VOICE RECOGNITION - HISTORICAL DEVELOPMENT AND MAIN TECHNIQUES

Hasanov Hasan, Burgas Free University, hasan.mehmedov@gmail.com

Georgieva Penka V., Burgas Free University, pgeorg@bfu.bg

Abstract: The natural language processing is one of the main areas of modern artificial intelligence. Voice recognition is an element of natural language processing and aims at transforming spoken words into written text by various techniques. Researchers in area of voice recognition face many challenges that have different sources.
In this study the historical development of voice recognition is presented, the voice recognition types are and the basic techniques used in the field are outlined.

Keywords: voice recognition, speech recognition, natural language processing, artificial neural networks, artificial intelligence

ГЛАСОВО РАЗПОЗНАВАНЕ - ИСТОРИЧЕСКО РАЗВИТИЕ И ОСНОВНИ ТЕХНИКИ

Хасан Хасанов, Бургаски свободен университет, hasan.mehmedov@gmail.com

Пенка В. Георгиева, Бургаски свободен университет, pgeorg@bfu.bg

Абстракт: Обработването на естествени езици е една от основните области на съвременния изкуствен интелект. Гласовото разпознаване е елемент на обработката на естествени езици, при който изречените думи се преобразуват в писмен текст с помощта на различни техники. Разпознаването на глас е област, в която изследователите се изправят пред множество предизвикателства от разнообразен характер.
В тази студия е направен обзор на историческото развитие на гласовото разпознаване, посочени са видовете гласово разпознаване и са представени основните техники, използвани в тази област.

Ключови думи: гласово разпознаване, обработка на естествени езици, изкуствен интелект

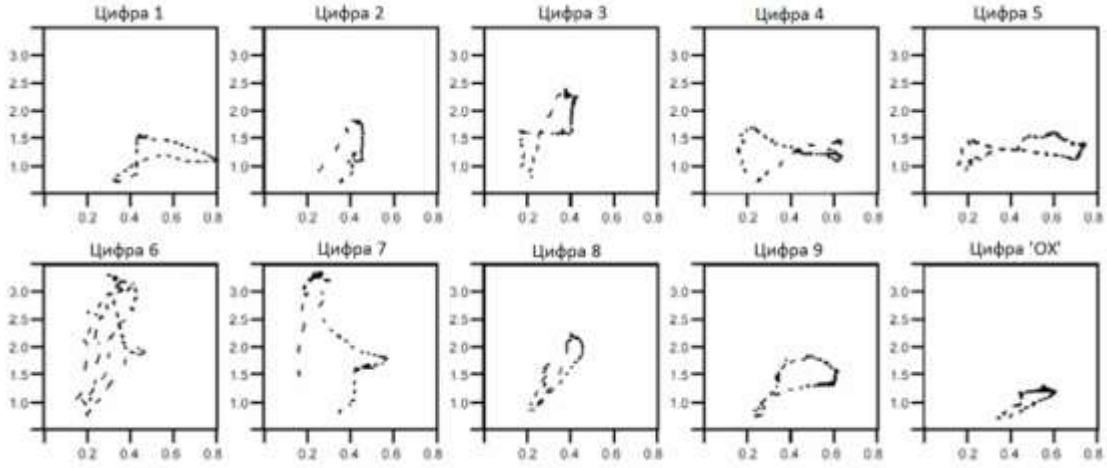
I. Въведение

Понятието *гласово разпознаване* (още познато като *разпознаване на глас*, *автоматично гласово разпознаване* и *компютърно разпознаване на глас*) се използва основно в два аспекта: 1) превръщането на изречените думи в писмен текст и 2) обучаването на дадена софтуерна система да разпознава конкретен глас.
Приложенията за разпознаване на глас включват гласови потребителски интерфейси, като гласово набиране (пр. *Обади се вкъщи!*), обработка на обажданията (пр. *Бих искал да проведа разговор.*), управление на домашните електроуреди, търсене (пр. намиране на подкаст, когато определени думи са изговорени), просто въвеждане на данни (пр. въвеждане на номер на кредитна карта), подготовка на структурирани документи (пр. доклади), гласова реч при обработка на текст (пр. текстообработка или електронни писма) и други.

II. ИСТОРИЧЕСКО РАЗВИТИЕ

Първи автоматични системи за речево разпознаване

Ранните опити за създаване на системи с автоматично речево разпознаване са в периода на 60-те и 70-те години на XX век. Те са силно повлияни от теорията за акустичната фонетика, която описва фонетичните елементи на речта (основните звуци на езика) и обяснява как тези елементи са акустично реализирани при говоримото изказване. Основните звуци са фонемите, но съществена роля имат и съответното им място и начинът на артикулация, използван за възпроизвеждане на звуците в различни фонетични контексти. Например, за пресъздаването на стабилен звук на гласна, гласните струни трябва да вибрират, а въздухът, който се разпространява през вокалния тракт се възпроизвежда като звук с естествени резонансни режими, аналогични на тези, които се появяват при акустичната туба. Тези естествени резонансни състояния се наричат форманти (формантни честоти). На фигура 1 са изобразени формантните траектории на първа и втора формантна честота при произнасяне на всяко от цифрите Тези траектории служат за референтен модел при определянето на идентичността на изговорена непозната цифра. [1].



Фиг. 1. Сравнение на възпроизвеждането на формант 1 и формант 2 на цифрите1, 2, …, 9 и непозната цифра OX

При други ранни системи за разпознаване, Олсън и Белар от лабораториите RCA създават система за разпознаване на 10 срички от един конкретен говорещ [2], а в Линкълн Лаб на Масачузетския технологичен институт, Дж. Форги и К. Форги проектират независима от говорителя система за 10-гласно разпознаване [3]. По същото време, независимо от тях японски изследователи създават специализиран хардуер, който да изпълнява задача по разпознаване на реч. Най-важни са разработките на Сузуки и Наката в Изследователската радио лаборатория в Токио [4], на Сакаи и Дошита от университета в Киото [5] и цифровия разпознавател на лабораториите NEC [6]. Работата на Сакаи и Дошита включва първата употреба на речеви сегментатор за анализ и разпознаване на реч в различни дялове от подадения изказ. Разработката на университета в Киото може да се счита като предшественик на системата с непрекъснато разпознаване на гласа.

Друга система за разпознаване създава Дийнс (Университетски колеж, Англия) и това е фонемен разпознавател, който разпознава 4 гласни и 9 съгласни [7]. Включвайки статистически данни за валидните фонемни последователности в английския език, те увеличават цялостната фонемна разпознаваща точност за думите, съдържащи две или

повече фонеми. В тази разработка за първи път е употребен статистически синтаксис на фонемно ниво при автоматично речево разпознаване.
Алтернатива на употребата на речеви сегментатор е концепцията за въвеждане на разнородна времева скала за изравняване на речевите модели, предложена независимо от Т. Мартин от лабораториите RCA [8] и Винчук от Съветския съюз [9]. Мартин открива необходимостта за справяне с времевата разнородност в повтарящи се речеви сегменти и предлага набор от решения, включително маркиране на крайните точки на изговорената последователност, което в голяма степен подобрява надеждността на действието на разпознавателя. Винчук предлага използването на динамично програмиране за времето между две изказвания, за да извлече състоятелна оценка за тяхната прилика. Неговата разработка, макар и в голяма степен непозната на запад, предшества тази на Сакое, Чиба [10] и други, представили по-формални методи, познати като динамични времеви изкривявания, при съгласуване на речевите шаблони. След публикацията на Сакое и Чиба, динамичното програмиране, в множество разнообразни форми (включително алгоритъма на Витерби [11]), се превръща в задължителна техника при автоматичното речево разпознаване.

Технологично развитие на комерсиални ситеми

Атал и Итакура независимо един от друг представят основните концепции на LPC (Linear Predictive Coding) [12], [13], като процедура за ефективно кодиране на речевата вълна, която е представена чрез времеви параметри, свързани с реакцията на вокалния тракт. До средата на 1970-те, основните идеи на тази технология са разработвани от Итакура [14], Рабинер и Левинсън [15] и др.
По същото време Том Мартин основава първата комерсиална компания Threshold Technology Inc., предлагащасофтуерна система за речево разпознаване - VIP-100 System. Системата се използва единствено в няколко приложения от фирми за производство на телевизионни фасадни плочи за контрол на количество и от FedEx за сортиране на пакети на конвейерна лента, но главната й значимост е начинът, по който указва влияние върху Агенцията за напреднали изследователски проекти (Advanced Research Projects Agency, ARPA) да започне финансиране по програмата за изследването на речево разпознаване Speech Understanding Research (SUR). Сред системите, създадени по тази програма е *Harpy* на университета Карнеги Мелон [16], която доказва способността си да разпознава реч, използвайки речник от 1011 думи, при това със сравнително голяма точност. В системата Harpy, входната реч, след като премине през параметричен анализ, се сегментира, след което сегментираната параметрична последователност се подлага на сравнение с телефонен шаблон, използвайки разстоянието на Итакура. Търсенето по графи компилира, предполага, изравнява и след това потвърждава разпознатата последователност от думи (или звуци), която удовлетворява ограниченията с най-високия възможен резултат. *Hapry* е първата система, която се използва мрежа от крайни състояния, за да намали изчисленията и ефективно да определи най-близкия съвпадащ низ. Методи за оптимизиране на получената крайна мрежа от състояния се появяват едва след 1990 г. [17]
Други системи, разработени под програмата SUR на ARPA са *Hearsay-II* на *CMU* и *HWIM* на *BBN* [18]. Нито една от двете системи не успява да реализира амбициозната цел на ARPA за ефективност. Въпреки това, подходът, предложен в Hearsay-II за използване на паралелни асинхронни процеси, симулиращи компоненти от източниците на знание в речевата система, е една пионерска концепция. В системата *Hearsay-II* знанията от паралелни източници се интегрира за да се получи следващото ниво на хипотеза. Системата *HWIM* се характеризира с мрежа за лексикално декодиране, обхващаща сложни фонетични правила с цел достигане на точност при разпознаване на фонемите,

способност за справяне с двусмислеността чрез решетка от алтернативни хипотези и концепция за верификация на думи на параметрично ниво. Друга комерсиална система е *DRAGON* на Джим Бекър, който стартира компания със същото име.

