

77-48211/429076 Разработка системы анализа биологических показателей человека на основе нейросетевых технологий

03, март 2012

авторы: Андрианов Д. А., Гаврилов А. И.

УДК 62-50: 681.513.00.004.8

Одним из наиболее эффективных средств системного анализа являются нейронные сети. Они способны обучаться на основе данных, извлекать существенные характеристики и обобщать данные. Длительный период эволюции придал мозгу человека много качеств, которые отсутствуют как в машинах с архитектурой фон Неймана, так и в современных параллельных компьютерах. К ним относятся:

- массовый параллелизм,
- распределенное представление информации и вычисления,
- способность к обучению и способность к обобщению,
- адаптивность,
- свойство контекстуальной обработки информации,
- толерантность к ошибкам,
- низкое энергопотребление.

Можно предположить, что приборы, построенные на тех же принципах, что и биологические нейроны, будут обладать перечисленными характеристиками.

Современные цифровые вычислительные машины превосходят человека по способности производить числовые и символьные вычисления. Однако человек может без усилий решать сложные задачи восприятия внешних данных (например, узнавание человека в толпе только по его промелькнувшему лицу) с такой скоростью и точностью, на которую не способен даже самый современный компьютер.

Одной из наиболее перспективных технологий, применяемых в настоящее время для решения задач интеллектуального анализа данных и обработки информации, является технология, основанная на применении искусственных нейронных сетей (ИНС), известных как самоорганизующиеся карты (Self - Organizing Feature Maps) Кохонена.

Самоорганизующиеся карты являются мощным нейросетевым средством анализа и визуализации многомерных данных. Они используются для отображения нелинейных статистических взаимосвязей на легко интерпретируемые (обычно двумерные) решетки, подчеркивающие топологические и метрические зависимости анализируемых данных.

Стратегия метода построения SOM заключается в представлении множества экспериментальных данных в виде топологической карты – решетки размерностью $n \times m$, состоящей из нейронов. Каждый нейрон представляет собой вектор, определяющий в d – мерном пространстве признаков некоторую характерную точку. В результате реализации алгоритма обучения вектора-нейроны распределяются в пространстве признаков, аппроксимируя статистические взаимосвязи между данными обучающего множества.

Рассмотрим общую форму алгоритма обучения:

Пусть задан набор векторов $M = \{m_i\}$ $i = \overline{1, K}$, где m_i — “нейроны” карты (синаптические весовые вектора), он задан на некоторой регулярной сетке, задающей топологию, на которую будут отображаться данные выборки.

Введем в пространстве векторов меру близости $d(x, y)$. Выбирая из обучающего множества вектора данных, найдем вектор-нейрон карты m_c , для которого выполняется условие $d(x_j, m_c) = \min(d(x_j, m_i))$. Этот нейрон и будет, пользуясь терминологией нейронных сетей, нейроном-“победителем” (в литературе по SOM, как правило, используется специальный термин “Best matching unit” — BMU).

Произведем настройку нейронов карты по рекурсивной формуле: $m_i(t+1) = m_i + \alpha(t)h_{ci}(t)[x(t) - m_i(t)]$, где $h_{ci}(t)$ — симметричная невозрастающая функция, определенная вокруг m_c , называемая “функцией близости” (она определяет степень влияния нейрона-“победителя” на соседние нейроны), $\alpha(t)$ — коэффициент обучения (убывающая функция времени), причем $0 < \alpha(t) < 1$.

Таким образом, на каждом шаге алгоритма обучения определяется ближайший (в силу заданной метрики) к поступившим на вход сети данным нейрон, который «подтягивается» к входным данным вместе со своими соседями, определяемыми функцией близости $h_{ci}(t)$. Графическая интерпретация функционирования алгоритма обучения представлена на рис.1.

