

УДК 004.67

Запропоновано метод синтезу поліноміальних нейронних мереж для вирішення задач прогнозування нестационарних часових рядів, що є альтернативою багатослойним перцептронам та радіально-базисним нейронним мережам використання яких обмежене при вирішенні багатьох практичних задач. Запропонований метод простий з точки зору чисельної реалізації і дозволяє ускладнювати архітектуру нейронної мережі без необхідності перерахунку вже налаштованих синаптичних ваг

Ключові слова: прогнозуюча модель, поліноміальна ортогональна нейронна мережа, поліноми Чебишева, орто-синапс, синаптичні ваги

Предложен метод синтеза полиномиальных нейронных сетей для решения задач прогнозирования нестационарных временных рядов, являющихся альтернативой многослойным перцептронам и радиально-базисным нейронным сетям, использование которых ограничивается при решении многих практических задач. Предложенный метод прост с точки зрения численной реализации и позволяет усложнять архитектуру нейронной сети без необходимости пересчета уже настроенных синаптических весов

Ключевые слова: прогнозирующая модель, полиномиальная ортогональная нейронная сеть, полиномы Чебишева, орто-синапс, синаптические веса

АДАПТИВНАЯ ПОЛИНОМИАЛЬНАЯ НЕЙРОСЕТЕВАЯ ПРОГНОЗИРУЮЩАЯ МОДЕЛЬ ВРЕМЕННЫХ РЯДОВ И ЕЕ ОБУЧЕНИЕ

Е. В. Мантула

Аспирант*

E-mail: elenamantula@gmail.com

С. В. Машталир

Кандидат технических наук, доцент*

E-mail: zerginio.m@gmail.com

*Кафедра информатики

Харьковский национальный

университет радиоэлектроники

пр. Ленина, 16, г. Харьков, Украина, 61166

1. Введение

Задача прогнозирования временных рядов часто встречается во многих приложениях, а для ее решения разработан целый арсенал методов – от простейших эмпирических до весьма сложных, основанных на интеллектуальном подходе, при этом инструментом прогнозирования в итоге является математическая модель – от простейшей регрессионной до сложной нейросетевой [1, 2].

Возможность использования того или иного метода определяется объемом априорной информации о решаемой задаче, при этом задача существенно усложняется такими факторами как нелинейность, нестационарность, стохастичность, хаотичность изучаемого явления. В этих случаях предыстория не может быть использована для нахождения параметров модели [3].

Интеллектуальные нейронные сети, используемые в подобных случаях, могут обеспечивать высокое качество предсказания. Однако, хотя они, обладая универсальными аппроксимирующими свойствами в процессе обучения, способны с заданной точностью восстановить практически любую нелинейную функцию, заданную набором наблюдений, их практическое использование ограничивается потребностью для настройки своих параметров иметь достаточно большую по объему обучающую выборку. И дело здесь не в том, что имеющихся в распоряжении исследователя наблюдений недостаточно. Просто в практических задачах временные ряды настолько часто меняют свои свой-

ства, что нейронная сеть просто не успевает обучиться на отдельных стационарных участках временного ряда [4].

В связи с этим возникает необходимость построения на основе нейросетевого подхода упрощенных прогнозирующих моделей, которые имели бы высокую скорость обучения в условиях ограниченной обучающей выборки, а также в случаях, когда данные на обработку подаются через произвольные заранее неизвестные интервалы времени, позволяющие усложнять архитектуру нейронной сети и обладающие возможностью перестраивания своих характеристик непосредственно по ходу решения задачи.

В данной работе предлагается адаптивная полиномиальная нейросетевая прогнозирующая модель и процедура ее обучения в условиях ограниченного объема априорной информации, обладающая высокими аппроксимирующими и экстраполирующими свойствами.

2. Литературный обзор и постановка проблемы

На сегодняшний день существует достаточно много математических методов решения задачи прогнозирования нестационарных временных рядов. К данным методам можно отнести различные классические статистические методы, применяющиеся для анализа стационарных процессов. К данным методам можно отнести регрессионные, корреляционные, спектраль-

ные, метод Бокса-Дженкинса. Однако к их недостаткам можно отнести тот факт, что они ограничены для анализа и моделирования непредсказуемых скачков, поэтому их целесообразно использовать для краткосрочных прогнозов. Также для решения данной задачи используются адаптивные методы, основанные на экспоненциальном сглаживании и интеллектуальные [5].