Паралелно с проектите на ARPA се оформят две обширни направления в разпознаването на реч и това са разработките на IBM и AT&T Bell Laboratories. Усилията на IBM, водени от Фред Джелинък, са насочени към създаването на „гласово активираща се пишеща машина“ (voice-activated typewriter, VAT), чиято основна функция се свежда до преобразуване на изговорено изречение в последователност от букви и думи, които да бъдат показани на екран или отпечатани на хартия [19]. Създадената разпознавателна система *Tangora* е силно зависима от говорителя - пишещата машина трябва да бъде тренирана от всеки индивидуален потребител. Технологичният фокус е върху познавателната лексика с основен акцент върху офис кореспонденцията и структурата на езиковия модел (граматиката), която е представена от статистически синтактични правила, описващи вероятността за появяване в речевия сигнал на последователността от езикови символи (т.е. фонеми или думи). Прието е такъв тип речево разпознаване да се нарича *транскрипция*. Наборът от статистически граматически или синтактични правила се нарича *езиков модел*, като най-често използван е *n-gram* моделът. Въпреки чв *n-gram* езиковият модел и традиционната граматика са проявления на езиковите правила, техните роли са фундаментално различни. Моделът *n-gram* характеризира връзката между думите в обхвата от *n* на брой думи и в този случай статистическо представяне на граматиката е удобно средство. Ефикасността му при търсенето по думи е демонстрирана в световно известната игра на Клод Шанон [20], в която човекът и компютърът трябва да познаят каква ще е следващата дума в произволно изречение. Човекът познава на базата на вродения опит с езика, а компютърът използва натрупаните статистики с думи, за да направи своя най-добър избор, базиран на максималната вероятност от изчислените честоти за поява на дадена дума. Демонстрирано е, че веднъж, след като наборът от думи надвиши 3 *(n>3)*, много по-вероятно е компютърът да победи човека. След тяхното въвеждане през 80-те, употребата на *n-gram* езиковите модели и на техните варианти, се превръща в неизменна част от системите за разпознаване на реч.

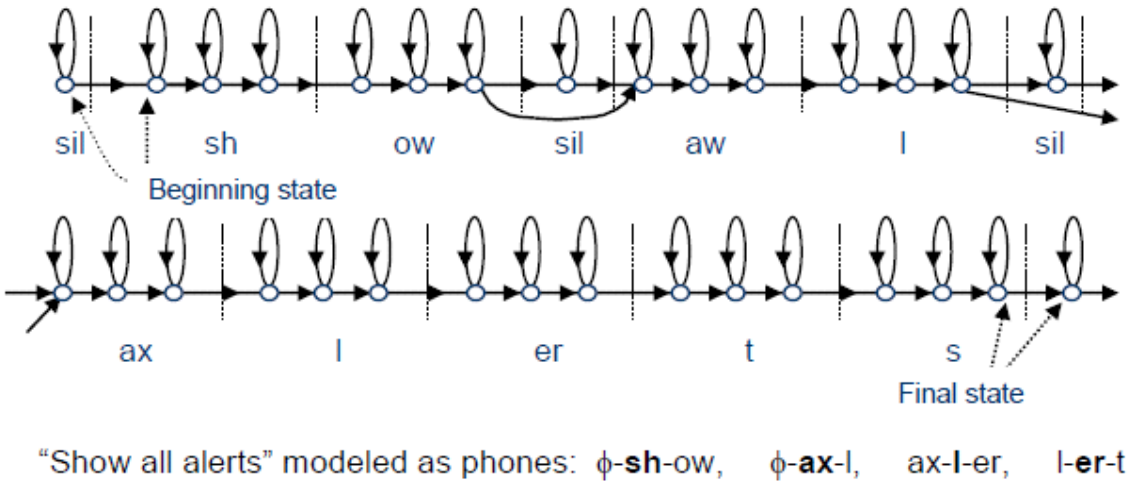
От друга страна, изследователската програма на AT&T Bell Laboratories има за цел да се предоставят автоматизирани телекомуникационни услуги на обществото (пр. гласово набиране) и управление на маршрутизирането на телефонните обаждания. Очакването е тези автоматични системи да работят добре за десетки милиони потребители без нуждата от индивидуален тренинг за всеки говорещ. Основният акцент на Bell Laboratories е системата да е независима от говорителя и да може да се справи с акустичната променливост, присъща на речевите сигнали, идващи от различни говорящи, често с коренно различни местни акценти. В последствие изследването на акустичното разнообразие на различни говорителите води до изучаването на съвкупност от измервания на спектралните разстояния и техники за статистическо моделиране [21], които в резултат дават значително точна репрезентация на произношенията на голям брой хора. Основна използвана техника при непрекъснатото разпознаването на речта е скритият модел на Марков [22], [23]. Тъй като приложения като гласово набиране и маршрутизиране на разговори обикновено изискват късо изговаряне на ограничена лексика и се състоят само от няколко думи, акцентът в изследването на Bell Laboratories се поставя по-скоро върху това, какво всъщност се има предвид под *акустичен модел* (т.е. спектралното възпроизвеждане на думи или звуци) отколкото върху същността на езиковия модел (т.е. представянето на граматиката или синтаксиса). Освен това, от голямо значение в подхода на Bell Laboratories е концепцията *за търсене на ключови думи* като примитивна форма за разбирането на речта [24]. Техниката за търсене на

ключови думи изисква разширяване на типичната парадигма за разпознаване на шаблони с такава, която включва тестване на хипотези.
Подходите на IBM и AT&T Bell Laboratories имат обща характеристика – и при двата ключова роля има математическият формализъм.

Системи за гласово разпознаване, използващи техники на изкуствения интелект

След 1980 г в областта на речевото разпознаване се наблюдава промяна на методологията от интуитивно-шаблонно ориентиран подход към статистическо моделиране, базирано на скрит модел на Марков (*CMM*). Основната идея на *CMM* е бече използвана от лабораториите на IBM и IDA [25] [26], но методологията е доразвита и се превръща в предпочитан метод за речево разпознаване, като тази тенденция остава постоянна през следващите две десетилетия.
CMM оперира със стохастични процеси и са подходящи за моделиране както на присъщата променливост на речевия сигнал и получаващите се спектрални особености, така и на структурата на говоримия език [27]. Реалният речеви сигнал по своята същност е изключително променлив поради вариации в произношението и акцента, както и заради фактора околна среда, като ехо и шум. Когато хората изговарят една и съща дума, акустичните сигнали не са идентични (всъщност могат да бъдат удивително различни), макар основната лингвистична структура по отношение на произношение, синтаксис и граматика може да останва същата. Формализмът на *CMM* използва верига на Марков за представяне на лингвистичната структура и лингвистични разпределения за променливостта на акустичното представяне на звуците. При даден набор от познати изказвания, представляващи достатъчна съвкупност от вариации на желаните думи (наречени тренировъчен набор), може да се използва ефикасен оценъчен метод, наречен алгоритъм на Баум-Велх [28], за да се получи „най-добрия“ набор от параметри, определящи съответстващия модел. Оценката на параметрите, дефиниращи модела е еквивалентна на трениране и обучение. Вероятностната мярка, представена от скрития модел на Марков е важен компонент при системите за речевото разпознаване, които следват подхода за разпознаване на статистически модели, и води началото си от теорията на Бейс за взимане на решение [29].
Идеята за използване на *CMM* възниква в Института за анализ на отбраната в Принстън. Баум определя *CMM* като набор от вероятностни функции на верига на Марков, която по определение включва две вложени разпределения, като едното се отнася до верига на Марков, а другото – до набора от вероятностни разпределения, всяко от които е съотнесено към състояние от верига на Марков [30]. Този двойно стохастичен процес е полезен в приложения за предвиждане на фондовия пазар и крипто-анализът на ротационен шифър, използван по времето на Втората световна война. Техниката на Баум за моделиране и оценка първо се прилага върху дискретни наблюдения, а след това случайните наблюдения се моделирани функции на вероятностната плътност. Двойно стохастичният процес на Баум намира приложение в системите за разпознаване на говор [31] [32], но по-късно става ясно, че ограничението на вида на функциите на плътността налага намаляване на производителността на системите, особено при задачи, независими от говорителя. В лабораториите Bell Laboratories въвеждат *CMM* със смесени вероятностни функции на плътността [33] [34], които оттогава нататък се утвърждават като изключително важна част от осигуряването на задоволителна точност при разпознаването, особено при лексикално големи задачи за разпознаване на реч, независими от говорителя.
CMM е неизменна част от големи декодиращи речта структури с включен езиков модел. Употребата на граматика от крайни състояния при продължителното речево разпознаване на голяма по обем лексика представлява разширяване на *CMM*, при което

структурата на езика се отчита на ниво взаимодействие между артикулацията и произношението. Въпреки, че тези структури са груби приближения на реалния речеви процес, изчислително те са ефективни и често достатъчни, за да се получи смислен начален резултат на изпълнение. Сливането на *CMM* (със своето предимство в статистическата съгласуваност, в частност при работа с акустичното разнообразие) и мрежата на крайни състояния (с нейното търсене и изчислителна точност, по-точно при работа с хипотези за последователност от думи) е важно технологично развитие през средата на 80-те години на XX век.

