

ОПТИМИЗАЦИЯ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ С ПОМОЩЬЮ БАЗЫ ДАННЫХ В ОПЕРАТИВНОЙ ПАМЯТИ

*Карнилов М.С., студ., Тихон В.Н., студ., Соколова А.С., асс.,
Черненко Д.В., ст. преп.*

*Витебский государственный технологический университет,
г. Витебск, Республика Беларусь*

Реферат. В статье рассмотрены возможности и новые способы оптимизации обучения нейронных сетей, работа в H2 mem DataBase на Java, принципы работы нейронных сетей, а также их обучение.

Ключевые слова: нейронные сети, базы данных, H2 mem DB, активационная функция, искусственный нейрон, искусственный интеллект, оптимизация.

База данных в памяти – это база данных, которая хранит весь набор данных в оперативной памяти. Это означает, что каждый раз, когда запрашивается база данных или обновляются данные в ней, предоставляется доступ только к основной памяти. Таким образом, в этих операциях отсутствует взаимодействие с диском, что ускоряет работу с памятью.

У таких баз данных самая высокая скорость обработки информации. Поэтому острая необходимость в таком хранилище возникает в двух случаях: когда необходимо работать с очень большим объемом информации, и в случае работы с большим количеством запросов к ней.

Рассмотрим возможность применения баз данных в памяти, на примере H2 memory DataBase и одним из модулей искусственного интеллекта – тренером нейронной сети.

При разработке искусственного интеллекта, в частности нейронных сетей, самой трудной и затратной задачей является обучение нейронной сети.

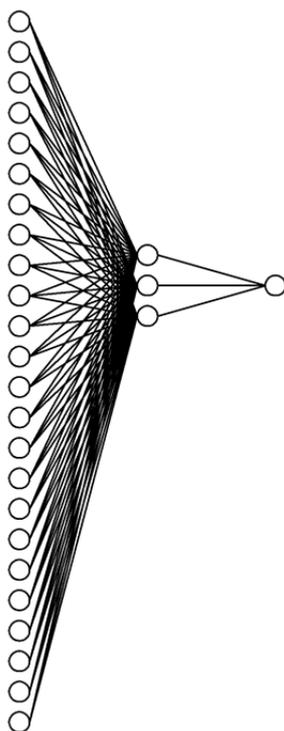


Рисунок 1 – Пример искусственной нейронной сети

На рисунке 1 приведен пример нейронной сети на 28 нейронов. Данная нейронная сеть имеет 3 слоя: input layer – входной слой, hidden layer – скрытый слой, output layer – выходной слой. Чтобы рассмотреть, как работает нейронная сеть необходимо разобраться, как устроен и работает искусственный нейрон.

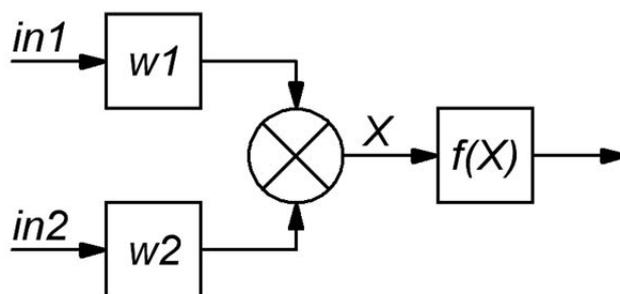


Рисунок 2 – Искусственный нейрон

Рассмотрим нейрон, имеющий 2 входа и один выход. Входные сигналы складываются с учетом их веса (важности)

$$X = in_1 w_1 + in_2 w_2$$

и подаются на блок, реализующий активационную функцию $f(X)$, которая определяет, срабатывает нейрон или нет. Входных сигналов может быть и больше.

Можно задавать весовые коэффициенты вручную. Однако это займет длительное время для одного нейрона, не говоря уже о нейронной сети в целом. Поэтому еще в прошлом веке придумали использовать тренера искусственной нейронной сети, который сам настраивает каждый нейрон.

Рассмотрим лишь метод обучения с учителем, т. е. когда имеются данные, которые необходимо обработать, и ответы. В самом начале при создании нейронной сети весовые коэффициенты нейронам назначаются случайно. Вследствие чего после первой эпохи (цикла обучения нейронной сети) результат работы будет неправильный. При каждом неправильном результате, вычисляется ошибка сети, используемая для определения в какую сторону необходимо подстраивать весовые коэффициенты нейронов.

На рисунке 1 представлена довольно простая нейронная сеть, и процесс ее обучения занимает минуты. Однако если имеется нейронная сеть на 1011 нейронов (столько нейронов в среднем содержит человеческий мозг), то ее обучение займет годы.

Поэтому для достижения максимально эффективного и быстрого обучения необходима оптимизация вычислительных процессов и процессов, связанных с сохранением и доступу к памяти.

Оптимизация вычисления достигается за счет рационального использования вычислительных мощностей. Такие гиганты как Intel и NVidia разработали специальные библиотеки, позволяющие эффективно использовать CPU и GPU.

Что касается работы с памятью, то самая низкая скорость работы с данными с использованием классического хранилища – HDD, так как она происходит с участием механических частей. Более скоростной является работа с SSD (твердотельным накопителем). Наивысший уровень данного направления оптимизации – работа с DRAM (оперативной памятью).

Несомненно, скорость будет теряться на программном уровне. Чтобы уменьшить потери, можно в оперативной памяти развернуть базу данных. Высокопроизводительной базой данных, хранящейся в памяти DRAM является H2.

Для оптимизации обучения искусственной нейронной сети в базе данных хранят весовые коэффициенты за предыдущие эпохи обучения. Это позволяет быстро откатывать нейронную сеть к состоянию, когда процент успешных решений был выше.

Еще одним достоинством базы данных в оперативной памяти является то, что она существует только тогда, когда она активна. Это способствует удалению ненужных данных после обучения. После обучения искусственная нейронная сеть сохраняет в себе только последние весовые коэффициенты и отключает базу данных, тем самым предоставляется возможность использования того адресного пространства, которое она занимала. А данных может быть большое количество.

Таким образом, использование базы данных в оперативной памяти способствует ускорению работы с данными и при этом позволяет эффективно использовать адресное пространство ЭВМ, тем самым оптимизируя процесс обучения нейронной сети.