

Лекция 12: Невронни мрежи

СЪЩНОСТ НА НЕВРОННИТЕ МРЕЖИ

Невронните мрежи (НМ) и по-точно, изкуствените НМ са най-типичният и най-развитият представител на т. нар. *конекционистки подход* в Изкуствения интелект. Основните градивни елементи на (изкуствените) НМ са силно опростени модели на биологичните неврони, от които е съставен човешкият мозък.

Основни компоненти на биологичните неврони:

- *дендрити* – разклонени структури, които осигуряват сензорен вход към тялото на неврона (входни устройства на неврона);
- *тяло на неврона* – сумира мембранните потенциали между синапсите и дендритите и изпраща по аксона активиращо напрежение с определен размер;
- *аксон* – провежда електрическия сигнал от тялото на неврона до неговите синапси;
- *синапси* – трансформират активиращото напрежение, получено по аксона, в електрически сигнали (импулси), които възбуждат или подтискат активността на съседните неврони (изходни устройства на неврона).

Отделните неврони са свързани помежду си (дендритите на всеки неврон получават “входна информация” от синапсите на други неврони и т.н.).

Всяка изкуствена НМ е система от обработващи елементи (processing units), които са силно опростени модели на биологичните неврони и затова често условно се наричат (изкуствени) неврони. Обработващите елементи са свързани помежду си с връзки с определени тегла. Всеки обработващ елемент преобразува входните стойности (входните сигнали), които получава, и формира съответна изходна стойност (изходен сигнал), която разпространява към елементите, с които е свързан. Това преобразуване се извършва на две стъпки. Най-напред се формира сумарният входен сигнал за дадения елемент, като за целта отделните входни стойности (сигнали) се умножават по теглата на връзките, по които се получават, и резултатите се събират. След това обработващият елемент използва специфичната за мрежата *активационна функция* (функция на изхода), която трансформира получения сумарен вход в изхода (изходния сигнал) на този елемент.

ОСНОВНИ ТИПОВЕ МОДЕЛИ НА НЕВРОННИ МРЕЖИ

Специалистите определят пет основни характеристики на НМ, които обикновено се използват за класификационни признаци при определянето на типовете модели на НМ: *топологията на мрежата, параметрите на обработващите елементи, типът на входните стойности, методът за обучение на мрежата и използваното обучаващо правило*.

Според топологията на мрежата

Най-често срещани са НМ, съставени от няколко обособени (последователни) слоя от елементи, при които елементите от най-ниския слой играят ролята на входни устройства на мрежата (те възприемат сигнали от външната среда), а елементите от

най-горния слой играят ролята на изходни устройства на мрежата (те извеждат резултата от работата на мрежата, който се получава на базата на входните сигнали и теглата на връзките между елементите). Обикновено при тези НМ връзките са еднопосочни и свързват елементите от един слой с елементи на слоя, разположен непосредствено над него. В зависимост от броя на слоевете в мрежата се говори за *двуслойни НМ* (при тях има само един входен и един изходен слой и липсват т. нар. *вътрешни* или *скрити слоеве*) и *многослойни НМ* (при тях има поне един скрит слой).

Според параметрите на елементите

Тук от значение са посоката и теглата на връзките, видът на активационната функция и др.

Както беше посочено, при повечето НМ връзките са еднопосочни и са ориентирани от елементите от даден слой към елементите от слоя, разположен непосредствено над него (това са т. нар. *прави мрежи*). Съществуват обаче и т. нар. *рекурентни мрежи* (такива са например мрежите на Хопфийлд), при които всеки елемент е свързан с двупосочни връзки с всички свои съседи. При тези мрежи понятието слой до голяма степен губи своя смисъл.

По отношение на теглата на връзките се говори за т. нар. *възбуждащи връзки* (връзки с положителни тегла) и *подтискащи връзки* (връзки с отрицателни тегла).

Най-често срещаните типове активационни функции са *стъпаловидна функция* (при персептрона) и *сигмоид* (при НМ с обратно разпространение на грешката).

Според типа на входните и изходните стойности

Входните стойности на мрежата (т.е. сигналите, които получават елементите от входния слой) могат да бъдат *двоични* (0 или 1) или *аналогови* (реални числа). Същото се отнася и за изходните стойности на мрежата (стойностите/сигналите, които елементите от изходния слой извеждат към “околната среда”). В случай, че изходните стойности на мрежата трябва да бъдат двоични, се налагат допълнителни изисквания към вида на активационната функция.

