(k))^T,$$

где k — номер наблюдений в исходном массиве. Тогда исходная задача кластеризации сводится к поиску набора центроидов c_1, \dots, c_M ; $c_j = (c_{j1}, \dots, c_{ji}, \dots, c_{jN})^T$, $j=1, M$, обеспечивающих минимум целевой функции:

$$E(x(k), c_j) = \sum_{k=1}^N \sum_{j=1}^M u_j(k) \|x(k) - c_j\|^2 = \sum_{k=1}^N \sum_{j=1}^M \mu_j(k) (x(k) - c_j)^T (x(k) - c_j), \quad (1)$$

где $\mu_j(k)$ — уровень четкой принадлежности k -го наблюдения j -му кластеру.

Уровень четкой принадлежности задается выражением

$$\mu_j(k) = \begin{cases} 1, & \text{если } x(k) \text{ принадлежит } j\text{-му кластеру,} \\ 0 & \text{в противном случае.} \end{cases} \quad (2)$$

В задачах нечеткой кластеризации уровень принадлежности принимает значения на отрезке

$$0 \leq \mu_j(k) \leq 1 \quad (3)$$

и удовлетворяет условию

$$\sum_{j=1}^M \mu_j(k) = 1 \quad \forall k = 1, 2, \dots, N. \quad (4)$$

Задача кластеризации резко усложняется в случае, если обработке подлежат не векторы, а матрицы

$$x(k) = \{x_{i_1, i_2}(k)\}, \quad i_1 = 1, 2, \dots, n; \quad i_2 = 1, 2, \dots, v.$$

Такая ситуация возникает в задачах обработки изображений, когда кластеризации подлежат отдельные кадры видеопоследовательности. Естественно, можно осуществить векторизацию матрицы $x(k)$, преобразовав ее в $(nv \times 1)$ -вектор. Однако в дальнейшем в процессе решения задачи оптимизации могут возникнуть трудности, связанные с так называемым эффектом проклятия размерности. Именно поэтому естественно рассматривать в качестве объектов кластеризации непосредственно матрицы $x(k)$, используя выражения для норм в матричной форме, согласованной с векторными формами этих норм. В частности, в статье [7] введен матричный аналог k -means, а в [8] — матричный аналог fuzzy c -means, в основу которых была положена сферическая норма Фробениуса. Представляется целесообразным для обработки изображений использовать и другие нормы, обеспечивающие процедуре кластеризации новые специфические свойства [9].

МАТРИЧНЫЙ МЕТОД HARMONIC k -MEANS

Развитием метода k -means, использующим евклидову норму, является процедура Harmonic k -means (НKM) [10–12], в основе которой — понятие гармонического среднего. Гармоническим средним для скалярного набора $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$ является выражение

$$\bar{x}_H = \frac{N}{\sum_{k=1}^N x^{-1}(k)}. \quad (5)$$

В монографии [1] показано, что процедура НKM приводит к результатам, близким к тем, которые обеспечиваются k -means. Однако эта процедура гораздо менее чувствительна к начальному заданию центроидов-прототипов, что особенно важно при обработке больших массивов информации. Если критерий качества кластеризации векторных наблюдений для k -means имеет вид (1), то целевую функцию для НKM можно представить как

$$E(x(k), c_j) = \sum_{k=1}^N \frac{M}{\sum_{j=1}^M \frac{1}{\|x(k) - c_j\|^2}} = \sum_{k=1}^N \frac{M}{\sum_{j=1}^M \frac{1}{(x(k) - c_j)^T (x(k) - c_j)}}. \quad (6)$$

Если обрабатываемая выборка $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$ образована $(n \times v)$ -матрицами $x(k) = \{x_{i_1, i_2}(k)\}$, то использование нормы Фробениуса в (6) соответственно приводит к выражению

$$E(x(k), c_j) = \sum_{k=1}^N \frac{M}{\sum_{j=1}^M \frac{1}{\|x(k) - c_j\|^2}} = \sum_{k=1}^N \frac{M}{\sum_{j=1}^M \frac{1}{\text{Sp}((x(k) - c_j)(x(k) - c_j)^T)}}, \quad (7)$$

где $\text{Sp}(\cdot)$ — символ следа матрицы, $c_j, j = 1, 2, \dots, M$, — матричные прототипы-центроиды.