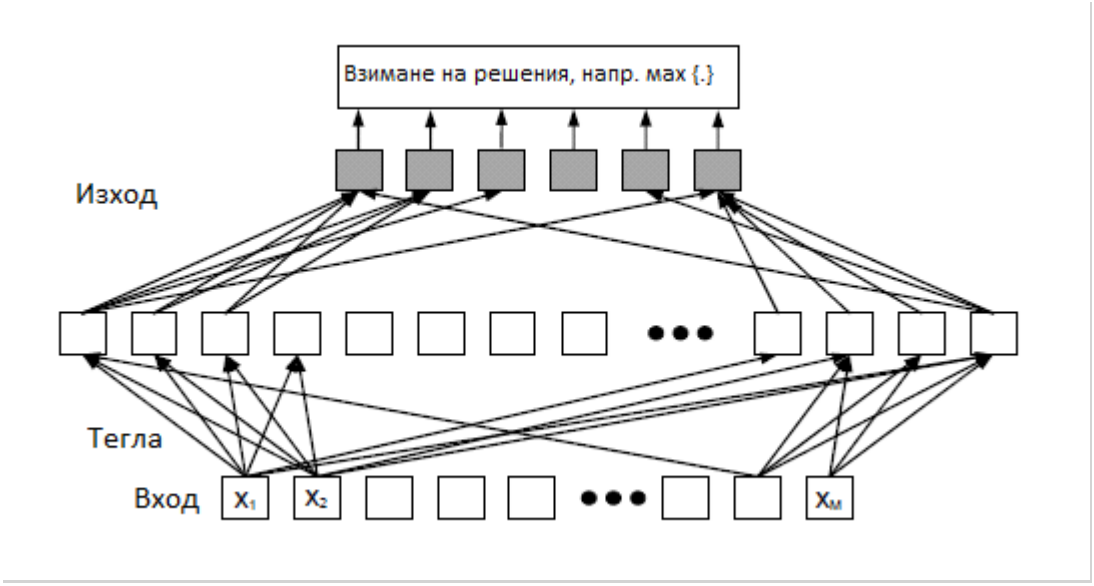


Фиг. 2. Комбинирана мрежа на крайните състояния за изричането на “show all alerts”

На фигура 2 е показан комбиниран модел на крайните състояния за изказването на “*show all alerts*”, съставен от няколко модели на части от думата, които зависят от контекста и представляват съответстващите фонемо-подобни единици (включително единица за тишина *sil*, която може да се появи в началото или в края на изречението, както и в края на всяка дума в изречението, подобно при пауза по време на говорене). Графът на крайните състояния се реализира като верига на Марков за пресмятане на вероятността от спектрално представяни във времето наблюдения на неизвестна реч. Всеки възел в графа се асоциира с вероятностно разпределение, което отговаря на променливостта при реализиране на съответния фонемо-подобен звук. Вероятността изказаното да е генерирано от мрежа на крайните състояния, представена от модела, се изчислява като последователна сума от локални вероятности, отнасящи се до елементарните единици на съставния модел, след като се изпълнява изравняване на състоянията чрез динамично програмиране, за да се максимизира попадението между означените единици и съответстващите части от наблюденията на речта дори и за модели с неправилни последователности от думи.

Във всеки интервал от време има набор от хипотетични единици, а определянето на еднаквостта на звука се основава върху възможно най-високата вероятностна стойност или резултат на съвпадение. Броят хипотетични единици за съвпадение и този на пътищата за търсене може да бъде много голям и да изисква ефикасни изчислителни алгоритми, за да се намери решение на задачата. В средата на 90-те години се разработва средство, наречено библиотека на крайните състояния (*FSM=finite-state machine library,*

което се превръща в неизменна част от всички модерни системи за разбиране на текст и речево разпознаване [35].
Друга технология, използвана обработката на естествени езици е идеята за изкуствена невронна мрежа. Невронните мрежи са представени още през 1943 г., но в началото не показват задоволителни резултати [36]. След появата на модела за паралелна разпределена обработка (parallel distributed processing), и по-конкретно на многослойния персептрон (фиг. 3), използването на невронни мрежи придобива широка популярност. Причината не е в аналогията с дейността на нервната система, а поради способността за апроксимация на произволна функция, при условие, че няма зададено ограничение за сложността на обработващата конфигурация. Ако разпознавателят на шаблони се онагледи като средство, извършващо свързване на функцията от входния модел с идентификатор от класа, тогава може да се каже, че многослойният персептрон би бил идеално средство за тази задача. Основната цел на първоначалните невронни мрежи за разпознаване на речт е съсредоточена върху прости задания като разпознаване на няколко фонеми или малък брой думи. Тъй като разпознаването на говора неизменно изисква боравене с изменения във времето, в своя ранен стадий невронните мрежи не са достатъчно добро средство [37]. Научноизследователска дейност по това време има за цел да интегрира невронните мрежи с главната структура на скрития модел на Марков.



Фиг. 3. Многослоен персептрон

През 90-те години задачата за разпознаването на шаблони постепенно се преобразува в оптимизационна задача за минимизиране на емпиричната грешка [38], защото главната цел при проектирането на разпознавателя е достигането на минимална грешка при разпознаване, а не най-добро приспособяване на функцията на разпределение към данните според критерия на Бейс. Концепцията за минимална емпирична грешка последователно дава като резултат набор от техники, една сред които е дискриминиращото обучение и базираните на ядра методи, като помощно векторните механизми (*SVM= support vector machines*) [38], [39].

Успехът на статистическите методи съживява интереса на ARPA към системите за разпознаване, включително системата *Sphinx* на CMU [40], системата *BYBLOS* на BBN [41] и системата *DECIPHER* на SRI [42]. В системата *Sphinx* успешно е интегриран

статистическия метод на *CMM* със предимствата на по-ранната система *Harpy* за търсене в мрежа. В резултат системата обучава и вгражда контекстово зависими телефонни модели в сложна лексикална декодираща мрежа, постигайки значителна точност при речевото разпознаване на голям брой от думи.
В ARPA полагат усилия и за създаване на критерии за оценка на технологиите за разпознаване на говор. Подобни оценки главно използват мерки за степента на грешка в дума и изречение, както и обхвата на системата за разпонзване. Оценки систематично се прилагат върху внимателно проектирани задания с нарастващо ниво на сложност, вариращи от разпознаване на последователен говор, произнесен със стилизирана граматическа структура (пр. при управление на военоморски ресурси) до транскрипции на живи телевизионни предавания и на ежедневна реч (пр. NAB включва лексика от над 20 млн. думи). Една от разработките е *ATIS* - задача, включваща прост спонтанен речеви разговор с автоматизирана система за извличане на информация от въздушния транспорт. Друга разработка *WSJ* се отнася до транскрипцията на набор прочетени параграфи от списанието Wall Street, като обемът на лексиката достига до 60 млн. думи. Най-голямото предизвикателство, предложено от ARPA е заданието *Комутационно табло (Switchboard)*. Речта е разговорна и спонтанна, с много инстанции на т. нар. несигурности, като например непълни думи, колебание, поправяне и т.н. Основното заключение, което може да се извлече от тези резултати е, че получаването на очаквания резултат при обработване на разговорна реч, която не е задължително да се подчинява на лингвистични ограничения, с значително по-трудно от обработване на реч, която следва стриктни граматически и семантични правила, В допълнение, оценъчната програма показва, че увеличението на количеството речеви данни, използвани за изчисление на разпознавателните параметри (т.е. размерът на обучаващата серия) винаги води до намаляне нивото на грешката в думите, като е прието е нивото на грешка да бъде бъде под 10%.

След 90-те години се наблюдава съществен подем на много изследователски програми по света. Всички системи за обработка на естествени езици стават по-сложни и точни. Добре структурираната базова софтуерна система е незаменима в по-нататъшното проучване и развитие за комбиниране на нови концепции и алгоритми. Системата *Hidden Markov Model Tool Kit (HTK)*, разработена от екипа на университета Кеймбридж с ръководител Стийв Янг, е една от най-широко използваните софтуерни системи при проучванията за автоматичното разпознаване на речта [43].