Исходной информацией для построения математической прогнозирующей модели является временной ряд, который описывает изучаемый показатель. При этом задача синтеза модели существенно усложняется нестационарностью показателей, высоким уровнем стохастичности или хаотичности, нелинейностью и тому подобными факторами [6].

В реальных задачах часто возникает случаи, когда исходная выборка наблюдения слишком мала по объему, или когда прогнозируемый процесс содержит как нерегулярные тренды, так и внезапные скачки. В этих случаях предыстория не может быть использована для нахождения параметров модели.

Еще более сложная ситуация возникает, когда наблюдения распределены на временной шкале неравномерно, т.е. квантование контролируемого процесса происходит с переменным шагом.

В этом случае спектр возможных подходов становится еще меньшим [7, 8].

Наиболее часто для решения подобных задач прогнозирования используются интеллектуальные нейронные сети - многослойный перцептрон и его модификации, лежащие в основе рекуррентных нейронных сетей, а также радиально-базисные ИНС [1].

Однако несмотря на то, что эти сети обеспечивают высокое качество предсказания, они обладают целым рядом недостатков, ограничивающих их использование при решении многих практических задач. Так, многослойные перцептроны, обучаемые на основе обратного распространения ошибок, достаточно медленно настраивают свои параметры и требуют больших объемов обучающей выборки.

Радиально-базисные ИНС обучаются с высокой скоростью благодаря тому, что их выходной сигнал линейно зависит от настраиваемых синаптических весов, однако наличие большого количества свободных параметров полей радиально-базисных активационных функций вынуждает решать дополнительные задачи кластеризации, а кроме того не следует забывать о достаточно часто возникающем в реальных задачах «проклятии размерности» [9].

Достойной альтернативой многослойным и радиально-базисным ИНС являются полиномиальные нейронные сети [2], использующие в качестве активационных функций те или иные полиномы и имеющие целый ряд преимуществ перед традиционными сетями: простоту с точки зрения численной реализации, высокую скорость обучения в условиях ограниченной обучающей выборки, а также в случаях, когда данные на обработку

подаются через произвольные заранее неизвестные интервалы времени, а архитектура нейронной сети может быть усложнена без необходимости пересчета уже настроенных синаптических весов.

В связи с этим в настоящей работе предлагается подход к синтезу адаптивных прогнозирующих моделей, который производится в условиях ограниченной обучающей выборки, при этом данные на обработку могут подаваться через произвольные заранее неизвестные интервалы времени.

3. Ортогональная полиномиальная искусственная нейронная сеть

В качестве активационных полиномиальных функций весьма перспективным представляется использование ортогональных полиномов, которые лежат в основе так называемых, ортогональных ИНС [3 – 5], основным преимуществом которых является простота обучения на основе метода наименьших квадратов с диагональной ковариационной матрицей.

Среди множества возможных ортогональных полиномов наиболее эффективными представляются полиномы Чебышева [6], обладающие целым рядом полезных свойств, связанных с традиционным в обучении ИНС квадратичным критерием.

На рис. 1 приведена архитектура полиномиальной нейронной сети, реализующую прогнозирующую модель нелинейной авторегрессии с экзогенными входами (NARX - модель) вида:

$$\hat{y}(k) = f(y(k-1), \dots, y(k-n_A), x(k-1), \dots, x(k-n_B))$$

или, переобозначая переменные,

$$\hat{y}(k) = f(z_1(k), \dots, z_{n_A}(k), z_{n_A+1}(k), \dots, z_p(k), \dots, z_{n_A+n_B}(k)),$$

где $\hat{y}(k)$ – прогнозируемая временная последовательность в момент дискретного времени $k=1,2,\dots$; $\hat{y}(k)$ – ее прогноз по имеющимся на $(k-1)$ -й момент реальным наблюдением; $x(k-1), \dots, x(k-n_B)$ – значения экзогенной (внешней) переменной, определяющие поведение прогнозируемого ряда; n_A, n_B – глубина используемой предыстории; $z_p(k) = y(k-p)$, если $p \leq n_A$ и $z_p(k) = x(k-p)$ при $p > n_A$.

