Конспект лекции

А.Е. Сулавко

Защищенное исполнение искусственного интеллекта

**Защищенное исполнение гибридных нейросетевых алгоритмов**

**ЗАЩИЩЕННОЕ ИСПОЛНЕНИЕ ГИБРИДНЫХ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ**

Активное развитие «широких» нейронных сетей в настоящее время идет по направлению комплексирования нейронов, имеющих принципиально разную архитектуру, в рамках одной гибридной сети.

Объединение различных нейронов (классических, квадратичных, гармонических и др.) в гибридную сеть потенциально позволяет значительно повысить энтропию выходов гибридного нейросетевого ПБК и снизить вероятность ошибочных решений. Гибридную сеть можно рассматривать как ансамбль слабых классификаторов. Снижение вероятности ошибок гибридной чети объясняется теоремой Кондорсе, которая утверждает: если мнения экспертов независимы, и вероятность правильного решения каждого из них больше 0,5, то с увеличением количества экспертов вероятность правильного решения комитета экспертов возрастает и стремится к единице. Причем, чем выше вероятность верного решения для каждого эксперта в отдельности, тем выше вероятность верного решения комитета. Отметим, что решение любого нейрона с бинарным выходом можно инвертировать, чтобы преодолеть барьер Кондорсе в 0,5 (известны также доказательства других теорем, позволяющих обойти барьер Кондорсе).

Однако на практике решения нейронов, играющих роль экспертов, в той или иной мере коррелированы, чем ниже коррелированность решающих правил, тем более ощутим положительный эффект при их комплексировании. Таким образом, имеются следующие параметры, которые влияют на эффективность гибридной нейронной сети:

* количество нейронов
* матрица коэффициентов корреляции Пирсона между решениями всех возможных пар нейронов
* мощность нейронов (их способность давать верные решения).

**Вариант модели гибридной ПБК.** Гибридная сеть из разнотипных нейронов, каждый из которых обрабатывает сочетания признаков с определённым уровнем взаимной зависимости или суммарной информативности, названа *гибкой сетью*. Такая сеть подстраивается под пространство признаков субъекта, определяя свою конфигурацию, исходя из обучающей выборки. Эти сети могут быть многослойными. Положительный эффект достигается, если на разных слоях используются различные типы нейронов. Однако аналогичного эффекта можно добиться, объединяя разнородные нейроны в один большой слой. В этом случае можно сохранить структурную пластичность сети и менять конфигурацию в процессе функционирования (добавляя, убирая нейроны), что затруднительно, если слоёв несколько (мутации гибких сетей в настоящей работе не рассматриваются).[4]

Если гибридный ПБК (ГПБК) стоится на базе однослойной гибкой сети, то классические нейроны, а также разностные и гиперболические нейроны Байеса принимают решение, исходя из порогового значения μ0 (1), для остальных нейронов предусмотрена иная функция активации (2).

(1)

(2)

где b – бит ключа, на который «настраивается» нейрон, *wz* – весовой коэффициент, влияющий на балансировку порогов всех нейронов z-типа.

Рассмотрим процесс создания и обучения гибридного ПБК (ГПБК) на базе однослойной гибкой сети:

1. Вычисляются параметры *mh,j, σh,j* (математическое ожидание и средне-квадратичное отклонение значений признака, соответствующего j-му входу, характерные для гипотезы h)*, Ij* (информативность) для каждого *j*-го признака
2. Определяется 10 групп информативности признаков G1–G10, начиная с наиболее информативных 0 < I*j* < 0,1 (G1), заканчивая наименее 0,9 < I*j* < 1 (G10), ΔI = 0,1. Признаки из G10 не учитываются, они почти не влияют на результат распознавания.
3. По обучающим данным «Свой» вычисляется матрица *RС* коэффициентов корреляции *rj,t* признаков.
4. Создаются нейроны для обработки независимых признаков. Для обработки каждой группы Gi создаются отдельные нейроны, причём при переходе от *Gi* к *Gi+1* количество входов каждого нейрона возрастает в 1,5 раза (*ni+1*= *Δn ꞏ ni, Δn = 1,5*): для квадратичных форм *n1*= 2, для «критериев» *n1*= 4.
5. Формируются нейроны, обрабатывающие зависимые признаки. Для «критериев» и «гравитационных» метрик устанавливается фиксированная размерность *nc* = 20 и *ng*= 2 соответственно. Зависимые признаки не нужно группировать по информативности, они имеют схожие формы функций плотности вероятности и близки по информативности.