Комуникация човек-машина

В ранните етапи от развитието на приложенията за разпознаване на реч изследователите се фокусират върху преобразуване на акустичната реализация на лингвистично събитие в думи. Концепцията, че процесът реч-към-текст е задължителна първа стъпка от комуникацията човек-машина, има своите последователи и до днес. Но при такава комуникация потребителите обикновено изговарят изречения, които в повечето случаи не удовлетворяват напълно граматическите ограничения на разпознавателя и включват нелексикални, а изговорените изкази често са влошени от лингвистично ненужни шумови компоненти като заобикалящ шум, неприсъщи акустични звуци, смущаваща реч и др. Освен това, както и при комуникацията човек-човек, употребата на речта често изисква диалог между потребителя и машината, за да се постигне някакво удовлетворително ниво на разбиране. Подобен диалог често изисква операции като заявка и потвърждение. Методът за търсене на ключови думи е решение за първия проблем, докато вторият насочва вниманието на изследователската общност върху обработката на диалога. Приложения, в реалистично се пресъздават комуникационните

възможности на човека са например системите *Pegasus* и *Jupiter* на екип от Института по технологии в Масачузец с ръководител на Виктор Зю [44] [45] и системата *How May I Help You (HMIHY)* на AT&T, дело на Ал Горин, която от 2000 година е неизменна услуга при обслужването на клиенти на фирмата [46].

Pegasus е система за разговорна реч, която осигурява информация за състоянието на въздушната линия над обикновена телефонна линия. *Jupiter* е подобна система, фокусирана върху достъпа до информация за локално и национално метеорологичното време. С добре проектирано управление на диалога, тези системи насочват потребителя да предостави нужната за изпълнение на заявката, сред малък и безрезервен набор от меню с опции, без излишно да искат детайли за заявката. Управлението на диалога често включва вградена информация за разпознати фрази и софт поправяне на грешки, така че потребителя да реагира така, сякаш в другия край на телефонната линия има реален човешки посредник, а не машина. Целта е да се проектира машина, която комуникира, а не само разпознава думи от изречен изказ.

Въпреки, че автоматичното разпознаване на говор и системите за разбиране на реч не са напълно съвършени по отношение на прецизността на думите/заданията, добре развитите приложения се възползват от съществуващата технология, за да предоставят реална стойност на клиента и броят и обхвата на такива системи постоянно се увеличава.

III. ВИДОВЕ ГЛАСОВО РАЗПОЗНАВАНЕ

Дискретно разпознаване на глас

При дискретното разпознаване на глас (*Discrete Voice Recognition*) се използва основно техниката *разпознаване на изолирани думи* (*Isolated Word Voice Recognition*). Разпознаването на изолирани думи е процес, при който след всяка дума трябва да бъде направена пауза за отделяне. Паузата, или липсата на звук е основен подход за определяне на началото и края на изговорена дума.

Дискретното разпознаване на глас извършва преобразуване на глас в текст дума по дума. В по-напредналите версии на дискретно гласово разпознаване се конвертират цели фрази, като това обикновено е едно изречение. Дискретно разпознаване на глас се използва предимно в приложения за диктовка и при гласова навигация с команди.

Непрекъснато разпознаване на глас

При непрекъснатото разпознаване на глас (*Continuous Voice Recognition/Connected Words Voice Recognitions*) началото и края на думата може да бъде определено, без да е необходима звукова пауза е вид разпознаване. Този вид разпознаване е сложна форма на разпознаване и изисква много по-голям ресурс.

Непрекъснатото разпознаване на глас също се нарича свързано гласово разпознаване, тъй като не изисква значителна пауза между думите. Вместо това говорителят може да говори по-естествено и въпреки това софтуерът за гласово разпознаване разбира къде започва думата и къде свършва.

Спонтанно гласово разпознаване на реч

Спонтанното гласово разпознаване на речта *(Spontaneous Speech Voice Recognition/ Natural Speech Voice Recognition)* е осъвършенствана форма на непрекъснатото разпознаване. Несъзнателни звуци, които хората издават, докато говорят като едновременно мислят и оформят нерепетирани изречения, могат да бъдат разбрани от софтуера и съответно да бъдат пропуснати при преобразуването. Ползата от спонтанното

гласово разпознаване на речта е, че не е нужно да се променя начина на говорене с компютъра.
Спонтанното гласово разпознаване на говор се нарича още "естествено разпознаване на гласа", тъй като поддържа разпознаване на естествения начин, по който говори човек.

Разпознаване на глас, зависещ от говорещия

Разпознаване на глас, зависещ от говорещия (*Speaker Dependent Voice Recognition*), е вид разпознаване, който зависи от гласа на говорещия човек. Този вид разпознаване изисква обучение, за да стане преобразуването на речта в текст по-точно. Обучението се извършва от говорителя. Резултатът е система за гласово разпознаване, която разбира специфичния акцент и глас на съответния човек.
Системите за разпознаване на глас, които зависят от говорителя, могат да бъдат дискретни или непрекъснати.

Разпознаване на глас, независещ от говорещия

Разпознаване на глас, независещ от говорещия (*Speaker Independent Voice Recognition*) е тип разпознаване на глас, което е обратното на разпознаване на глас, зависещ от говорещия. Този вид разпознаване не изисква обучение и може да разбере реч от широк кръг говорители.
Системите за разпознаване на глас, независими от говорещия, изискват от системите повече възможности за обработка и в общия случай не са толкова точни при преобразуването на речта в текст.

Разпознаване на глас от естествен език

Разпознаване на глас от естествен език (*Natural Language Voice Recognition*) всъщност е добавка към непрекъснатото спонтанно разпознаване на глас и се отнася до способността на компютъра да разбере въпрос или команда, казана по естествен начин, а не по структурирания начин, който очаква системата. Обикновено това се отнася до способността на компютъра и да отговаря на говорещия с естествен език, като че ли се провежда разговор.
Естественото разпознаване на глас е това, което управлява съвременните виртуални асистенти като *Сири*, *Алекса* или *Кортана*.

IV. ТЕХНИКИ ЗА ГЛАСОВО РАЗПОЗНАВАНЕ

Смисловата обработка на естествения език (която включва анализ и синтез) първоначално се свързва със ситаксиса и синтактичния анализ. Скоро обаче се установява, че оставането само на синтактично равнище не дава задоволителни резултати, защото понякога прилагането на едно или друго синтактично правило зависи от смисъла на самия текст. Появява се необходимостта от синтактично-смислов анализ и разбиране на текста и следователно извличане на знанията от текста и представянето им в някаква система. При решаването на този проблем се налагат редица ограничения, например ограничения върху предметната област, на знанията, на допустимите конструкции в езика, на обема на анализираната информация и т.н.

Морфологичен анализ

Морфологичният анализ означава определяне на думата като част на речта и на всички нейни граматически характеристики. Всички форми на думата образуват нейната парадигма. Промяната на думата от една словоформа в друга става чрез промяна на структурата на думата, т.е. чрез прибавяне и отнемане на морфеми (суфикси). Морфемата е най-малката значеща единица в езика.
Думата морфема има гръцки произход и означава значимата част на една дума (корен, представка, наставка, вставки и окончания). При морфологичния анализ се анализират морфемите на думите, т.е. техните компоненти и признаци, и неучастващите в думите признаци (например препинателните знаци – пунктоация) се отделят от думите [47].

Синтактичен анализ

Синтаксисът е дял от граматиката, който се занимава със строежа на изречението и с неговите части. В тази стъпка на гласовото разпознаване изреченията, които са линейни последователности от думи, се преобразуват в структури, които показват как думите се отнасят една към друга. Тази стъпка на граматически разбор, наречена парсване (*parsing)*, превръща линейния списък от думи в изречението в структура, която дефинира представените единици от този списък.

Семантичен анализ

Семантиката е дял от граматиката за значението на думите, на техните части и съчетания. При семантичния анализ на структурите, създадени от синтактичния анализатор, се присвояват стойности. В повечето езикови пространства словосъчетанието *“Безцветната зелена жаба”* ще бъде отхвърлена като семантично неправилна. Тази стъпка съпоставя на отделните думи подходящи обекти от базата знания и трябва да създаде коректни структури, които показват по какъв начин значенията на отделните думи се комбинират едно с друго.

Интегриране в съответен контекст

Значението на дадено изречение може да зависи смислово от изреченията, които го предхождат. Обектите, въведени в изречението, трябва или да са експлицитно определени, или да се отнасят към обекти от предишни изречения. Целият контекст трябва да бъде ясен и това е целта на тази стъпка.

Прагматичен анализ

Структурата, представяща какво е било казано преди, се интерпретира отново, за да се определи, какво всъщност означава.

Обработка на сигнала

Обработката на сигнала представлява взимане на говорима част от езика и да се изрази в последователност от думи. Тук се включва дигитализиране на сигнала и различаване на сегментите на думата, които могат да бъдат асемблирани в цели думи. Езиковите елементи, обработвани като сигнали, се наричат фонеми.

АЛГОРИТМИ

Акустичното моделиране и езиковото моделиране са важни части от съвременното базирано на алгоритми статистическо разпознаване на реч. Скритите модели на Марков

са широко използвани в много системи. Езиковото моделиране има и много други приложения като интелигентна клавиатура и класификация на документи.

Скрити модели на Марков

Модерна система с общо предназначение за разпознаване на реч се основава на скритите модели на Марков, тъй като речевият сигнал може да се разглежда като отделни фрагменти от стационарния сигнал, т.е. като част от целия сигнал. За малък времеви интервал (пр. 10 милисекунди), говорният сигнал може да се приеме за почти стационарен процес.