Оптимизация целевой функции (6) позволяет получить следующую оценку [13]:

$$c_j = \frac{\sum_{k=1}^N \|x(k) - c_j\|^{-4} \left(\sum_{j=1}^M \|x(k) - c_j\|^{-2} \right)^{-2} x(k)}{\sum_{k=1}^N \|x(k) - c_j\|^{-4} \left(\sum_{j=1}^M \|x(k) - c_j\|^{-2} \right)^{-2}} =$$

$$= \frac{\sum_{k=1}^N \text{Sp}((x(k) - c_j)(x(k) - c_j)^T)^{-2} \left(\sum_{j=1}^M \text{Sp}((x(k) - c_j)(x(k) - c_j)^T)^{-1} \right)^{-2} x(k)}{\sum_{k=1}^N \text{Sp}((x(k) - c_j)(x(k) - c_j)^T)^{-2} \left(\sum_{j=1}^M \text{Sp}((x(k) - c_j)(x(k) - c_j)^T)^{-1} \right)^{-2}}. \quad (8)$$

При этом уровень нечеткой принадлежности на основе (8) задается в виде

$$\mu_j(k) = \frac{\|x(k) - c_j\|^{-4}}{\sum_{j=1}^M \|x(k) - c_j\|^{-4}} = \frac{\text{Sp}((x(k) - c_j)(x(k) - c_j)^T)^{-2}}{\sum_{j=1}^M \text{Sp}((x(k) - c_j)(x(k) - c_j)^T)^{-2}}. \quad (9)$$

Данная оценка близка к оценке принадлежности на основе FCM [7]:

$$\mu_j(k) = \frac{\|x(k) - c_j\|^{-2}}{\sum_{j=1}^M \|x(k) - c_j\|^{-2}}. \quad (10)$$

Несложно увидеть, что с вычислительной точки зрения процедура НКМ лишь ненамного сложнее стандартного метода FCM. Это объясняется тем, что вычисление следа матрицы не требует значительных вычислительных ресурсов, являясь линейной функцией, и соответственно не увеличивает сложности алгоритма.

КЛАСТЕРИЗАЦИЯ МАТРИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С ИСПОЛЬЗОВАНИЕМ HARMONIC *k*-MEANS

Задача кластеризации резко усложняется, если объектом исследования являются не отдельные изображения, а их последовательности, представленные, например, видеорядом. Задача сегментации и кластеризации для векторных последовательностей рассмотрена в [14, 15]. При этом предполагалось, что в распоряжении исследователя имеется единственный временной ряд, который разбит на однородные участки с помощью тех или иных алгоритмов нечеткой кластеризации. Ситуация существенно усложняется, если обработке подлежит множество рядов. При этом на первый план выносятся вопросы оценки расстояния между последовательностями различной длины. Здесь может быть использована процедура Dynamic Time Warping (DTW) [16]. Однако такой кластеризации могут быть подвергнуты только короткие последовательности [17, 18]. Если обрабатываемые сигналы содержат большое число отсчетов (кадров видеопоследовательности), то такой подход становится неэффективным. В этой ситуации весьма полезной может оказаться процедура Iterative Deepening Dynamic Time Warping (IDDTW), описанная в [19]. Однако ее применение требует существенной модификации для случая, если объектом исследований являются достаточно длинные матричные временные ряды.

Стандартная DTW основана на вычислении расстояния (схожести) между двумя временными матричными последовательностями: $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$,

$x(k) = \{x_{i_1, i_2}(k)\}$, и $Y = \{y(1), y(2), \dots, x(l), \dots, x(M)\}$, $y(l) = \{y_{i_1, i_2}(l)\}$, при $N \neq M$ с использованием сферической метрики

$$d(x(k), y(l)) = (\text{Sp}(x(k) - y(l))(x(k) - y(l))^T)^{1/2} \quad (11)$$

между двумя отдельными кадрами рядов X и Y .

