Конспект лекции

А.Е. Сулавко

Защищенное исполнение искусственного интеллекта

**Защищенное исполнение нейросетевых алгоритмов на базе квадратичных нейронов**

**ЗАЩИЩЕННОЕ ИСПОЛНЕНИЕ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ НА БАЗЕ КВАДРАТИЧНЫХ НЕЙРОНОВ**

Помимо сетей персептронов (сетей, выстроенных на классических нейронах), для генерации кода можно использовать сеть иных функционалов, в частности, квадратичных форм. В таком случае *квадратичный нейрон* может быть определен как сумматор входных биометрических данных, обученный обогащать эти данные в квадратичном пространстве с последующим применением к результату обогащения функции активации.

*Обучение квадратичного нейрона* представляет собой подбор параметров сдвига математических ожиданий биометрических данных, положения порогов его квантователя, а также весовых коэффициентов, обеспечивающих практически однозначное преобразование примеров образа «Свой» в выходной код доступа и хеширование (перемешивание) данных образов «Чужой». При этом первым шагом такого обучения является вычисление положения откликов нейрона при замене математического ожидания биометрических параметров нулевыми значениями.

*Оценка вероятности появления ошибок первого рода* (отказа в доступе при предъявлении примеров образа «Свой») осуществляется предъявлением примеров, не участвовавших в обучении. При использовании примеров – *m* и выявлении ошибочного отказа в доступе – *e* вероятность оценивают через следующее соотношение:

(1)

В случае отсутствия обнаруженных ошибок в тестовой выборке е = 0 вероятность появления ошибок первого рода оценивают как

(2)

Если необходима более точная оценка, то следует добавить шум к примерам образа «Свой» для просмотра окрестностей проверяемых примеров.

*Оценка вероятности ошибок второго рода* выполняется путем использования тестовой базы образов «Чужой» с вычислениями по ГОСТ Р 52633.3–2011 в пространстве расстояний Хэмминга. Отличие состоит только в том, что расстояния Хэмминга вычисляются по модулю K. Переход от вычисления расстояний Хэмминга по модулю два к вычислению расстояний Хэмминга по модулю К обусловлен заменой бинарных квантователей линейных нейронов на К уровневые квантователи квадратичных нейронов.

Классическая квадратичная форма предполагает обращение корреляционных (ковариационных) матриц (размерность матрицы равна количеству признаков). В биометрических приложениях обращение этих матриц затруднительно, если их размерность оказывается высокой. В этом случае возникает эффект, называемый *«проклятием размерности»*. В общем случае, в машинном обучении, под этим термином понимают экспоненциальное возрастание обучающих данных и/или связанных с их обработкой вычислений.

Сети квадратичных нейронов обучаются в полностью *автоматическом режиме*. Они имеют общие черты с сетями радиально-базисных функций[6], однако в отличие от последних не компрометируют биометрический эталон. Квадратичные нейроны имеют иную архитектуру[6], что позволяет *скрывать статистические моменты признаков*, вычисляемых при обучении, то есть осуществлять обогащение биометрических данных с сокрытием математических ожиданий биометрических параметров случайными числами, а также скрывающими действительное значение стандартных отклонений обрабатываемых им биометрических параметров. Данная особенность квадратичным нейронов позволяет говорить о *защищенном исполнении сетей квадратичных нейронов*.

Квадратичный нейрон с четырьмя входами дает эффект, сравнимый с линейным нейроном, имеющим 16 входов. Таким образом, при обработке 512 признаков сеть квадратичных нейронов будет давать на выходе 512/4=128 бит ключа (каждый нейрон по одному биту). Это существенный рост длины криптографического ключа (но этого во многих случаях недостаточно). В реальных задачах количество входов нейрона зависит от информативности признаков и их коррелированности. Для построения преобразователя биометрического образа в 256 битный ключ требуется не менее 1024 достаточно информативных признаков. Но во многих типах биометрических образов такого количества признаков просто нет.

Основным недостатком квадратичных нейросетевых функционалов является то, что выходной код для образов «Чужой» обладает низкой энтропией, связанной с отсутствием баланса состояний «0» и «1» в разрядах выходного кода. Данный недостаток можно устранить за счет применения выходного квантователя на выходе сумматоров нейронов с тремя и более выходными состояниями [7].

Меры близости, которые могут лежать в основе квадратичных нейронов и сетей радиально-базисных функций обобщаются с помощью меры Минковского:

(3)

где *aj* – значение j-го признака из вектора ā, представляющего собой распознаваемый образ (биометрический или иной); n – количество признаков; *mj* и *σj* – математическое ожидание и среднеквадратичное отклонение значений j-го признака для того класса образов, с которым сравнивается образ ā; *p* – степенной неотрицательный коэффициент, определяющий уровень *«искривления» пространства признаков*.

Важным понятием для работы с пространством признаков, является понятие ***спрямляющего пространства.*** Для образования нового понятия надо построить соответствующую разделяющую гиперповерхность. Каждой гиперповерхности пространства X в пространстве Y с координатами y1 = φ1(x), …, ym = φm(x) соответствует гиперплоскость . Введение пространства Y позволяет заменять рассмотрение разделяющих гиперповерхностей разделяющими гиперплоскостями. Поэтому пространство векторов Y получило название спрямляющего. В спрямляющем пространстве каждому объекту ставится в соответствие вектор y = (y1, …, ym). Этот вектор относится к первому классу, если он лежит по одну сторону от разделяющей гиперплоскости, и ко второму, если по другую.

*Искривление пространства признаков возникает из-за наличия корреляционных связей между ними*. Как правило, пространство признаков не является ни плоским, ни в равной степени искривленным. Утверждение о том, что это пространство имеет локальные неоднородности, также не вполне корректно. Скорее уровень искривления пространства признаков меняется относительно наблюдателя. Различные классы образов, как правило, имеют отличающиеся матрицы коэффициентов корреляции между признаками, поэтому относительно образов, относящихся к разным классам, пространство признаков искривлено по-разному.

Помимо коэффициентов корреляции с другими признаками важным показателем для признака является уровень его информативности, который можно оценить по данным обучающей или тестовой выборки. Для каждого класса образов информативность любого признака может существенно отличаться. Под *информативностью признака* для определенного класса образов понимается количество собственной информации от площади пересечения функций плотности вероятности (ФПВ) «Свой» и «Чужие». ФПВ «Свой» характеризует значения признака строго для определенного класса образов, ФПВ «Чужие» характеризует значения этого же признака для всех классов образов в целом. Хотя термины «Свой» и «Чужие» (классы, образы) относятся к области биометрии, их можно распространить на любую задачу классификации, при этом из смысл останется тем же.

Изменяя параметр *p* в (3) можно добиться *снижения количества ошибок классификации*, если признаки имеют невысокий и примерно равный уровень коррелированности. В реальных задачах корреляция между признаками различна, поэтому следует создавать *сеть или комитет нейронов (в данном случае квадратичных)*, каждый из которых обрабатывает отдельный набор признаков, имеющих определенный в некотором интервале уровень корреляционной зависимости.

Всегда существует оптимум параметра *p,* при котором достигается наименьший показатель EER. Если признаки независимы, то пространство признаков можно считать *«плоским»* (не искривленным) и мера близости Пирсона/Евклида дает хороший результат. Чем выше корреляция между признаками, тем больше возрастает оптимальное значение *p* и шире интервал, на котором лежит оптимум.

**Список источников**