От гледна точка на теорията на вероятностите един случаен процес притежава свойство на процес на Марков, ако условното разпределение на вероятностите на бъдещи състояния на наблюдавания процес, при предварително известни текущи и минали състояния, зависи само от състоянието към този момент, и не зависи от миналите такива. С други думи бъдещите състояния на един процес са условно независими от миналите състояния, при развитие на наблюдавания процес. Процес, който притежава това свойство се нарича процес на Марков. Най-популярните процеси на Марков са веригите на Марков. Верига на Марков е процес на Марков, който приема стойности от дискретното множество, наречено пространство на състоянията, като стойността му се променя в точно определени моменти от време. Тези вериги са широко разпространени и използвани в теорията на разпознаване на реч, тъй като заемат много малко памет за моделиране на динамични процеси и освен това са лесни за реализация.

В статистическото моделиране скрит модел на Марков е процес на Марков с неизвестни параметри. Всеки определен параметър е функция на вероятностите на дадено състояние. Всяко състояние може да се възпроизведе, което не позволява да се определи поредицата от състояния до момента на наблюдение. С други думи поредицата от състояния не може да бъде наблюдавана и остава скрита.

В общия модел на Марков, състоянието е директно видимо от наблюдателя и за това вероятните състояния на промяна са единствените параметри. В скрития модел на Марков състоянието не е пряко видимо, но променливите, повлияни от състоянието, са видими.

Скритите Марковски модели са особено известни с приложенията си в разпознаването на реч, ръкопис, жестове, и др.

Динамично програмиране

Разпознаването на реч основано на *DTW (Dynamic Time Warping)* е ефективен метод при разпознаващи системи с малък речник от думи.

Алгоритъм за динамично програмиране [48]

Стъпка 1: Инициализация

$D(1,1) = d(1,1), B(1,1) = 1, for\ j = 2, \dots, M$ и се изчислява $D(1,j) = \infty$

Стъпка 2: Итерации

при $i = 2, \dots, N$

{

за $j = 1, \dots, M$ се изчислява

{

$$D(i,j) = \min_{1 \le p \le M} [D(i-1,p) + d(p,j)]$$

$$B(i,j) = \arg \min_{1 \le p \le M} [D(i-1,p) + d(p,j)]$$

}

}

Стъпка 3: Връщане назад и преустановяване на процеса

Оптималното разстояние е $D(N, M)$ и оптималния път е $(s_1, s_2, \ldots, s_N)$, където

$$s_N = M \text{ и } s_i = B(i+1, s_{i+1}), i = N-1, N-2, \ldots, 1.$$

Алгоритъм Forward за оценка на скрит модел на Марков [49]

Стъпка 1: Инициализация

$$\alpha_1 = \pi_1 b_1(X_1), 1 \le i \le N$$

Стъпка 2: Индукция

$$\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i)\alpha_{ij}\right] b_j(X_t), \qquad 2 \le t \le T;\ 1 \le j \le N$$

Стъпка 3: Преустановяване на процеса

$$P(X|\Phi) = \sum_{t=1}^{N} \alpha_T(i)$$

Ако е необходимо да се завърши в крайно състояние,

$$P(X|\Phi) = \alpha_T(s_F)$$

Декодиране на модел на Марков с алгоритъм на Витерби

Стъпка 1: Инициализация

$$V_1(i) = \pi_i b_i(X_1)$$
$$B_1(i) = 0$$

Стъпка 2: Въведение

$$V_t(j) = \max_{1 \le i \le N}[V_{t-1}(i)a_{ij}]b_j(X_t), \qquad 2 \le t \le T;\ 1 \le j \le N$$
$$B_t(j) = arg\max_{1 \le i \le N}[V_{t-1}(i)], \qquad 2 \le t \le T;\ 1 \le j \le N$$

Стъпка 3: Прекратяване

$$The\ best\ score = \max_{1 \le i \le N}[V_t(i)]$$
$$S_T^*(j) = arg\max_{1 \le i \le N}[B_t(i)]$$

Стъпка 4: Връщане назада

$$S_t^* = B_{t+1}(S_{t+1}^*), \qquad t = T-1, T-2, \ldots, 1$$

$S^* = (S_1^*, S_2^*, \ldots, S_T^*)$ е най-добрата последователност

Оценяване на параметрите на модел на Марков с Forward-backward алгоритъм на БаумУелч

Стъпка 1: Инициализация

Избира се първоначалната оценка Φ.

Стъпка 2: Е-Стъпка

Изчисляване на спомагателната функция $Q(\Phi, \widehat{\Phi})$.

Стъпка 3: М-Стъпка

Изчисляване на $\widehat{\Phi}$, така че да максимизира спомагателната Q-функция.

Стъпка 4: Итерация

Полага се $\Phi = \widehat{\Phi}$ и алгоритмът се повтаря от стъпка 2 до достигане на желаната сходимост.

Ограничения на скритият модел на Марков

В стандартните скрити модели на Марков има редица ограничения като:

- използване на непрекъснато експоненциално разпределение;
- вероятността на прехода зависи само от източника и предназначението;
- всички наблюдавани състояния са зависими само от състоянието, което ги поражда, без значение от наблюдаваните съседни състояния.

Изследователите са предложили редица техники за справяне с тези ограничения, макар че това не гарантира значителни подобрения в точността на разпознаване на реч, като използване на непрекъснато моделиране, условно независими предпоставки и други.

Изменения на речевия сигнал. Изчисляване на грешката при разпознаване.

Характеристики на речевите сигнали

За постигане на ефективност при разпознаване на реч задължително се взимат под внимание следните три фактора:

1) изменения в стила на речевата вълна:
   a. произнасяне на отделни думи или кратки фрази с пауза между тех;
   b. непрекъснато произнасяне;
   c. скорост на говорене;
   d. сила на речевата вълна - шепнене, нормален говор и викове;
2) индивидуални гласови характеристики- всеки глас е уникален, различен и подчинен на физическите особености, които влияят върху дължината, ширината и размера на вокалния тракт, пола, възрастта, диалекта, здравословното състояние, образованието и не на последно място маниера на говорене;
3) влияние на околната среда.

Оценка на грешката

При разпознаване на реч основно се разграничават три вида грешки:

- замяна - правилната дума е заменена с неправилна;
- изтриване - дадена дума е пропусната;
- вмъкване - добавена е допълнителна дума.

Задачата за максималното съответствие се състои в съпоставяне на броя на разпознати думи към чроя правилни думи и пресмятане на броя на заменените (*Subs*), изтритите (*Dels*), и вмъкнатите (*Ins*) думи.

Степента на сгрешени думи (*Word Error Rate*) се дефинира като:

$$Word\ Error\ Rate = 100\% * \frac{Subs + Dels + Ins}{N},$$

където N е броят на всички думи в изречението.

Алгоритъм за пресмятане на степента на сгрешени думи (word error rate):

Стъпка 1: Инициализация

$R[0,0] = 0$
$R[i,j] = \infty$
$if\ (i < 0)\ or\ (j < 0)\ B[0,0] = 0$

Стъпка 2: Итерации

$for\ i = 1, \dots, n\ \{$
  $for\ j = 1, \dots, m\ \{$

$$R[i,j] = \min \begin{bmatrix} R[i-1] + 1 & \text{(изтриване)} \\ R[i-1, j-1] & \text{(съответства)} \\ R[i-1, j-1] + 1 & \text{(замяна)} \\ R[i, j-1] + 1 & \text{(вмъкване)} \end{bmatrix}$$

$$B[i,j] = \begin{cases} 1, & \text{ако думата е изтрита} \\ 2, & \text{ако думата е вмъкната} \\ 3, & \text{ако думата съответства} \\ 4, & \text{ако думата е заменена} \end{cases}$$
$$\}$$
$$\}$$

Стъпка 3: Връщане назад и прекратяване

$$\text{Степен насгрешени думи} = 100\% * \frac{R(n,m)}{n}$$

$$\text{Оптимален обратен път} = (S_1, S_2, \dots, 0) \text{ където } S_1 = B[n,m]$$

$$S_t = \begin{bmatrix} B[i-1,j], & \text{ако } S_{t-1} = 1 \\ B[i,j-1], & \text{ако } S_{t-1} = 2 \\ B[i-1,j-1], & \text{ако } S_{t-1} = 3,4 \end{bmatrix} \text{ за } t = 2, \dots, \text{докато } S_t = 0$$

Акустично моделиране. Извличане на характеристики от речевия сигнал и прилагане на адаптивни техники за намаляване на грешката

Преобразуване на свойствата на речевата вълна

За нормализиране на характеристиките на говорещия се използват:

- линеен дискриминантен анализ;
- картографиране, основано на невронни мрежи;
- изкривяване на честотния спектър при нормализиране на дължината на говорния тракт, чрез честотно мащабиране на Мел и билинейна трансформация.

За редуциране на външните различия между различни говорещи се използва коефициент на изкривяване, пресметнат за всеки от тях.

Избор на априорни единици при фонетичното моделиране

Фонетичните системи са свързани с конкретния език и неговите лингвистични правила. Важните елементи, които трябва да бъдат моделирани, са начина на представяне на акустичната тишина при липса на речеви сигнал и фонетичната информация за езика. Единиците, които се моделират трябва да бъдат:

- точни - да представят акустичните особености на фонемите в различните им употреби;
- обучени - да имат достатъчно данни за правилно оценяване на съответния елемент;
- породени - всяка нова дума да може да бъде получена като производна на предварително дефинирани елементи в системите за разпознаване на реч.