Според обучаващото правило

Най-често НМ се използват за решаване на задачи, свързани с разпознаване (класификация). При това създаването на НМ, предназначена за решаване на дадена конкретна задача, обикновено се извършва по следната обща схема. Най-напред се определя топологията на мрежата, т.е. броят на слоевете и броят на елементите във всеки слой. Броят на елементите от входния слой се определя от размерността на входните данни, а броят на елементите от изходния слой – от броя на разпознаваните класове. Броят и размерите на скритите слоеве се определят итеративно в зависимост от предметната област и конкретната задача. След определянето на топологията на мрежата се преминава към нейното обучение, т.е. към определянето на подходящи стойности на теглата на връзките между елементите. За целта най-често отначало се задават случайни стойности на търсените тегла, след което те итеративно се променят (уточняват) по такъв начин, че на следващата стъпка новият изход на мрежата да бъде по-добър от стария. Правилото, по което се променят теглата на връзките, се нарича *обучаващо правило* (обучаващ алгоритъм) на мрежата.

Примери за по-известни обучаващи правила: правило за обучение на персептрона (алгоритъм за обучение с фиксирано нарастване), алгоритъм за обучение с обратно разпространение на грешката, правило на състезателното обучение.

Според метода на обучение

Съществуват два основни типа обучение на НМ – *обучение с учител* и *обучение без учител*. При обучението с учител се следи разликата между получения и очаквания изход на мрежата (учителят задава очаквания изход на мрежата) и итеративно се извършват корекции на теглата на връзките съгласно избраното обучаващо правило. При обучението без учител липсват данни за правилния изход на мрежата. Теглата на връзките се настройват така, че представянето на данните в мрежата да е най-добро съгласно използвания (зададения) критерий за качеството на представянето.

ПОПУЛЯРНИ МОДЕЛИ НА НЕВРОННИ МРЕЖИ И АЛГОРИТМИ ЗА ОБУЧЕНИЕ

Персептрон

Персептронът (по-точно, двуслойният персептрон) е двуслойна НМ с един елемент в изходния слой, който често се нарича сумиращ процесор на персептрона.

Резултатът от работата на персептрона е решаването на задача за разпознаване при дадени два класа, разделими с хиперравнина (в двумерния случай - разделими с права линия, т.е. линейно разделими).

Алгоритъм за обучение на персептрона – т. нар. *алгоритъм за обучение с фиксирано нарастване*. Доказана е теорема за сходимост, според която ако съществува множество от тегла, определящи хиперравнина, която разделя двата класа, то с помощта на този алгоритъм могат да бъдат намерени подходящи стойности на търсените тегла на връзките между елементите от входния и изходния слой.

Забележка. Нека е дадено множество от обекти, които подлежат на разпознаване/класификация. Предполага се, че всеки обект се характеризира с набор от съществени признаци. На базата на използваните признаци се определя *признаковото пространство (пространството на признаците)*. Всеки обект е точка в това пространство. Тогава задачата за дефиниране на класовете се свежда до това за всеки клас C_i да се намери функция $g_i(\vec{x})$, която е такава, че $\vec{x} \in C_i$ тогава и само тогава, когато $g_i(\vec{x}) > g_j(\vec{x})$ за всяко $j \neq i$. При известни функции $\{g_i\}$ задачата за класификацията на даден обект \vec{x} се свежда до намиране на това k , за което $g_k(\vec{x}) > g_j(\vec{x})$ за всяко $j \neq k$ (тогава $\vec{x} \in C_k$). При наличие само на два класа горната задача се свежда до намиране или използване на т. нар. *разделящо правило (разделяща функция) g* със свойството: $\vec{x} \in C_1$, когато $g(\vec{x}) > 0$, и $\vec{x} \in C_2$, когато $g(\vec{x}) < 0$. В този случай точките, за които $g(\vec{x}) = 0$, образуват границата между класовете.

Невронни мрежи с обратно разпространение на грешката

Това са НМ с поне един скрит слой, при които съществуват връзки от всеки елемент от даден слой към всеки елемент от непосредствено следващия слой.

Активационната функция на този тип мрежи има вида

$$\text{output} = \frac{1}{1 + e^{-\text{sum}}}, \text{ където } \text{sum} \text{ е сумата от всички входни стойности, умножени с}$$

теглата на съответните връзки.

Алгоритъм за обучение с обратно разпространение на грешката: включва множество *епохи*, в рамките на всяка от които последователно се разглеждат всички примери от обучаващото множество. За всеки обучаващ пример се извършват т. нар. *прав пас* (разпространение на активационните стойности) и *обратен пас* (анализ на

грешките на изхода и разпространение на корекциите на теглата на връзките между елементите от различните слоеве). Има ниска скорост на сходимост, която води до ниска скорост на обучение на невронната мрежа и необходимост от голям брой обучаващи примери.

НМ с обратно разпространение на грешката се използват широко поради голямата им мощност.