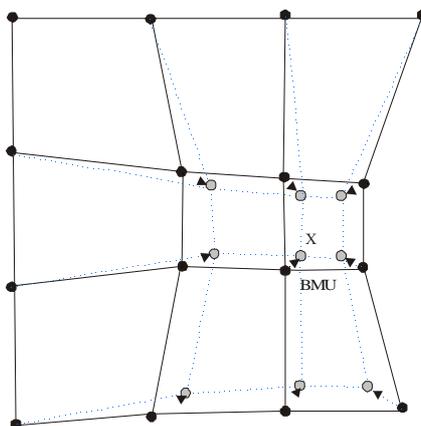


Рис.1. Графическая интерпретация основной идеи алгоритма обучения SOM

Метод построения модели системы на основе SOM заключается в представлении множества экспериментальных данных в виде топологической карты — решетки размерностью $n \times m$, состоящей из нейронов. Каждый нейрон представляет собой вектор, определяющий в d – мерном пространстве признаков некоторую характерную точку. В результате реализации алгоритма обучения вектора-нейроны распределяются в пространстве признаков, аппроксимируя статистические взаимосвязи между данными обучающего множества.

При использовании SOM для анализа (мониторинга) различных систем и процессов необходимо реализовать следующие процедуры:

- сбор данных (измеряемых входных, выходных сигналов и характеристик процесса);
- предварительная обработка данных (фильтрация);
- выделение признаков с целью наилучшего определения проблемной среды с позиций анализа;
- нормализация (взвешивание);
- обучение и “разметка” карты;
- подтверждение работоспособности (принятие решения об адекватности модели); визуализация.

Причем выполнение этапов 1,5,6,7 является обязательным, остальных – опциональным.

В данной работе рассмотрена игровая задача выяснения профпригодности на основе биометрических признаков. А именно, статистика спортивных соревнований и изучено влияние различных параметров на вероятность выигрыша. Для решения этой задачи используются Самоорганизующиеся карты Кохонена, с помощью toolbox SOM (Self-Organizing Maps).

Проведен анализ статистики профессиональных боксеров в супертяжелой весовой категории. Для обучения нейросети вводим 50 действующих боксеров, находящихся на вершине рейтингов. Каждого из них рассматриваем как отдельный вектор. В качестве координат используем возраст в годах, рост в сантиметрах, массу в килограммах, размах рук в сантиметрах, количество побед, количество поражений, результативность в пяти последних боях (количество побед в них) и, собственно, сам результат в этом бою. Он будет дискретен: выиграл – 1, проиграл – 0. Статистические данные будем брать с известного сайта, посвященного боксу, boxrec.com.

На рис.1. Представлены результаты обучения и визуализации параметров карты.