Сравнение на различните елементи

Независимо от това дали се използват контекстно зависими, или контекстно независими модели на думи, е необходимо дефинирането и автоматичното определяне на всички възможни фонологични вариации в конкретния език. Когато тези модели на думи са адекватно обучени, те предоставят най-добрите резултати в процеса на разпознаване. Следователно при разпознаване на малък речник, моделите от цели думи са много по-широко използвани, тъй като са точни и не е необходимо да бъдат обобщени и всеобхватни. Моделите на думи са точни при наличието на достатъчен набор от данни, като така биват обучени само върху ограничен набор от задачи, но така тяхното реално приложение е ограничено.

Компромисът между моделите, основани на думи и тези, основани на фонетика е, че при вторите се използва по-големи единици (пр. срички). Тези единици включват обособени групи, различаващи отделните фонетични единици (фонетични клъстери), които обхващат всички изменения на една фонетична единици, в зависимост от конкретиката на контекста. Центърът на клъстера е точка, независеща от контекста, докато началната и крайната точки са силно определени от контекста.

Контекстно зависими елементи

Използването на контекстно зависими елементи значително подобрява точността в процеса на разпознаване, при условие че има достатъчно данни за оценяване на тези контекстно-зависими параметри. Контекстно-зависимите фонеми са широко използвани за разпознаване на реч при голям речник от думи, тъй като значително подобряват тяхната точност и ефективност. При този метод от значение са съседните фонеми, намиращи се в ляво и в дясно.

Ако използваният фонетичен модел е контекстно зависим, качеството на разпознаване значително може да бъде подобрено при условие, че има достатъчно данни за обучение и всички контекстно-зависими параметри са определени. Контекстно-зависимите фонеми са широко използвани в системите за разпознаване на реч при голям речник.

Съществува фонетичен модел, който взема под внимание както левите, така и десните съседни фонеми и това е трифонният модел. Ако две фонеми са едни и същи, но с различен ляв или десен контекст, те се разглеждат като различни трифони. Различните реализации на една фонема се наричат алофони. Трифоните са частен случай на алофоните. Моделиране на контекстно-зависими фонеми намиращи се в отделни думи е сложен процес.

Клъстеризиране на акустично-фонетични единици

Балансирането на способността за обучение и точността между фонетичния модел и модела, използващ думи, може да се обобщи чрез моделиране на подфонетични събития. В действителност резултатите, постигнати както с фонетичните, така и с подфонетичните данни, са сравнително еднакви. Това е ключовото предимство пред моделирането на цели думи. [49]

При подфонетичното моделиране се разглеждат състояния във фонетични скрити модели на Марков. След допълнително обобщавана на изходния клъстер, резултатите са разпределени според различните фонетични модели. По този начин всеки клъстер представлява набор от сходни Марковски състояния, наричен сенон. Така моделът от поддуми, който се състои от последователност от сенони, след тяхното клъстеризиране е завършен. Оптималният брой от сенони за една система се определя от наличнта база за обучение и може да бъдат допълнен в последствие.

Адаптивни техники и намаляване на несъответствията

Разработени са редица техники за използването на адаптивни техники с цел намаляване на несъответствията. Може да се използва неинтрузивен обучителен процес, който през цялото време работи на фонов режим. При обучение без учител параметрите на модела могат непрекъснато да бъдат променяни, така че всички несъответствия се отстраняват.

Системите, които транскрибират реч не в реално време могат да използват различни комбинации на техники за разпознаване.

Тъй като резултатите от разпознаването могат да бъдат непълни, съществува възможност за отклонения, ако нивото на грешката при разпознаване е високо.

Дори и нивото на грешка е ниско, адаптираните резултати не са толкова добри, колкото при обучение с учител. Чрез този вид обучение могат да се управляват широка гама от параметри:

- проверка на фоновия шум, когато дикторът не говори;
- регулиране на усилването на микрофона при нормален говорене на диктора;
- адаптиране на акустичните параметри чрез прочит на няколко изречения от диктора;
- промяна на декодиращи параметри, така че да се постигне най-добрата скорост без загуба на точност;
- динамично създаване на нови записвани изречения въз основа на модели за грешки, специфични за потребителя.

Максимална апостериорна оценка

Максималната апостериорна оценка *(Maximum aposteriori – MAP)* може ефективно да се справи с проблема, свързан с недостатъчни данни, тъй като може да се възползва от предварителната информация за съществуващите модели. Параметрите могат да бъдат коригирани при повторното обучение, така че ограничените нови обучаващи данни биха могли да променят параметрите на модела, ръководен от предварителните знания, и така да се компенсира резултатът от несъответствията. Подобряване на параметрите предотвратява големите отклонения, и освен това позволяват подобряване на крайните резултати в процеса на разпознаване.

Максимална вероятностна линейна регресия

Максималната вероятностна линейна регресия *(Maximum likelihood linear regression – MLLR)* отговаря на основните критерии за създаване на скрити модели на Марков, в случай, че броят на свободните параметри се запазви. Тъй като параметрите на преобразуването могат да бъдат оценени при сравнително малко адаптирани данни, то максималната вероятностна линейна регресия е подходяща за ускоряване на адаптирането. Този метод е широко използван за адаптиране на модели, като при нови диктори, така при промяна на средата.

Сравнение на максимална вероятностна линейна регресия и максималната апостериорна оценка

Максималната вероятностна линейна регресия може да бъде комбинирана с максималната апостериорна оценка. Това гарантира, че с увеличаване на обема от данни за обучение ще има не само компактна максимална линейна регресионна преобразуваща функция за бързо адаптиране, но също така и директно преобразувани параметри на модела.

Клъстерни модели

Разпределението по отделни клъстери отчита разнообразни вариации спрямо различните фонетични класове. Освен това е възможно допълнително класифициране на подфонетично ниво.

Доверителни оценки

Един от повечето критични компоненти в практическите системи за разпознаване на реч е свързана с надеждността на доверителните оценки. В системите за разпознаване на реч, зависещи от диктора и тези, които не зависят от диктора, за определяне на доверителни оценки се използват записи на дикторите за премахване на думите, които могат да бъдат объркани. Това е критично в системите с обучение без учител на дикторите, позволявайки избирателно използване на резултатите от разпознаването, така че траскрипциите с по-малка сигурност и яснота могат да бъдат отхвърлени от процеса на адаптация.

Изкуствени невронни мрежи

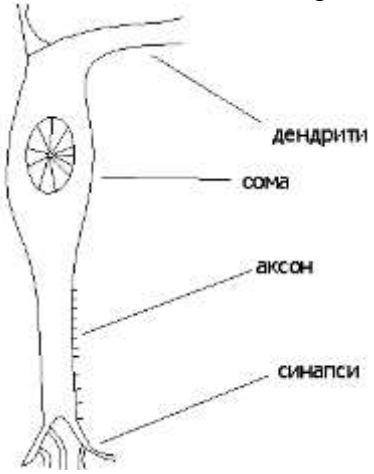
Изкуствените невронни мрежи са модели на биологичните невронни мрежи.
Невронната мрежа е система за паралелно обработване на информация, която има възможност да съхраненява и използва експериментални данни и знания. Тя моделира дейността на своя биологичен еквивалент - мозъка в следните два аспекта:

- информацията се натрупва в мрежата чрез процес на обучение;
- силата на връзките между отделните възли (неврони) се моделира с тегла на съответните връзки, които се използват за съхранение на информацията.

Най-общо невронните мрежи се състоят от прости елементи за обработка на информация наречени неврони или възли. Невроните са свързани и теглата на връзките между тях определят силата на съответните връзки. Входната информация за всеки неврон е претеглената сума от сигналите от останалите неврони. Тази информация се акумулира в неврона, като изходния му сигнал се определя посредством активационна (предавателна) функция.

Биологичен неврон

Опростен модел на биологичен неврон е показан на фигура 4. Той се състои от синапси, дендрити, сома (тяло на клетката) и аксон. Дендритите и аксона са израстъци на клетката, чрез които тя е свързана с останалите клетки. Посредством дендритите, клетката получава сигнали от аксоните на другите клетки в мозъка. В синапсите се осъществява връзката и сигналите се преобразуват чрез сложни електрохимични процеси. Сигналите от синапсите се предават по дендритите към тялото на клетката, където се агрегират. Когато нивото на агрегирания сигнал надхвърли определена граница, се генерира сигнал по аксона на клетка и така тя въздейства на останалите нервни клетки.



