

ТАБЛИЧНАЯ РЕАЛИЗАЦИЯ В EXCEL НЕЙРОННОЙ СЕТИ КОХОНЕНА

В.И. Аникин, А.А. Карманова

Кафедра информационного и электронного сервиса, Поволжский Государственный Университет Сервиса
ООО «НетКрэкер»

Тольятти, Российская Федерация

E-mail: anikin_vi@mail.ru, turaeva.alexandra@gmail.com

В работе рассматривается табличная реализация в среде Microsoft Excel самоорганизующейся карты Кохонена, построенной посредством авторской оригинальной технологии алгоритмического табличного моделирования. Доказано, что электронные таблицы можно использовать в качестве удобной и эффективной среды алгоритмического моделирования нейронных систем.

ВВЕДЕНИЕ

Последние десятилетия ИНС Кохонена представляет собой известный и широко используемый в различных прикладных областях алгоритм кластеризации данных.

В статье описывается использование итерационной табличной модели среды Excel для реализации сети Кохонена [1]. Обучение сети происходит в пакетном режиме посредством клеточного автомата [2] с использованием технологии алгоритмического табличного моделирования [3].

Обучение с помощью клеточного автомата всегда выполняется на максимальной скорости обучения, что приводит к сокращению времени обучения сети в несколько десятков раз.

ТАБЛИЧНАЯ РЕАЛИЗАЦИЯ СЕТИ КОХОНЕНА

Интерфейс управления структурой обучающего клеточного автомата содержит 30 связанных нейронов на двумерной решетке размером 5*6 нейронов. Пользователь в соответствии с нужной структурой клеточного автомата (одномерный клеточный автомат, двумерный клеточный автомат с окружением Мура, многосвязный клеточный автомат) активирует нужные нейроны; остальные нейроны не используются.

Особенностью алгоритма обучения нейронной сети Кохонена клеточным автоматом в отличие от классического алгоритма является небольшое число настраиваемых параметров, к которым относятся следующие:

- число прогонов – число эпох обучения нейронной сети; в каждой эпохе учебные образцы подаются на вход нейронной сети последовательно в порядке возрастания их номеров;
- скорость обучения – параметр, регулирующий скорость обучения сети Кохонена клеточным автоматом; обычно этот параметр равен 1, однако замечено, что его уменьшение до величины 0.5..0.7 благоприятствует лучшей самоорганизации нейронов сети (в ущерб скорости обучения);
- количество эпох взаимодействия – доля эпох обучения, в течение которых связи

между соседними нейронами сети включены.

Соответственно процесс обучения нейронной сети состоит из двух этапов: на первом этапе (фаза взаимодействия нейронов) сеть обучается клеточным автоматом, на втором этапе связи между соседними нейронами разрываются, и сеть дообучается по алгоритму WTA (Winner Takes All – победитель забирает все). Длительность первой фазы обучения подбирается таким образом, чтобы финитная конфигурация клеточного автомата/нейронов сети достигла устойчивого состояния.

Табличная модель сети Кохонена реализована на четырех рабочих листах Excel. На рабочем листе «Выборка» содержится многомерная выборка учебных образцов, координаты которой нормированы к интервалу [-1,1]. На рабочем листе «Обучение» реализован алгоритм обучения сети Кохонена с помощью клеточного автомата. Рабочий лист «Кластеризация» содержит результаты кластеризации образов обученной сетью Кохонена. На рабочем листе «Обработка» выполняются анализ и статистическая обработка полученных результатов кластеризации.

Процесс обучения ИНС Кохонена в Excel включает следующие последовательные этапы:

- настройка параметров обучения через пользовательский интерфейс табличной модели сети, таких как структура сети, скорость обучения, доля эпох взаимодействия, число прогонов, конфигурация обучающего клеточного автомата, начальное состояние нейронов сети;
- запуск процесса обучения сети в течение заданного числа прогонов или ее пошаговое обучение;
- фиксация весов нейронов обученной сети в случае необходимости или приведение нейронной сети в исходное необученное состояние.

С кнопками ассоциированы простые универсальные макросы. Эти макросы универсальны в том смысле, что они предназначены исключительно

для управления моделью и с небольшими вариациями встречаются во всех итерационных табличных моделях, построенных по технологии алгоритмического табличного моделирования [4].

Новый детерминированный алгоритм обучения сети Кохонена с помощью клеточного автомата продемонстрировал способность к распознаванию структуры многомерных данных.

Табличная реализация сети Кохонена позволила выполнить процесс самообучения сети и распределить близкие по своим входным характеристикам образцы по отдельным кластерам. Средние значения входных параметров по выделенным кластерам совпадают с абсолютными координатами центров кластеров, что говорит о нахождении правильных координат центров кластеров, соответствующих физическим центрам тяжести кластеров.

С помощью механизма сводных таблиц для каждого из кластеров определены описательные характеристики — средние, минимальные, максимальные значения, несмещенная дисперсия и несмещенное отклонение входных параметров в целом по выборке и в разрезе кластеров.

Сводная таблица является одним из инструментов для проведения визуального анализа кластеров. В работе была использована такая возможность Excel как условное форматирование, которое с помощью выбранной цветовой шкалы автоматически разбивает выделенный диапазон на нужное количество равных интервалов и окрашивает ячейку для каждого значения цветом интервала, в которое попало число.

Отсортировав кластеризованную исходную выборку по номеру кластера и применив условное форматирование к каждому из столбцов измерений, мы получаем информацию о каждом образце исходной выборки. Условное форматирование в Excel позволяет получить так называемый атлас, хранящий в себе визуальную информацию о зависимостях между входными параметрами для каждого образца и кластеров в целом.

ЗАКЛЮЧЕНИЕ

Причинами выбора табличного процесса стали удобство реализации адаптированного процесса обучения нейронной сети, гибкость и высокая скорость обучения.

Использование Excel позволяет любому пользователю табличного редактора без глупо-

ких знаний языка программирования детально изучить в работе как стандартные алгоритмы нейронных сетей, в том числе сети Кохонена, и их модификации, так и реализовать собственные идеи. Принципиально важным здесь является то, что сам алгоритм обучения ИНС Кохонена реализуется чисто табличными средствами без написания программного кода на языке VBA и использования пользовательских функций, что является нетривиальной задачей. Стоит заметить, что пользовательские функции работают в Excel значительно медленнее встроенных функций, поэтому их отсутствие в табличной модели нейронной сети Кохонена позволило нам достичь максимального быстродействия в процессе обучения последней.

На рынке программного обеспечения наиболее известны такие среды по работе с нейронными сетями как STATISTICA [5] и Deductor [6]. Эти продукты являются платными и закрытыми для изменений, в них нет возможностей по оптимизации используемых алгоритмов, и модификация стандартного алгоритма сети Кохонена в них невозможна.

Технология табличного моделирования может быть основой для поиска решения с применением нейронных сетей без необходимости использования специальных программных сред во многих прикладных областях. Табличные модели в настоящее время широко распространены и представляют собой экономически выгодную альтернативу для целей обучения моделированию и использованию техник принятия решений.

1. Кохонен, Т. Самоорганизующиеся карты / Т. Кохонен. Пер. 3-го англ. изд. – М.: БИНОМ. Лаборатория знаний, 2008. – 655 с.: ил.
2. Аникин, В. И. О возможности обучения искусственной нейронной сети Кохонена с помощью клеточного автомата / В. И. Аникин, А. А. Тураева. Вектор науки ТГУ, 2011. – № 3(17). – с. 22-24.
3. Аникин, В. И. Алгоритмическое табличное моделирование в Microsoft Excel: итерационные модели / В. И. Аникин, О. В. Аникина. – «Информатика и образование», 2009. – № 9, с. 88-95.
4. Аникин, В. И. Визуальное табличное моделирование клеточных автоматов в Microsoft Excel: монография / В. И. Аникин, О. В. Аникина. – Тольятти: Изд. –во ПВГУС., 2013. – 324 с.
5. Академия Анализа Данных StatSoft [Электронный ресурс] / Режим доступа: www.statsoft.ru.
6. Deductor Аналитическая платформа [Электронный ресурс] / Режим доступа: www.basegroup.ru.