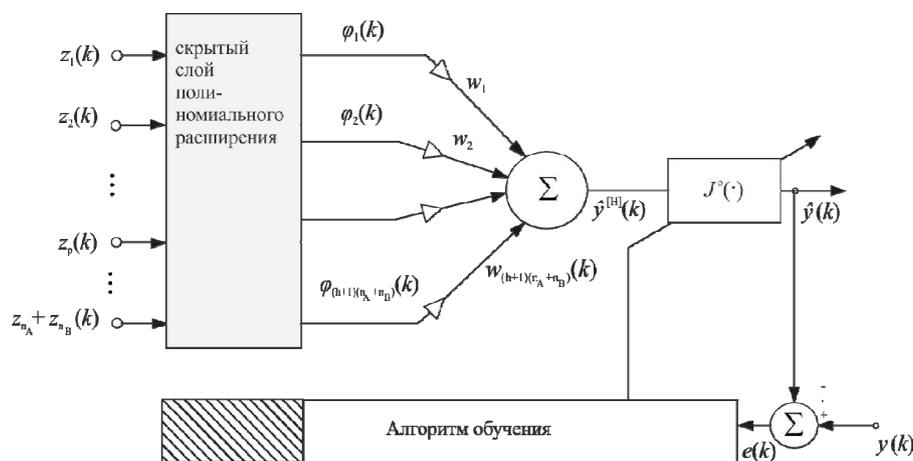


Рис. 1. Полиномиальная ортогональная нейронная сеть

Входные сигналы $z_1(k), \dots, z_p(k), \dots, z_{n_A+n_B}(k)$ с рецепторного слоя подаются на первый скрытый слой сети, именуемый слоем полиномиального расширения, где реализуется нелинейное чебышевское преобразование

$$\begin{aligned} f_0(z_p(k)) &= 1, \\ f_1(z_p(k)) &= z_p(k), \\ f_2(z_p(k)) &= 2z_p^2(k), \\ f_h(z_p(k)) &= 2z_p(k)f_{h-1}(z_p(k)) - f_{h-2}(z_p(k)), \end{aligned} \tag{1}$$

на основе которого входные сигналы $z_p(k)$, $p=1, 2, \dots, n_A+n_B$ преобразуются так, что на выходах первого скрытого слоя формируются сигналы $\phi_1(k) = f_0(z_1(k)), \dots, \phi_{(h+1)(n_A+n_B)}(k) = f_h(z_{n_A+n_B}(k))$.

Второй скрытый слой образован адаптивным линейным ассоциатором [1] с $(h+1)(n_A+n_B)$ входами и таким же количеством настраиваемых синаптических весов $w = (w_1, w_2, \dots, w_{(h+1)(n_A+n_B)})^T$. На выходе второго скрытого слоя в результате формируется скалярный сигнал

$$\hat{y}^{[H]}(k) = \sum_{l=1}^{(h+1)(n_A+n_B)} w_l \phi_l(k) = w^T \phi(k),$$

где $\phi(k) = (\phi_1(k), \phi_2(k), \dots, \phi_{(h+1)(n_A+n_B)}(k))^T$.

В ортогональной нейронной сети, введенной в [5], вместо адаптивного линейного ассоциатора предлагается использовать элементарный перцептрон Розенблатта с фиксированной активационной функцией гиперболического тангенса

$$\hat{y}(k) = \varphi^0(\hat{y}^{[H]}(k)) = \tanh(\hat{y}^{[H]}(k)).$$

Понятно, что единственная активационная функция не может обеспечить никакие дополнительные аппроксимирующие свойства.

Сама же кривая гиперболического тангенса может выполнять только роль ограничителя

$$-1 < \hat{y}(k) < 1.$$

При этом, при произвольном нелинейном преобразовании $\varphi^0(\circ)$, нейронная сеть в целом производит отображение

$$\hat{y}(k) = \varphi^0(w^T \phi(k)) = \varphi^0 \left(\sum_{l=1}^{(h+1)(n_A+n_B)} w_l \phi_l(k) \right). \tag{2}$$

Для улучшения аппроксимирующих и экстраполирующих свойств сети мы предлагаем формировать выходной слой не на основе единственной фиксированной активационной функции, а на основе так называемого орто-синапса [6], схема которого приведена на рис. 2.

В качестве активационной функции здесь используются те же полиномы Чебышева (1).

Кроме того, имеется $h+1$ настраиваемых синаптических весов, что улучшает качество решения рассматриваемой задачи.