Фиг. 4. Биологичен неврон

Абстрактен неврон

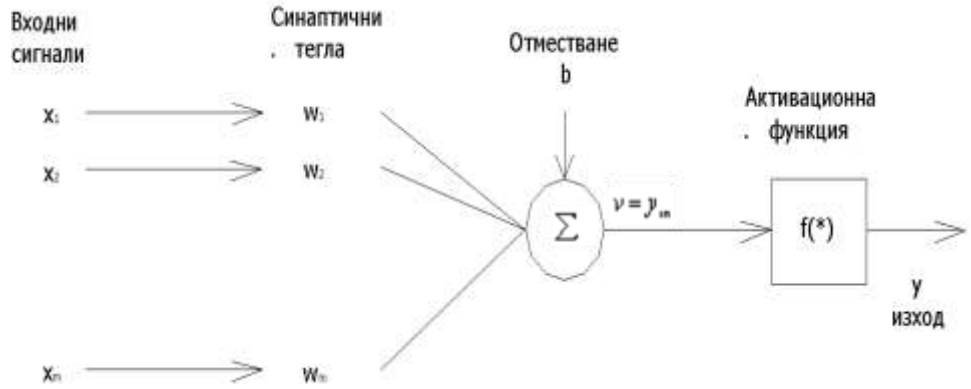
Изкуствените невронните мрежи съвсем не са точни и пълни модели на техните биологични аналози. Те са силно опростени и реализират само малък брой от техните добре изучени и изяснени структурни и функционални характеристики (фиг. 5).
В общия случай всеки неврон има много входове, които са модел на дендридите и един изход, който е модел на аксона. На входовете $x_1, x_2, \dots, x_m$ постъпват сигналите към неврона. Те могат да са външни сигнали или сигнали от изходите на други неврони. С всеки вход е свързано тегло $w_j, j = 1, 2, \dots, m$ (синаптично тегло), което моделира силата на връзката при предаване на сигнала посредством синапса, свързан със съответния дендрит. Агрегирането на входните сигнали в тялото на неврона се моделира със сума за пресмятане на изходния сигнал:

$$v = y_{in} = \sum_{j=1}^{m} w_j x_j + b,$$

където b е сигнал с постоянна стойност, наречен отместване.
Тялото на нервната клетка се моделира със суматора и блока на активационната функция. В общия случай активационната функция е нелинейна и изходният сигнал се изчислява по формулата:

$$y = f(y_{in}) = f\left(\sum_{j=1}^{m} w_j x_j + b\right).$$

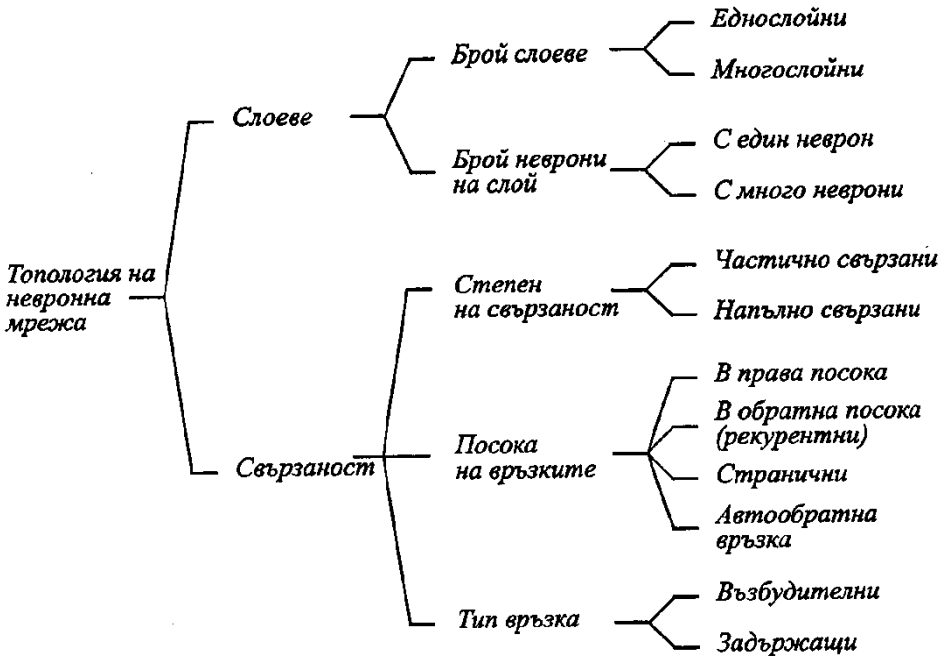
В зависимост от начина на формулиране на активационната функция, изкуствените неврони могат да имат статично и динамично поведение. [50]



Фиг. 5. Модел на абстрактен неврон

Топология и Архитектури на изкуствените невронни мрежи

Топологията на невронните мрежи определя начина, по който множеството неврони в мрежата са подредени и свързани. Тя е най-важната характеристика на невронните мрежи, защото определя функционалните им особености и областите на приложението им. Признаците, които определят различните топологии са представени на фигура 6.
Броят слоеве и неврони в даден слой зависи от сложността на решаваната задача. Невроните могат да се свързват по различен начин и в зависимост от това да имат различно поведение спрямо информационния поток в мрежата. Когато даден неврон е свързан с малък брой други възли, мрежите се изграждат като частично свързани. Но в повечето приложения те се проектират като напълно свързани, а след трениране на мрежата несъществените връзки получават нулеви тегла, което е равносилно на отсъствие на връзка.

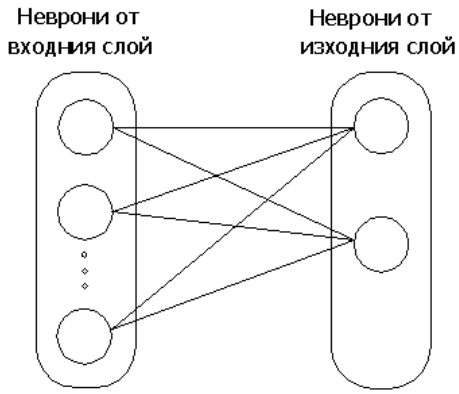


Фиг. 6. Топология на изкуствените невронни мрежи

Мрежите с *прави връзки* предават информация само в една посока – от входа към изхода. Ако има връзки в обратна посока, мрежите се наричат с *обратна връзка* или *рекурентни*. Възможно е да има и странични връзки между невроните в един слой. Тогава мрежите се наричат *странични*. Когато в мрежата се предвиди самите неврони да имат обратна връзка, те се наричат мрежи с *автообратна връзка*.
Връзките могат да бъдат *възбуждащи* или *поглъщащи*. Това се постига чрез задаване съответно на положителни или отрицателни стойности на теглата на съответната връзка. Например, при мрежите със състезателно обучение има както възбуждащи така и поглъщащи връзки. [51]

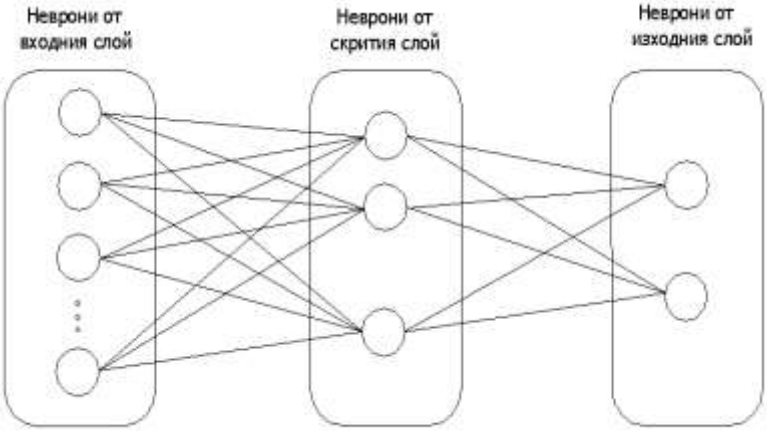
В приложенията най-често срещаните архитектури на невронните мрежи са:

- еднослойни мрежи с еднопосочно предаване на сигнала (фиг. 7.);



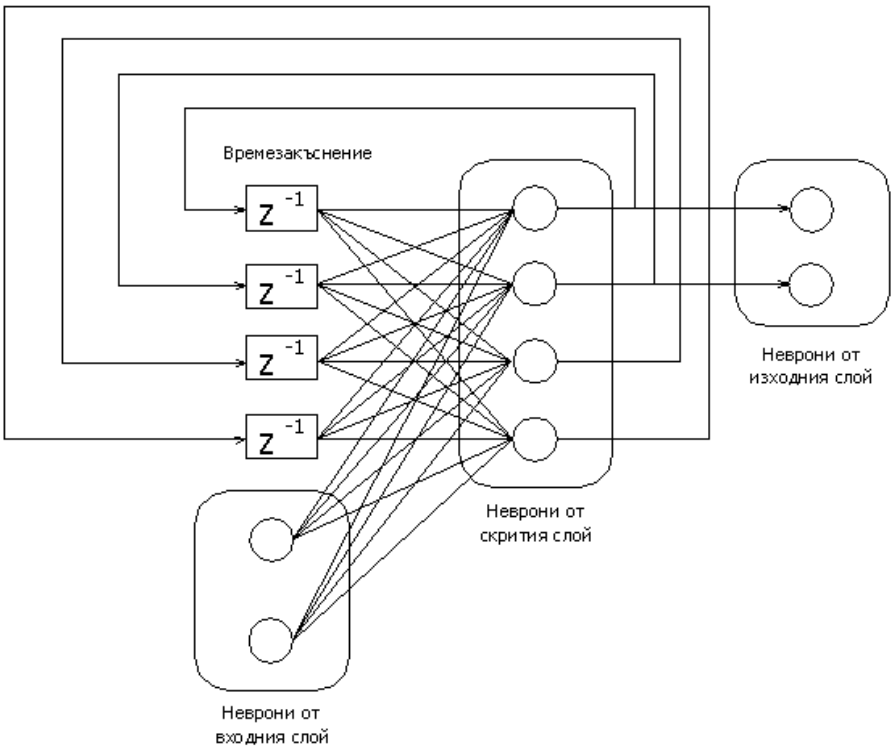
Фиг. 7. Еднослойна мрежа с еднопосочно предаване на сигнала

- многослойни мрежи с еднопосочно предаване на сигнала (фиг. 8.);



Фиг. 8. Многослойна мрежа с еднопосочно предаване на сигнала

- рекурентни мрежи с обратни връзки (фиг. 9).