Таким образом, исходной информацией для расчетов является $(N \times M)$ -матрица расстояний (11) между всеми отсчетами двух рядов.

В качестве меры близости рядов выбирается так называемый оптимальный искривленный путь (optimal warping path), который обозначим $D(N, M)$. Такой путь строится с помощью процедуры динамического программирования:

$$D(k, l) = d(x(k), y(l)) + \min \{D(k-1, l), D(k, l-1), D(k-1, l-1)\}, \quad (12)$$

$$k = 1, 2, \dots, N; \quad l = 1, 2, \dots, M.$$

Очевидно, что при больших N и M использование методов динамического программирования для поиска $D(N, M)$ становятся неэффективным в силу большой размерности данного класса задач.

Преодолеть эти трудности возможно с помощью IDDTW, однако при использовании стандартной процедуры также возникают проблемы. При этом исходные ряды разбиваются на четыре участка (отсчета), а на каждом из них эти последовательности заменяются своими средними значениями. Далее проводится стандартное оценивание расстояния между матричными рядами, содержащими по четыре отсчета. Если получаемая оценка расстояния неудовлетворительна, в рассмотрение вводится 8, 16 и т.д. участков. Это означает, что количество отсчетов возрастает в геометрической прогрессии.

Для определения расстояния можно использовать как выражение (12), так и более простую с вычислительной точки зрения оценку (11), поскольку в последнем случае рассматривается $N+2, N+3, \dots$ участков обрабатываемых рядов, а не 2^N , где $N = 2, 3, \dots$

Таким образом, если обработке подлежит Q временных рядов $x_1, \dots, x_q, \dots, x_Q$, каждый из которых содержит $N_1, \dots, N_q, \dots, N_Q$ отсчетов, получаем набор матричных последовательностей $\bar{x}_1, \dots, \bar{x}_q, \dots, \bar{x}_Q$, содержащих фиксированное количество K отсчетов, т.е.

$$\bar{x}_q = \{\bar{x}_q(1), \dots, \bar{x}_q(k), \dots, \bar{x}_q(K)\}, \quad \bar{x}_q(k) = \{\bar{x}_{qi_1, i_2}(k)\}.$$

При этом квадрат расстояния между любыми двумя отсчетами разных рядов с использованием выражения (11) задается как

$$d^2(\bar{x}_q(k), \bar{y}_p(l)) = \text{Sp}(\bar{x}_q(k) - \bar{y}_p(l))(\bar{x}_q(k) - \bar{y}_p(l))^T. \quad (13)$$

Рассматривая M центроидов $c_j, j = 1, 2, \dots, m$, в виде $(n \times vk)$ -матриц с элементами, подлежащими оцениванию, можно предложить следующую процедуру кластеризации матричных временных рядов произвольной длины:

$$c_j = \frac{\sum_{q=1}^Q d^{-4}(\bar{x}_q, c_j) \left(\sum_{j=1}^M d^{-2}(\bar{x}_q, c_j) \right)^{-2} \bar{x}_q}{\sum_{q=1}^Q d^{-4}(\bar{x}_q, c_j) \left(\sum_{j=1}^M d^{-2}(\bar{x}_q, c_j) \right)^{-2}}, \quad (14)$$

$$\mu_j(q) = \frac{d^{-4}(\bar{x}_q, c_j)}{\sum_{j=1}^M d^{-4}(\bar{x}_q, c_j)}.$$

Имеем обобщение метода Harmonic k -means на случай длинных матричных последовательностей, содержащих разное число отсчетов.

ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ СЕГМЕНТАЦИИ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ

Анализ предложенного подхода к сегментации–кластеризации многомерных данных осуществляется с использованием видеороликов из документального цикла “Destroyed in Seconds” фильмов канала Discovery. Выбор подобных исходных данных объясняется тем, что в этом случае на относительно коротких видеопоследовательностях происходит неоднократное изменение сцены как из-за смены ракурса снимающей камеры, так и из-за быстрых изменений, происходящих в результате некоего инцидента (катастрофы) в этих видео. На рис. 1 дан пример видеопоследовательностей и их пространственных сегментаций, полученных автоматически (без вмешательства человека), которые необходимы для формирования массива векторных наблюдений X .

Таким образом, в качестве исходных данных для эксперимента выступает множество результатов пространственных сегментаций видеопоследовательностей различной длины. Здесь каждая последовательность содержит совокупность сегментов, которые обладают схожими характеристиками и слабо коррелируют с соседними сегментами.

Рассмотрим пример сегментации–кластеризации видеопоследовательности, представленной на рис. 1. В общей сложности видеопоследовательность разбивается на пять сегментов. На рис. 2 для четырех из них приведены характерные кадры, откуда видно, насколько отличаются сегменты между собой. При этом каждый сегмент имеет свои значения характеристик, по которым формируется кластер. Здесь N — номер кадра видеопоследовательности, D — расстояние между кадрами.

С учетом того, что в основе подхода к кластеризации лежит Iterative Deepening Dynamic Time Warping, нами рассмотрена ситуация сравнения двух видеопоследовательностей разной длины (рис. 3). Сравнение результатов кластеризации–сегментаций позволяет сделать вывод, что поиск значений характеристик временного ряда не зависит от его длины. Так, в случае, когда сравнивае-