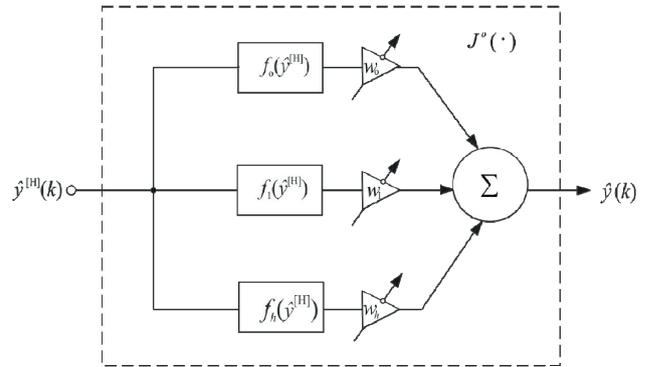


Рис. 2. Орто-синапс выходного слоя полиномиальной нейронной сети

Преобразование, реализуемое орто-синапсом, может быть записано в виде

$$\hat{y}(k) = \sum_{j=0}^h w_j^0 f_j^0(\hat{y}^{[H]}(k)), \tag{3}$$

а отображение (2), реализуемое сетью в целом:

$$\hat{y}(k) = \sum_{j=0}^h w_j^0 f_j^0 \left(\sum_{l=1}^{(h+1)(n_A+n_B)} w_l \phi_l(k) \right). \tag{4}$$

Такая сеть содержит $(h+1)(n_A+n_B) + h+1$ настраиваемых в процессе обучения синаптических весов.

4. Обучение ортогональной полиномиальной искусственной нейронной сети

Обучение сети производится на основе обратного распространения ошибок [1], при этом сначала настраиваются веса $w^0 = (w_0^0, \dots, w_h^0)$ выходного слоя, а затем вектор весов w второго скрытого слоя.

Первый скрытый слой сети является непараметрическим и не содержит настраиваемых параметров.

Для настройки вектора весов выходного слоя w^0 целесообразно воспользоваться оптимальными по быстродействию алгоритмами, например, аддитивно-мультипликативной модификацией алгоритма Качмажа [9 – 12].

$$\begin{aligned} w^0(k) &= w^0(k-1) + \gamma \frac{y(k) - \hat{y}(k)}{\beta + \|F(k)\|^2} F(k) = \\ &= w_0(k-1) + \gamma \frac{e(k)F(k)}{\beta + \|F(k)\|^2} = \\ &= w_0(k-1) + \gamma \frac{y(k) - w^{0T}(k-1)F(k)}{\beta + \|F(k)\|^2} F(k), \end{aligned} \tag{5}$$

где $F(k) = (f_0^0(\hat{y}^{[H]}(k)), f_1^0(\hat{y}^{[H]}(k)), \dots, f_h^0(\hat{y}^{[H]}(k)))^T$, $0 < \gamma < 2$, $\beta \geq 0$ – параметры алгоритма, выбираемые из эмпирических соображений.

Здесь важно отметить, что, поскольку выходной сигнал сети $\hat{y}(k)$ линейно зависит от вектора синаптических весов выходного слоя w^0 , их настройка может быть произведена максимально быстро.

Поскольку в реальных задачах анализируемый ряд $y(k)$ «загрязнен» различными возмущениями и по-

мехами, для их фильтрации вместо одношагового алгоритма (5) можно воспользоваться многошаговыми процедурами типа метода наименьших квадратов на скользящем окне.

Тогда оценка $w^0(k)$ на скользящем окне из s наблюдений может быть записана в виде:

$$w^0(k) = \sum_{\alpha=k-s+1}^k F(\alpha)F^T(\alpha)^{-1} \sum_{\alpha=k-s+1}^k F(\alpha)y(\alpha) \quad (6)$$

или в рекуррентной форме:

$$\begin{cases} P_s(k) = P_s(k-1) - \frac{P_s(k-1)F(k)F^T(k)P_s(k-1)}{1+F^T(k)P_s(k-1)F(k)}, \\ P_s(k) = P(k) + \frac{P(k)F(k-s)F^T(k-s)P(k)}{1-F^T(k-s)P(k)F(k-s)}, \\ p_s(k) = p_s(k-1) + F(k)y(k) - F(k-s)y(k-s), \\ w^0(k) = P_s(k)p_s(k). \end{cases} \quad (7)$$

Синаптические веса второго скрытого слоя в описании сети (4) входят нелинейно, следовательно, для их настройки необходимо использовать алгоритмы, основанные на обратном распространении ошибок.