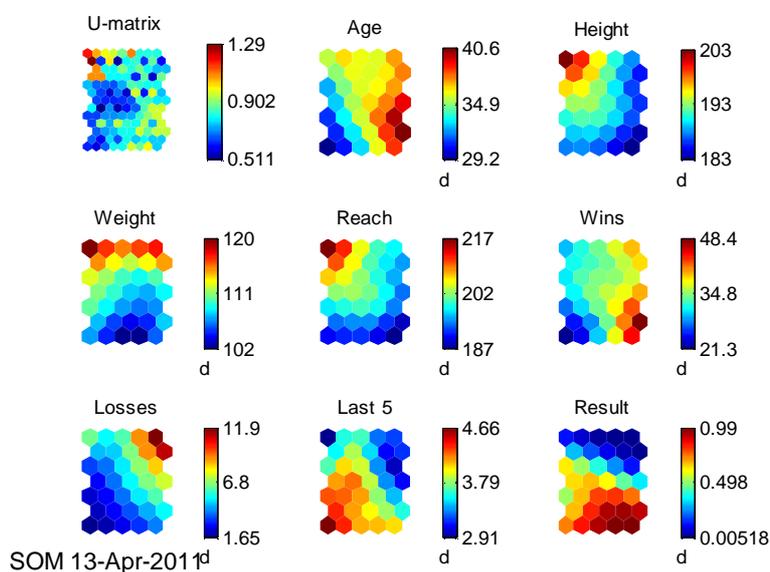


Рис. 2. Распределение нейронов на карте размером 7 на 5

Глядя на эту карту, можно выявить несколько взаимосвязей. Во первых, видна четкая взаимосвязь между параметром 2(рост) и параметром 4(размах рук). Вес с ними тоже связан, но не так сильно. В целом это достаточно очевидно: чем больше рост, тем больше размах рук. Вторая взаимосвязь менее четкая: параметр 1(возраст) и параметр 5(число побед). Сложно одержать много побед, если ты новичок. Интересно что взаимосвязи между возрастом и числом поражений не видно. И в самом деле, многие ветераны по-прежнему в обойме. Например, Виталий Кличко в настоящий момент является чемпионом, не смотря на возраст 39 лет. Еще можно выявить легкое влияние на вероятность выигрыша следующих параметров: 1) вес – видно, что излишки веса негативно влияют на результат. 2) число поражений – влияние тоже негативное. Действительно, многие боксеры «ломаются» после поражений. 3) пять последних боев – влияние позитивное. Естественно, боксеры, выигрывающие бои, находятся на подъеме.

Также посредством этой карты можно составить модель идеального боксера и, наоборот, худшего. Наилучший боксер, имеющий вероятность победы в следующем бою 0.742, имеет возраст 35 лет, рост 187 см, массу 102 килограмм, размах рук 190, 35-40 побед, 1-2 поражения и 4 победы в 5 последних боях. Если немного округлить, то получается боксер, сильно похожий на Томаша Адамека (показательно, что Томаш

выиграл бой спустя неделю после первого запуска программы). Наихудший боксер, имеющий вероятность выигрыша 0.334, имеет возраст 37 лет, рост 190 см, вес 117 кг, размах рук 192 см, 35-37 побед, 8-12 поражений и 3 победы в последних 5 боях. 100% подходящего этим параметрам боксера автор не знает.

А теперь о том, чего на этой карте не видно: на ней нет «аномалий». Например возраст размечен до 40.6, тогда как возраст Эвандера Холифилда, входящего в обучающую выборку, на момент последнего боя 47 лет (сейчас уже 48). В обучающую выборку также входил и гигант Джулиус Лонг. Его рост 216 см, масса 136 кг, размах рук 229 см. Тогда как максимумы 203, 120 и 217 соответственно. Также не присвоены отдельные нейроны боксерам, отличающимся необычными пропорциями тела. Например, Хасим Рахман имеет рост 189 см, вес 115 кг и размах рук 208 см.

Поэтому мы попробуем карты других размеров (рис.3, рис.4).

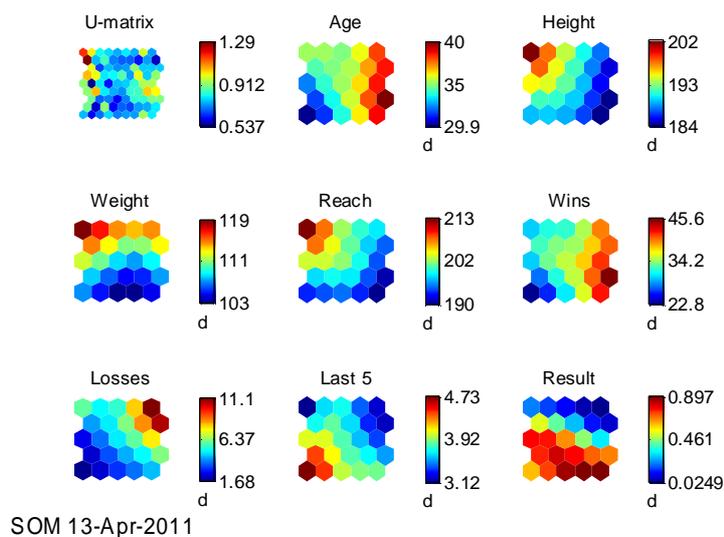


Рис. 3. Распределение нейронов на карте размером 5 на 5

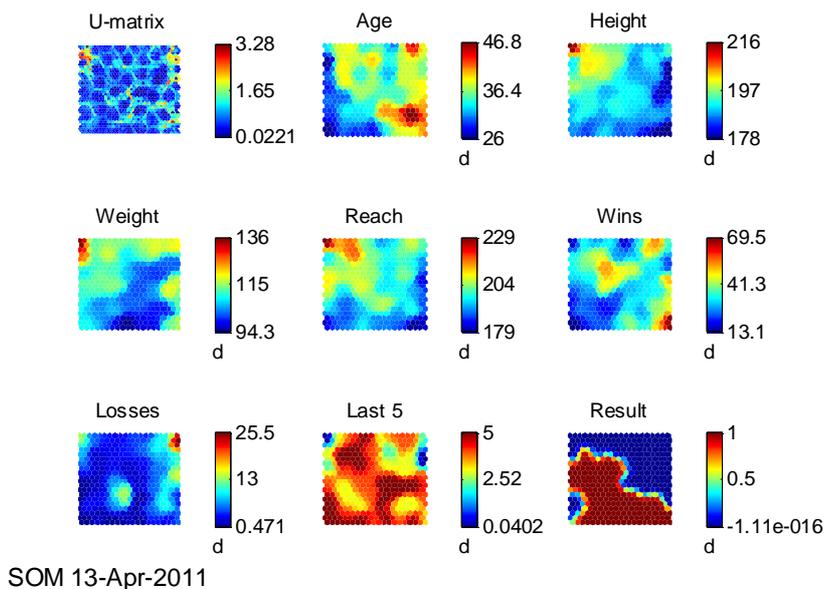


Рис. 4. Распределение нейронов на карте размером 20 на 20

При любом варианте отмеченные зависимости сохраняются, но с возросшей размерностью повышается точность результатов и пределы различных параметров.

Сделаем выводы по карте. Боксер, потерпевший за 15 поражений, выигрывать уже не будет. Также карта очень низко оценивает шансы гигантов вроде Джулиуса Лонга.

Также самоорганизующиеся карты Кохонена можно использовать для прогнозирования. В качестве примера рассмотрим шансы на удачное возвращение известного российского боксера Николая Валуева. Николая обладает поистине уникальными данными: рост 213 см, размах рук 216 см. В последнем бою он весил 143 кг, этот вес и возьмем в качестве настоящего. В настоящий момент Валуеву 37 лет, он одержал 50 побед на профессиональном ринге, потерпел 2 поражения. Из пяти последних боев Николая выиграл 4.

Поскольку Николай боксер уникальный, то точно соответствующего ему нейрона найти не удалось. Наиболее близким оказался нейрон, отмеченный на рис. 19 белым. Он имеет следующие параметры: возраст 36 лет, рост 196 см, вес 112 кг, размах рук 204 см, 45 побед, 2 поражения, 4 победы в последних 5 боях и, самое интересное, вероятность выигрыша 24%. Следовательно, Валуеву возвращаться уже не стоит. Очевидно, что основным препятствием для максимально точного прогноза стали уникальные биометрические данные российского боксера. Причина в том, что в обучающей выборке подобных людей не было. Но статистика выступлений наиболее близкого нейрона довольно хорошо отражает реальную картину.

Основные результаты работы

Проведен анализ современных технологий обработки многомерных данных.

- Исследованы вопросы применения искусственного интеллекта в задачах системного анализа.
- Исследована структура, математическое описание, алгоритмы обучения и принципы функционирования карт Кохонена.
- Разработана структура, алгоритмическое и программное обеспечение системы оценки биометрических показателей.
- Подтверждена эффективность функционирования системы прогнозирования результатов соревнований.

Литература

1. Kohonen T. Self-Organization and Associative Memory. - New-York: Springer - Verlag, 1988. – 620 p.
2. Kohonen T., Hynninen J., Kangas J., Laaksonen J. [SOM_PAK: The Self-Organizing Map Program Package](#). Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finland, 1996.
3. Nene S. A., Nayar S. K., Murase H., - Columbia Object Image Library (COIL-20), Technical Report CUCS-005-96, February 1996.
4. Alhoniemi E. et al. "Process monitoring and Modeling Using the Self-Organizing Map" Integrated Computer-Aided Engineering, IEEE. - 1999. - №6(1).