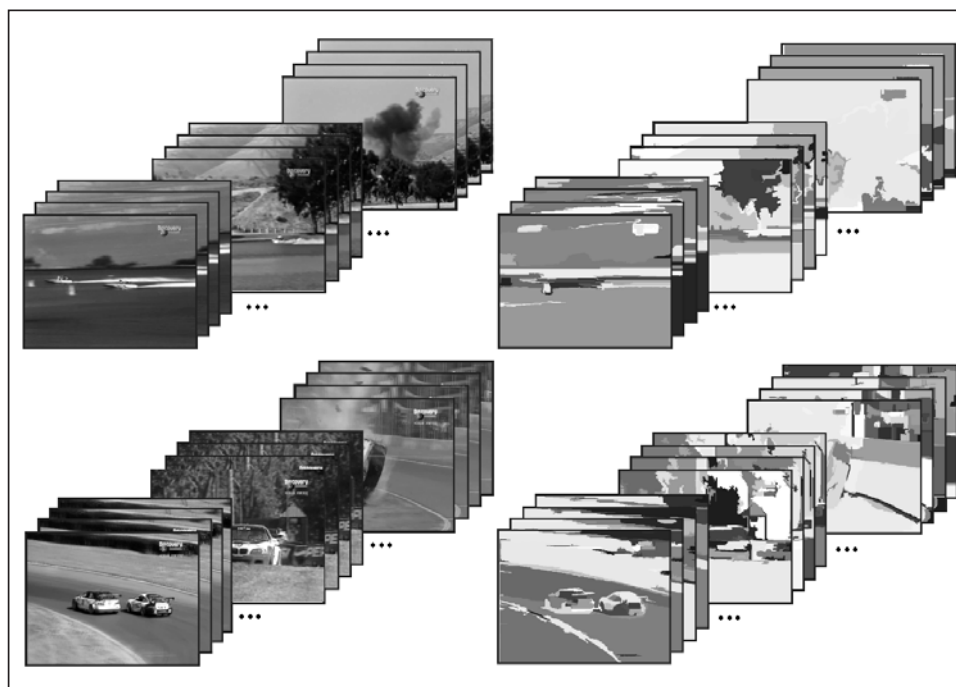


Рис. 1. Видеопоследовательности и их пространственные сегментации

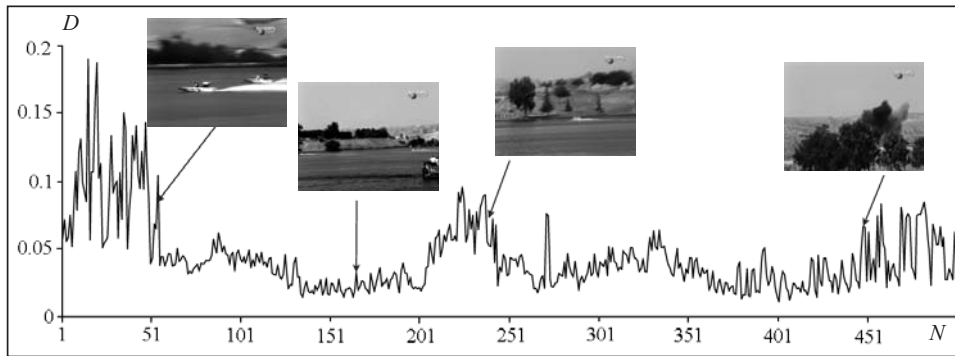


Рис. 2. Временная сегментация видеопоследовательности

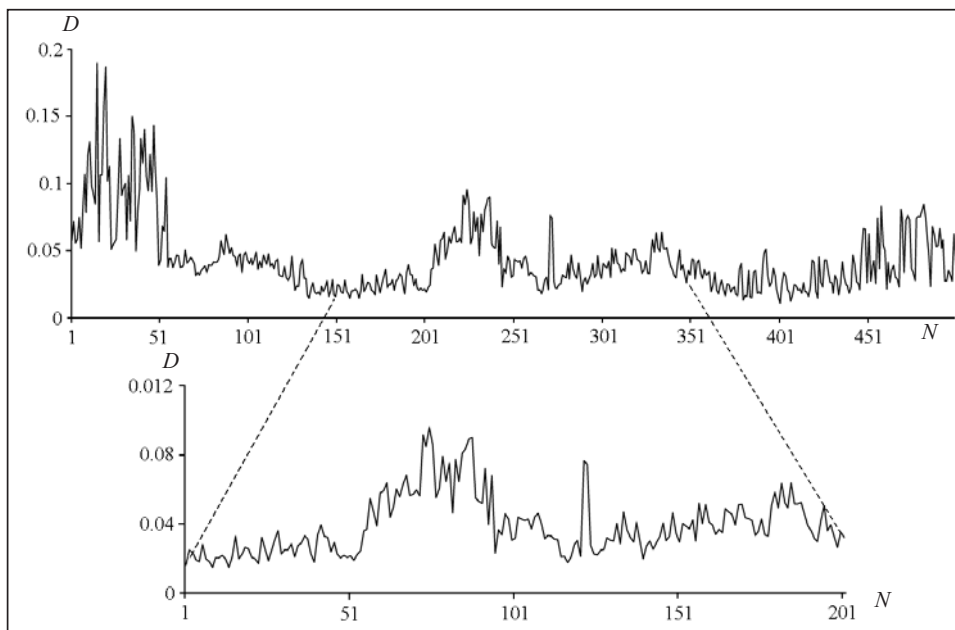


Рис. 3. Сравнение кластеризаций–сегментаций видеопоследовательностей различной длины

мый ряд меньшей длины является подмножеством исходной последовательности, получаем те же характеристики, что и в соответствующем многомерном временном ряде большей длины. Таким образом, предложенный подход позволяет разбивать видеопоследовательности различной длины на совокупность кластеров, содержащих последовательности кадров с аналогичными характеристиками. Центроид кластера при этом может служить ключевым кадром этого сегмента. Можно утверждать, что совокупность полученных предложенным подходом центроидов будет определять видеопоследовательность в целом.