Указанные значения гиперпараметров (*ΔI, ni, Δn, nc, ng*) были получены эмпирически и близки к оптимальным. Каждый признак может входить в несколько нейронов, но в один нейрон не более одного раза.

Нейроны, относящиеся к одной категории (например, квадратичные), настраиваются идентично. Поэтому допустимо формировать единую конфигурацию для всех типов нейронов, относящихся к категории (например, создав «подсеть» разностных нейронов Байеса, она полностью копируется и для гиперболических). Это ускоряет настройку. Желательно, чтобы количество совпадающих синапсов для нейронов, базирующихся на одной мере близости, не превышало 25 %. Иначе ИНС будет избыточной.

В состав ГПБК может входить НПБК (ГПБК+).

Настройка многослойной гибкой сети отличается тем, что нейроны последующих слоёв воспринимают выходы предыдущего как признаки, поэтому обучающая выборка «пропускается» через очередной слой для настройки следующего. Процесс повторяется циклически, пока не будут настроены все слои.

Описанный процесс обучения абсолютно устойчив. Время обучения *гибкой сети* сопоставимо со временем обучения широкой сети (при равном числе нейронов и объёме выборки), однако при этом количество ошибочных решений у гибкой сети гораздо меньше. Сравнимый с НПБК уровень ошибок для ГПБК получается на почти двукратно меньшем числе обучающих примеров «Свой» (8 – 9), что эквивалентно увеличению скорости обучения.

Важным аспектом обучения любого вида сети, в том числе гибридной, является факт изменения со временем биометрических образов пользователей, что продиктовано множеством как внешними по отношению к человеку факторами, так и его внутренним состоянием. В таком случае можно говорить о так называемом *концептуальном дрейфе модели* (*дрейфе данных*)[7]. Под данным термином, в общем случае, понимается любое изменение отношений между входными и выходными данными модели машинного обучения с течением времени.

*Создать ПБК на основе многослойной нейронной сети с использованием итерационных алгоритмов обучения на сегодняшний день затруднительно*. Дел в том, что формирование и обучение ПБК, как правило, должно быть автоматическим (по крайней мере, в биометрии и многих других приложениях информационной безопасности), но практически все итерационные алгоритмы имеют существенную склонность в переобучению. Как следствие, приходится постоянно следить за процессом обучения и периодически проводить проверку качества решений на валидационной выборке. Чем ниже информативность биометрического образа (т.е. уникальность и стабильность признаков), тем больший объем выборки нужен и тем выше склонность в переобучению. Например, для обучения *сверточных нейронных сетей* в задаче биометрической идентификации личности по лицу достаточно пяти примеров изображения лица от каждого идентифицируемого субъекта (в различных ракурсах) для того, чтобы процесс настройки был относительно робастным. Образ лица весьма информативен, гораздо информативней, чем, например, рукописный или голосовой образ.

Отдельным подходом можно рассматривать обучение глубоких нейронных сетей с множеством выходов для извлечения признаков из биометрических образов. Недостаток такого подхода заключается в том, что генерируемый из биометрических данных код на выходе нейронной сети коррелирован с входными биометрическими данными, что небезопасно.

Извлечение признаков с помощью многослойных нейронных сетей нельзя приравнивать к генерации ключа на основе биометрических данных, так как к криптографическим ключам и паролям предъявляются определенные требования, связанные с их длиной и энтропией. Признаки, непосредственно извлеченные из биометрического образа (или любого другого распознаваемого образа), не обладают соответствующими свойствами. Извлечение признаков – это лишь этап обработки образа в системе связывания предварительно сгенерированного ключа и биометрических параметров субъекта.

Помимосверточных многослойных сетей, в качестве экстрактора признаков можно также рассматривать ***вариационные автокодировщики***[11, глава 5]. В общем случае вариационный автокодировщик представляет собой автокодировщик (генеративная модель, которая учится отображать объекты в заданное скрытое пространство (и обратно)) основанный на вариационном выводе. При попытке использования обыкновенного автокодировщика для генерации новых образов (желательно из того же априорного распределения, что и датасет) возникает следующая проблема: случайной величиной с каким распределением проинициализировать скрытые векторы, для того, чтобы образ, после применения декодера, стал похож на образы из датасета, но при этом не совпадал ни с одной из них? Ответ на этот вопрос не ясен, в связи с тем, что обыкновенный автокодировщик не может ничего утверждать про распределение скрытого вектора и даже про его область определения. В частности, область определения может быть даже дискретной. Вариационный автокодировщик в свою очередь *предлагает пользователю самому определить распределение скрытого вектора*[10].

**Список источников**