Введем в рассмотрение целевую функцию обучения

$$\begin{aligned} J(k) &= \frac{1}{2} e^2(k) = \frac{1}{2} (y(k) - \hat{y}(k))^2 = \\ &= \frac{1}{2} (y(k) - \sum_{j=0}^h w_j^0 f_j^0(\hat{y}^{[H]}(k)))^2 = \\ &= \frac{1}{2} (y(k) - w^{0T} F(k))^2 = \\ &= \frac{1}{2} (y(k) - \sum_{j=0}^h w_j^0 f_j^0(\sum_{l=1}^{(h+1)(n_A+n_B)} w_l \varphi_l(k)))^2 = \\ &= \frac{1}{2} (y(k) - w^{0T} F(w^T \varphi(k)))^2, \end{aligned}$$

которая может быть минимизирована по w_1 с помощью градиентного алгоритма

$$\begin{aligned} w_1(k) &= w_1(k-1) - \eta(k) \frac{\partial J(k)}{\partial w_1} = \\ &= w_1(k-1) + \eta(k) e(k) \frac{\partial \hat{y}(k)}{\partial w_1} = \\ &= w_1(k-1) + \eta(k) e(k) \sum_{j=0}^h w_j^0(k) \frac{\partial f_j^0(\hat{y}^{[H]}(k))}{\partial \hat{y}} \varphi_1(k), \end{aligned}$$

или в векторной форме:

$$w(k) = w(k-1) + \eta(k) e(k) w^{0T}(k) \left(\frac{\partial \bar{F}(k)}{\partial \hat{y}} \right) \varphi(k), \quad (8)$$

где $\eta(k) > 0$ – параметр шага обучения, $\left(\frac{\partial \bar{F}(k)}{\partial \hat{y}} \right)$ – $(h+1) \times 1$ -вектор, образованный частными производными $\left(\frac{\partial f_j^0(\hat{y}^{[H]}(k))}{\partial \hat{y}} \right)$.

Алгоритм (8) может быть также переписан в более привычной в обучении нейронных сетей форме δ -правила обучения [1]:

$$w(k) = w(k-1) + \eta(k) \delta(k) J(k), \quad (9)$$

где $\delta(k) = e(k) w_0^T(k) \left(\frac{\partial \bar{F}(k)}{\partial \hat{y}} \right)$ – так называемая локальная ошибка (δ -ошибка) обучения многослойной сети.

Необходимо отметить также, что в силу нелинейной зависимости выхода сети $y(k)$ от синаптических весов $w(k)$ второго скрытого слоя, скорость сходимости алгоритма (7) существенно ниже, чем скорость процедуры (5), поэтому для обучения нейронной сети в целом можно использовать обучение по эпохам, применяемое в обучении многослойных сетей.

Тем не менее, поскольку обучается только один скрытый слой, рассматриваемая нейросетевая модель все равно настраивается быстрее, чем стандартный трехслойный персептрон.

Данная методика значительно упрощает построение прогнозирующей модели на основе нейросетевого подхода, поскольку для обучения требуется минимальный по объему набор данных, что особенно полезно для решения таких весьма специфических задач, как, например, проблема контроля и прогнозирования показателей загрязнения окружающей среды, где сложность заключается в том, что временные ряды настолько часто меняют свои свойства, что нейронная сеть просто не успевает обучиться на отдельных стационарных участках временного ряда.

5. Выводы

1. В настоящей работе рассмотрена задача прогнозирования нестационарных нелинейных временных рядов в условиях ограниченного объема априорной информации.

2. Для решения данной задачи предложен подход, основанный на синтезе полиномиальных нейронных сетей, который является альтернативой многослойным персептронам и радиально-базисным нейронным сетям, имеющих ряд недостатков, ограничивающих их использование при решении различных практических задач. К преимуществам данного подхода перед традиционными нейронными сетями можно отнести вычислительную простоту и высокую скорость обучения в условиях ограниченной обучающей выборки, а также в случаях, когда данные на обработку подаются через произвольные заранее неизвестные интервалы времени. Подход позволяет усложнять архитектуру нейронной сети без необходимости пересчета уже настроенных синаптических весов.

3. Предложена процедура обучения данной нейронной сети, основанная на использовании ортогональных полиномов в качестве активационных функций и базирующаяся на обучении по эпохам (данный вид обучения используется для многослойных сетей). Именно поскольку обучается только один скрытый слой, рассматриваемая нейросетевая модель настраивается быстрее, чем стандартный трехслойный персептрон.