Представляет интерес сравнение предложенной процедуры с известными методами кластеризации видеоданных. Однако большинство исследователей, как правило, приводят данные вычислительных экспериментов без предоставления конкретного исходного видеоматериала и его характеристик, ограничиваясь указанием лишь общих данных. Этот факт делает некорректными любые прямые сравнения работы алгоритмов особенно при анализе видеоданных произвольной природы. Поэтому в настоящей статье основное внимание уделено теоретическому обоснованию эффективности предложенного подхода и экспериментальным подтверждением его работоспособности. В дальнейшем предполагается осуществить сравнительный анализ известных подходов на основе выбора тестовых видеоданных.

ЗАКЛЮЧЕНИЕ

Рассмотренный метод кластеризации матричных последовательностей произвольной длины является модификацией и обобщением метода Harmonic k -means. Данная процедура предназначена для обработки (сегментации, обнаружении изменений свойств, кластеризации без учителя) больших объемов видеоданных и с вычислительной точки зрения более эффективна. В отличие от традиционного метода k -means в задачах видеообработки рассматриваемая процедура не критична к начальному выбору прототипов-центроидов и позволяет осуществлять кластеризацию данных в условиях перекрывающихся классов [20]. Результаты представляют особый интерес при анализе видеоданных с нечетко выраженными переходами между сценами. Таким образом, представляется возможность разработки новых подходов по быстрому контекстному поиску в базах видеоданных произвольной природы. Перспективным при обработке и кластеризации видеоданных представляется также использование конфигурационных пространств геометрических объектов [21, 22].

СПИСОК ЛИТЕРАТУРЫ

1. Gan G., Ma C., Wu J. Data clustering: Theory, algorithms and applications. Philadelphia: SIAM, 2007. 466 p.
2. Jain A.K., Dubes R.C. Algorithms for clustering data. Englewood Cliffs, N.J.: Prentice Hall, 1988. 320 p.
3. Olson D.L., Dursun D. Advanced data mining techniques. Berlin: Springer, 2008. 180 p.
4. Hulianytskyi L., Malyshko S. Big data in information analytical system "NEWSCAPE". In: *Data Stream Mining & Processing: Proc. IEEE First Int. Conf. on Data Stream Mining & Processing* (23–27 August, 2016, Lviv, Ukraine) 2016. P. 382–386.
5. Hulianytskyi L., Riasna I. Formalization and classification of combinatorial optimization problems. *Optimization Methods and Applications*. 2017. Vol. 130. P. 239–250.
6. Hulianytskyi L., Riasna I. Automatic classification method based on a fuzzy similarity relation. *Cybernetics and Systems Analysis*. 2016. Vol. 52, N 1. P. 30–37.
7. Ведмедь А.Г., Машталир С.В., Сакало Е.С. Матричный алгоритм самообучения карты Кохонена в задачах обработки изображений. *Системи управління, навігації та зв'язку*. 2009. Вип. 1. С. 188–192.
8. Bodyanskiy Ye., Volkova V., Skuratov M. Matrix neuro-fuzzy self-organizing clustering network. *Scientific Journal of Riga Technical University. Computer Science. Information Technology and Management Science*. 2011. Vol. 49. P. 54–58.
9. Mashtalir V.P., Yakovlev S.V. Point-set methods of clusterization of standard information. *Cybernetics and Systems Analysis*. 2001. Vol. 37, N 3. P. 295–307.
10. Han J., Kamber M. Data mining: Concepts and techniques. San Francisco: Morgan Kaufmann, 2006. 26 p.
11. Xu R., Wunsch D. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*. 2005. Vol. 16, N 3. P. 645–678.
12. Güngör Z., Ünler A. K -harmonic means data clustering with tabu-search method. *Applied Mathematical Modelling*. 2008. Vol. 32, N 6. P. 1115–1125.
13. Zhang B., Hsu M., Dayal V. K -harmonic means: A spatial clustering algorithm with boosting. *First International Workshop on Temporal, Spatial and Spatio-Temporal Data Mining*. London: Springer-Verlag, 2000. P. 31–45.
14. Abonyi J., Feil B., Nemeth S., Arva P. Fuzzy clustering based segmentation of timeseries. *Lecture Notes in Computer Science*. 2003. Vol. 2810. P. 275–285.
15. Liao T.W. Clustering of time series data: A survey. *Pattern Recognition*. 2005. Vol. 38, N 11. P. 1857–1874.
16. Berndt D.J., Clifford J. Using dynamic time warping to find patterns in time series. In: *Proc. 3rd International Conference on Knowledge Discovery and Data Mining. KDD Workshop*, 1994. P. 359–370.

