

СБОР ДАННЫХ ДЛЯ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Севостьянюк М.А.

Белорусский государственный университет информатики и радиоэлектроники
Институт информационных технологий,
г. Минск, Республика Беларусь

Скудняков Ю.А. - к.т.н., доцент

В работе для полноценного обучения нейронной сети предложен новый подход сбора и обработки больших массивов данных с помощью разработанного программного обеспечения.

В настоящее время широкое применение решение задач с помощью нейронных сетей [1-2]. Достаточно актуальной становится проблема сбора данных для обучения.

Обучающий набор данных – это набор наблюдений, где записаны все значения как входных, так и выходных переменных. В первую очередь нужно решить, какие переменные будут использоваться и какое количество наблюдений нужно собрать.

Первоначальный выбор переменных осуществляется интуитивно, исходя из опыта разработчика в данной предметной области. Существуют специальные программные пакеты, которые облегчают выбор переменных и отмену их предыдущих выборов [3-4].

Разработчик включает все переменные, которые считает необходимыми, а программный пакет в дальнейшем помогает сократить это множество.

Работа с числовыми данными может создать проблемы, когда данные имеют нестандартный масштаб и в них имеются незаполненные значения, а также когда данные не являются числовыми.

Использование дополнительных программных пакетов помогает организовать числовые данные и масштабировать их в нужный диапазон.

Пропущенные цифровые значения можно заменить на средние.

Трудоемкой задачей является работа с текстовыми данными. Например, использование национальности (BLR, RUS, UKR, POL). Эти переменные также можно представить в виде чисел и идентификаторов с помощью дополнительного программного пакета для работы с данными.

Например, нейронная сеть должна собрать множество данных и оценить возможность риска заболеваний позвоночника в пожилом возрасте.

Это очень зависит от множества факторов, таких как: занимается ли человек спортом, имел ли травмы спины в прошлом, работает ли на работе, где нужно поднимать тяжести и другие. Травмы спины можно разделить на несколько различных групп по степени тяжести, группы по степени тяжести можно разделить на другие подгруппы.

Травмы спины в прошлом также можно отнести к различным периодам жизни человека, таких как детство (до 12 лет), подростковый возраст (12-16 лет), молодость (16-30 лет) и т.д.; Все это важно, для учета процента риска в будущем.

В данный момент известны некоторые эвристические правила, увязывающие число необходимых наблюдений с размерами сети. Согласно правилам, число наблюдений должно быть в десять раз больше числа связей в сети.

Это зависит также от сложности отображения, которое нейронная сеть стремится воспроизвести. С ростом переменных количество требуемых наблюдений растет нелинейно и при большом числе переменных может потребоваться огромное число наблюдений.

Также при работе с данными иногда появляется проблема некорректных данных. При распознавании значения некоторых переменных могут быть искажены или отсутствовать.

Программный пакет для работы с нейронными сетями должен иметь специальные средства обработки пропущенных значений.

В целом, нейронные сети способны самостоятельно обрабатывать искаженные данные, заменяя их на среднестатистические.

В целом можно сделать вывод о следующем:

- разработчику при сборе данных для нейронной сети всегда нужно выбирать такие переменные, которые точно должны влиять на конечный результат, выбирать только значащие переменные;
- переменные текстового типа по возможности преобразовать в числовые или отметить, как незначащие;
- для анализа нужно иметь от сотен или тысяч наблюдений и чем больше в задаче переменных, тем больше нужно иметь наблюдений;
- избегать наблюдений, которые содержат пропущенные значения.

Данные рекомендации помогут точно и правильно произвести анализ данных и обучение нейронной сети.

Список использованных источников:

1. Хайкин, С. Нейронные сети: полный курс *Neural Networks: A Comprehensive Foundation* / С. Хайкин. – М.: Вильямс, 2006. – 2-е изд. – 1104 с.
2. Каллан, Р. Основные концепции нейронных сетей / Р. Каллан - М.: Издательский дом «Вильямс», 2001. - 928 с.
3. Портал о машинном обучении [Электронный ресурс] – Режим доступа: [http:// machine learning.ru/](http://machinelearning.ru/) – Дата доступа: 20.03.2021.
4. Statsoft [Электронный ресурс] – Режим доступа: <http://statsoft.ru/> – Дата доступа: 20.03.2021.