Література

1. Хайкин, С. Нейронные сети: полный курс [Текст] / С. Хайкин. – М.: Изд. дом «Вильямс», 2006. – 1104 с.
2. Pao, Y. H. Adaptive Pattern Recognition and Neural Networks [Text] / Y. H. Pao. – Reading, MA: Addison-Wesley, 1989 – 320 p.
3. Yang, S.-S. An orthonormal neural network for function approximation [Text] / S.-S. Yang, C.-S. Tseng // IEEE Transactions on Systems, Man, and Cybernetics. – 1996. – Vol. 26, № 12. – P. 925–935.
4. Lee, T. T. The Chebyshev polynomial-based unified model neural networks for function approximation [Text] / T. T. Lee, J. T. Jeng // IEEE Transactions on Systems, Man, and Cybernetics. – 1998. – Vol. 28, № 12. – P. 925–935.
5. Patra, J. C. Nonlinear dynamic system identification using Chebyshev functional link artificial neural networks [Text] / J. C. Patra, A. C. Kot // IEEE Transactions on Systems, Man, and Cybernetics. – 2002. – Vol. 32, №4. – P. 505–511.
6. Бодянский, Е. В. Искусственные нейронные сети: архитектуры, обучение, применение [Текст] / Е. В. Бодянский, О. Г. Руденко // Харьков. ТЕЛТЕХ, 2004. – 372 с.
7. Бидюк, П. И. Методы прогнозирования [Текст]: Т. 1 / П. И. Бидюк, О. С. Меняйленко, О. С. Половцев. – Луганск: Альма-матер, 2008 – 301 с.
8. Бидюк, П. И. Методы прогнозирования [Текст]: Т. 2 / П. И. Бидюк, О. С. Меняйленко, О. С. Половцев. – Луганск: Альма-матер, 2008 – 305 с.
9. Райбман, Н. С. Построение моделей процессов производства [Текст] / Н. С. Райбман, В. М. Чадеев. – М.: Энергия, 1975. – 376 с.
10. Бодянский, Е. В. Ортоинакс, ортонейроны и нейропредиктор на их основе [Текст] / Е. В. Бодянский, Е. А. Викторов, А. Н. Слипченко // Системи обробки інформації. – 2007. – Вип. 4 (62). – С. 139–143.
11. Бодянский, Е. В. Субоптимальное управление стохастическими процессами [Текст] / Е. В. Бодянский, С. Г. Удовенко, А. Е. Ачкасов, Г. К. Вороновский. – Харьков: Основа, 1997. – 140 с.
12. Перельман, И. И. Оперативная идентификация объектов управления [Текст] / И. И. Перельман. – М.: Энергоатомиздат, 1982. – 272 с.

В роботі представлено загальний метод кластеризації об'єктів, що використовує нечіткі бінарні відношення для визначення міри близькості векторів ознак об'єктів за «кутовою» та «довжинною» напівметриками. Даний метод реалізований у вигляді трьох алгоритмів. Програмна реалізація даного методу показала його ефективність при розв'язанні різних прикладних задач та простоту в застосуванні

Ключові слова: кластерний аналіз, кластер, нечіткі бінарні відношення, розбиття об'єктів, кластеризація об'єктів

В работе представлено общий метод кластеризации объектов, использующий нечеткие бинарные отношения для определения меры близости векторов признаков объектов по «угловой» полуметрике и полуметрике длины. Данный метод реализован в виде трех алгоритмов. Программная реализация данного метода показала его эффективность при решении различных прикладных задач и простоту в применении

Ключевые слова: кластерный анализ, кластер, нечеткие бинарные отношения, разбиение объектов, кластеризация объектов

УДК 519.8

ДЕЯКІ МЕТОДИ АВТОМАТИЧНОГО ГРУПУВАННЯ ОБ'ЄКТІВ

Н. Е. Кондрук

Кандидат технічних наук, доцент
Кафедра кібернетики і
прикладної математики
Ужгородський
національний університет
пл. Народна, 3, м. Ужгород,
Україна, 88000
E-mail: kondrukne@gmail.com

1. Вступ

В останні десятиліття спостерігається ріст інтересу до нового напрямку в обробці інформації – інтелектуальному аналізу даних (Data Mining).

В запропонованій роботі розглядається часткова задача інтелектуального аналізу даних – задача

кластерного аналізу, відома як задача автоматичного групування об'єктів, класифікації без учителя або таксономії.

Кластерний аналіз (англ. Data clustering) – задача розбиття заданої вибірки об'єктів на підмножини (кластери), так, щоб кожен кластер складався з схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися.