

САМООРГАНИЗУЮЩИЕСЯ ОДНО И ДУМЕРНЫЕ КАРТЫ КОХОНЕНА ДЛЯ КЛАСТЕРИЗАЦИЯ НЕСТРУКТУРИРОВАННЫХ ДАННЫХ

К.т.н. Е.А Арсирій, А.П. Василевська

Одесский национальный политехнический университет
Украина, г. Одесса
sauterelle@mail.ru

Сложность анализа бизнес процессов определяется неструктурированностью исходных данных. К неструктурированным относятся данные (тексты, графика, мультимедиа) представленные в произвольной форме. Предварительным этапом для анализа таких данных является их структуризация — упорядочение данных. Кластеризация — процедура группировки объектов по схожести их признаков либо определения метки кластера для объекта, как одна из процедур интеллектуального анализа данных широко применяется в бизнес-аналитике для структурирования данных, сокращения их объема, обнаружения аномалий, а также в качестве предварительной обработки при моделировании поведения экономических объектов [1]. В работе предлагается использовать нейросетевой подход основанный на использовании одно и двумерных самоорганизующихся карт Кохонена для решения задачи кластеризации. При этом одномерная карта (слой) Кохонена представляет собой структуру неупорядоченных нейронов с конкурирующей функцией активации, а двумерная карта состоит из упорядоченных (прямоугольная или гексагональная решетка) конкурирующих нейронов, которые отражают структуру данных таким образом, что близким кластерам на карте соответствуют близко расположенные нейроны [2].

На рис 1 показаны результаты использования слоя Кохонена в NNTool Matlab состоящего из 5 нейронов для кластеризации неструктурированных данных изображений гидродинамических потоков [3]. На рис 2 показан результат структуризации (выделения признаков) неструктурированных данных для типичных экземпляров классов изображений

гидродинамических потоков.

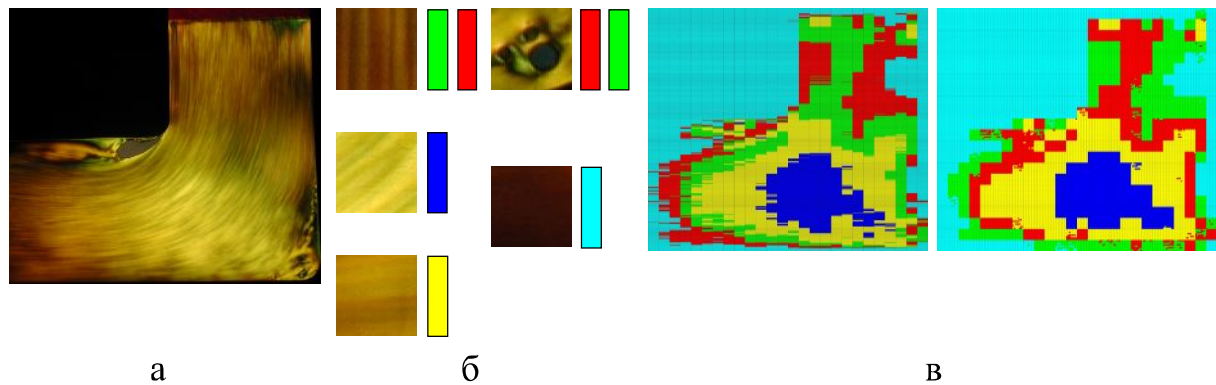


Рис 1 – Кластеризации изображений гидродинамических потоков (а – исходное изображение, б – изображения экземпляров классов с соответствующими им цветовой разметкой, в – результирующее изображение)

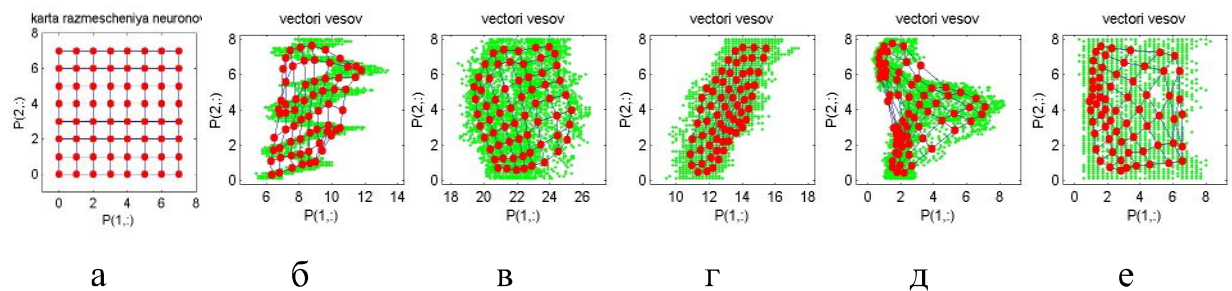


Рис 2 – Структуризация изображений экземпляров классов (а – расположение нейронов в исходной карте, б,в,г,д,е – расположение нейронов в картах экземпляров классов)

Предлагаемое использование слоя Кохонена позволяет структурировать за счет кластеризации неструктурированные данные и сократить их объем в 50 раз. Использование самоорганизующейся карты Кохонена в качестве предварительной обработки с выделением признаков позволяет сократить объем исходных данных в 64 раза.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Паклин Н.Б., Орешков В.И. Бизнес аналитика: от данных к знаниям – С.: Питер, 2010 – 704с.
2. Медведев В.С., Потемкин В.Г. Нейронные сети MATLAB 6 – М.: ДИАЛОГ-МИФИ, 2002 – 496с.
3. Арсирый В.А., Арсирый Е.А., Власенко В.А. Метод и информационная технология визуализации структур гидродинамических потоков // Труды Одесского политехнического университета. – 1997. – Вып.1. – С. 242–247.