

ПРОВЕРКА ГИПОТЕЗЫ О ЕДИНОЙ ЗАКОНОМЕРНОСТИ В ДАННЫХ

Мучкина Ю. С.

Научный руководитель — профессор Миркес Е. М.

Сибирский федеральный университет

В наши дни возрастает необходимость в системах, которые способны не только выполнять однажды запрограммированную последовательность действий над заранее определенными данными, но и способны сами анализировать вновь поступающую информацию, находить в ней закономерности, производить прогнозирование и т. д. В этой области приложений самым лучшим образом зарекомендовали себя нейронные сети – математические модели, а также их программные или аппаратные реализации, построенные по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

Задачи, решаемые нейронными сетями, можно классифицировать следующим образом:

- классификация;
- кластеризация;
- аппроксимация функций;
- предсказание (прогноз);
- оптимизация;
- ассоциативная память;
- управление.

На сегодняшний день разработан ряд технологий, позволяющих решать с помощью нейронных сетей дополнительные по отношению к стандартной постановке задачи. Примером такой задачи является определение однородности обучающей выборки (построение традиций).

При решении задачи выявления закономерностей в данных предполагается, что все данные принадлежат к одной закономерности. При этом в большинстве методов аппроксимации данных единственным способом проверить эту гипотезу является применение метода. Если метод работает, то значит гипотеза верна.

Обобщенная аппроксимационная теорема Стоуна утверждает, что если в задачнике нет явных противоречий, то существует сеть, способная решить эту задачу вне зависимости от правильности гипотезы о единой закономерности. То есть нейронная сеть всегда может построить сложную закономерность, которая будет едина для всех данных задачника. Однако исследователей, как правило, интересуют простые закономерности, поскольку именно они позволяют наиболее эффективно работать с данными.

Предлагаемый метод построения традиций позволяет явным образом проверить гипотезу о единой закономерности. Суть метода состоит в следующем. Выбираем небольшое число примеров и обучим нейронную сеть решать их. Выявляем примеры, принадлежащие этой закономерности во всем задачнике. Снова обучаем нейронную сеть. Повторяем процедуру до тех пор, пока не будут выявлены все примеры, принадлежащие этой закономерности. Совокупность таких примеров называется традицией.

Ниже описан алгоритм построения одной традиции (выявления одного типа данных):

- Шаг 1: Формирование обучающего множества.
Обучающее множество состоит из двух примеров – по одному из каждого класса.
- Шаг 2: Обучение нейронной сети.
- Шаг 3: Тестирование нейронной сети
Тестирование проходят все примеры задачника, не входящие в обучающее множество.
- Шаг 4: Проверка.
Если есть правильно решенные примеры (не входящие в ОМ), то Шаг 5. Иначе – построение «Традиции» закончено.
- Шаг 5: Формирование нового обучающего множества.
Обучающее множество формируется из обучающего множества, полученного на предыдущей итерации с добавлением правильно решенных примеров.

К сожалению, содержательная интерпретация получаемых с помощью нейронных сетей типов данных может быть выполнена только специалистом в предметной области. Тем не менее, нейронные сети позволяют сформулировать вопрос интерпретации и сфокусировать внимание специалиста в предметной области на том, о чем он иначе даже не задумался бы.

Описанный алгоритм построения традиций на данный момент был опробован на двух задачах: А.Н. Горбанем на задаче о предсказаниях результатов выборов президентов США и Д.А. Россиевым на задаче о прогнозировании заболевания вторичным иммунодефицитом. В обоих случаях построенные традиции получили содержательную интерпретацию, а в случае с диагностикой вторичного иммунодефицита был выделен отдельный класс больных на завершающей стадии заболевания.

В данной работе автором разработана программная реализация алгоритма построения традиций для внедрения в учебный процесс и широкую практику.