17. Setlak G., Bodyanskiy Y., Pliss I., Vynokurova O., Peleshko D., Kobylin I. Adaptive fuzzy clustering of multivariate short time series with unevenly distributed observations based on matrix neuro-fuzzy self-organizing network. *Advances in Intelligent Systems and Computing*. 2018. Vol. 643. P. 308–315.
18. Bodyanskiy Y., Kobylin I., Rashkevych Y., Vynokurova O., Peleshko D. Hybrid fuzzy clustering algorithm of unevenly and asynchronously spaced time series in computer engineering. In: *Proc. 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering* (20–24 February 2018, Lviv–Slavske, Ukraine). P. 930–935.
19. Chu S., Keogh E., Hart D., Pazzani M. Iterative deepening dynamic time warping for time series. In: *Proc. of the 2nd SIAM International Conference on Data Mining*, 2002. P. 195–212.
20. Gerasin S.N., Shlyakhov V.V., Yakovlev S.V. Set coverings and tolerance relations. *Cybernetics and Systems Analysis*. 2008. Vol. 43, N 3. P. 333–340.
21. Stoyan Y.G., Yakovlev S.V. Configuration space of geometric objects. *Cybernetics and Systems Analysis*. 2018. Vol. 54, N 5. P. 716–726.
22. Yakovlev S.V. On some classes of spatial configurations of geometric objects and their formalization. *Journal of Automation and Information Sciences*. 2018. Vol. 50, N 9. P. 38–50.

Надійшла до редакції 26.07.2018.

С.В. Машталір, М.І. Столбовий, С.В. Яковлев
КЛАСТЕРИЗАЦІЯ ПОСЛІДОВНОСТЕЙ ВІДЕОДАНИХ НА ОСНОВІ
ГАРМОНІЧНИХ k -СЕРЕДНІХ

Анотація. Розглянуто сегментацію–кластеризацію відеопослідовностей за допомогою аналізу багатовимірних часових послідовностей. Запропоновано підхід для використання глибокої ітеративної часової деформації разом з матричним методом гармонічних k -середніх. Така процедура сегментації–кластеризації на відміну від традиційного підходу не є чутливою до початкового вибору центрів, що особливо зручно в умовах аналізу довільних даних великих обсягів.

Ключові слова: сегментація, кластеризація, багатовимірні послідовності, відео, динамічна деформація.

S.V. Mashtalir, M.I. Stolbovyi, S.V. Yakovlev
VIDEO SEQUENCES CLUSTERING BY THE k -HARMONIC MEANS

Abstract. The study is devoted to segmentation–clustering of video sequences by the analysis of multidimensional time sequences. An approach is proposed for using an iterative deepening time warping in conjunction with the matrix harmonic k -means. This segmentation–clustering procedure, unlike the traditional approach, is insensitive to the initial centroids selection, which is especially useful in the analysis of arbitrary mass data.

Keywords: segmentation, clustering, multidimensional sequences, video, dynamic warping.

Машталір Сергей Владимирович,
 доктор техн. наук, профессор кафедры Харьковского национального университета радиоэлектроники,
 e-mail: sergii.mashtalir@nure.ua.

Столбовий Михаил Иванович,
 аспирант Харьковского национального университета радиоэлектроники, e-mail: st.mihail92@gmail.com.

Яковлев Сергей Всеволодович,
 доктор физ.-мат. наук, профессор, профессор кафедры Национального аэрокосмического университета им. Н.Е. Жуковского «Харьковский авиационный институт», e-mail: svsyak7@gmail.com.