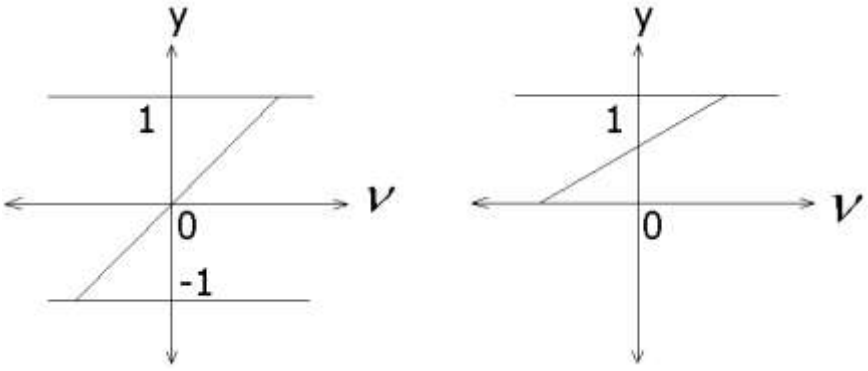
Фиг. 9. Рекурентниа мрежа

Активационни функции

Направените изследвания са показали, че невронните мрежи работят устойчиво само при ограничени активационни функции, точно както и биологичните им еквиваленти формират изходен сигнал след като сумарното стимулиращо въздействие достигне определена граница. Ако изходният сигнал се изменя в интервала [0 ; 1] те се наричат мрежи с еднополярни активационни функции, а изменението на стойностите е в интервала [−1 ; 1] - мрежи с биполярни активационни функции). [50]

Две са основно използваните активационни функции :

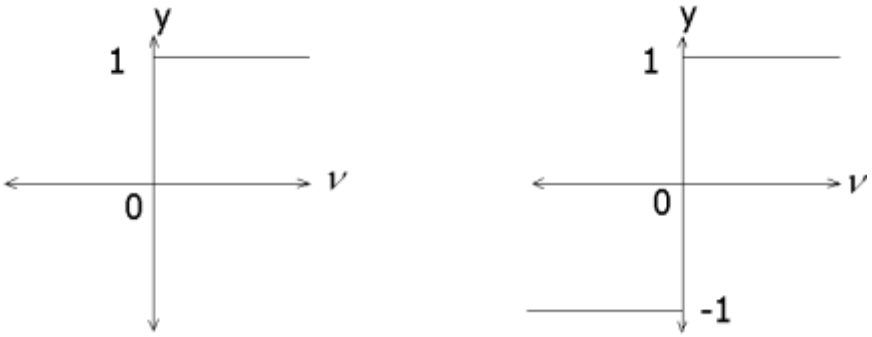
- *линейна функция* – еднополярна (фиг. 10a) и биполярна (фиг. 10b);



a) еднополярна *b) биполярна*

Фиг. 10. Линейна активационна функция

- *прагова фунцкия* - еднополярна (фиг. 11a) и биполярна (фиг. 11b). Праговата логическа функция (наричана още релейна, бинарна или сигнум функция) се използва за реализация на прости бинарни мрежови възли. Прилага се в еднополярен вариант (изходът има стойност 0 или 1) и двуполярен вариант (изходът приема стойност -1 или 1). Мрежите с бинарни неврони са най-лесни за хардуерна реализация, но често са много ограничени спрямо възможностите си за обучение [51];



a) еднополярна *b) биполярна*

Фиг. 11. Прагова фунцкия

- *сигмоидална и радиално базисна функция*. На фигура 12a) е представена еднополярна сигмоидална функция, а на фигура 12b) – биполярна сигмоидална функция. Те се задават с уравнението

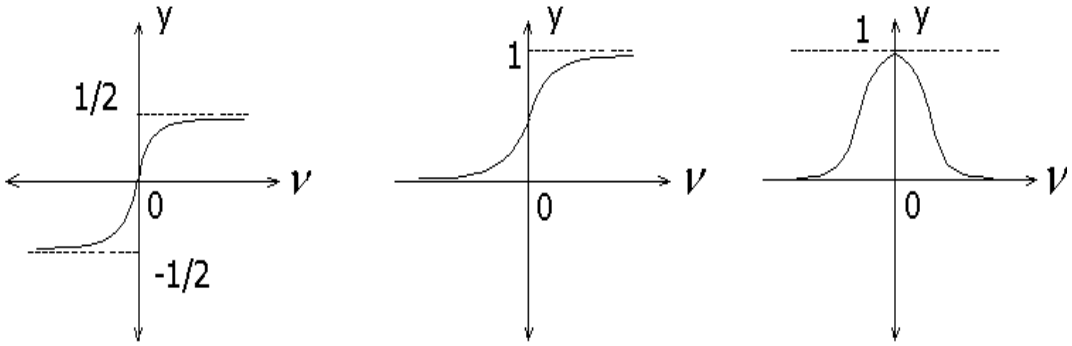
$$y = \frac{1}{1 + e^{-\alpha . x}},$$

където α е константа, чрез която регулира ширината на прехода, като обикновено интервала на функционални стойности е $(0\,;1)$, $(-1\,;1)$ или $\left(-\frac{1}{2}\,;\frac{1}{2}\right)$. Сигмоидални функциие се използват при а обучение с обратно разпространение на грешката.

Радиална базисна функция е представена на фигура 12c) и се задава с формулата:

$$y = e^{-\sum_{m=1}^{n}\frac{(x-c_m)^2}{2\sigma_m^2}},$$

където c_m задава центъра на активационната функция на m-тия неврон, σ_m - нейната изпъкналост. За разлика от останалите активационни функции, радиалната базисна функция придобива максималната си стойност единствено когато всички входни сигнали са със стойности близки до определената за тях централна стойност и е практически *0*, когато стойностите се различават съществено от зададената централна. Невронни мрежи с радиално базисни функции се използват успешно в задачите за идентификация на обекти и управление. [51]



a) еднополярна b) биполярна c) радиална базисна

Фиг. 12. Сигмоидална и радиална базисна функция

Обучение на невронни мрежи

Обучението на невронните врежи е процес, при който чрез целенасочено изменение на тегловните коефициенти на връзките, мрежите придобиват желаните свойства или поведение. Подходите за обучение могат да се обединят в три групи. Всяка от тях наподобява определен аспект от поведенческото обучение в биологичните системи.

1. Обучение с учител

При *обучение с учител (supervised learning)* началните стойности на теглата са зададени предварително и така се подава вектор с входни данни на входа на мрежата. Получената реакция на изхода се сравнява с предварително зададена и се пресмята грешката. Нейната стойност е критерий за промяна на тегловните коефициенти. Този процес продължава до минимизиране на грешката, затова обучението с учител е известно и под наименованието *обучение с минимизиране на грешката*. Прилагането на метода изисква задаване на серии от входни данни - образци и желани изходни сигнали, които се наричат тренировъчни комплекти. Чрез тях системата се обучава да реагира по желания начин. При мрежите, използващи обучение с учител се разграничават два етапа – етап на трениране, при който теглата се адаптират за да бъде научено желано поведение и на опериране, при който намерените на първия етап тегловни коефициенти са фиксирани и мрежата работи с тях при различни входни данни. Този подход за обучение широко се използва при невронни мрежи с прави връзки и някои рекурентни мрежи.

2. Обучение без учител

При *обучение без учител (unsupervised learning)* теглото на дадена връзка се определя в зависимост от корелацията между активациите на свързаните чрез нея два неврона, откъдето произлиза и другото название - *корелационно обучение*. Първият вариант на такова обучение е предложен от Хеб и извършва актуализация на стойността на всеки тегловен коефициент в мрежата чрез добавяне към текущата му стойност на нов член, пропорционален на произведението от стойностите на активация на двата свързани неврона. арактерно за това обучение е, че то не изисква тренировъчни комплекти и не съществува разграничение между етапите на обучение и опериране на мрежата.

Обучението без учител намира широко приложение при задачи, които изискват асоциативна памет.

3. Обучение с поощрение

При *обучение с поощрение (reinforcement learning)*, подобно на обучението с учител, трябва да има примери с образци от входни данни, но не е необходимо да се предоставят образци с желаните изходни реакции. Въвежда се обща мярка за адекватност на получените резултати, която води мрежата към желаното поведение. Тази мярка се нарича усилващ или поощряващ сигнал, който се подава като обратна връзка в мрежата, за да се поощрят коректните и съответно накажат некоректните поведения. Това става чрез увеличаване на онези тегла, които са допринесли за доброто поведение и намаляване предизвикалите лош резултат. [51]

Приложение на невронните мрежи

При алгоритмичния подход за решаване на различни задачи в компютърните науки се прави абстрактен модел на разглежданата задача, след като се разработва алгоритъм за решаването и, и накрая алгоритъмът се реализира с компютърна програма. Алгоритъмът е ефективен метод, който при даден списък от коректно дефинирани команди за изпълнение на задача и дадено начално състояние, преминава през коректно дефинирана поредица от последователни състояния и завършва в едно крайно състояние. Преходът между състоянията не е задължително да е детерминиран, тъй като някои алгоритми, известни като вероятностни алгоритми, съдържат елемент на случайност. В компютърните науки алгоритмите се изпозват за пресмятане, обработка на данни и други. [52]

За разлика от алгоритмичния подход, използването на невронни мрежи се базира на обучение чрез примери, с което се прави опит за моделиране на биологичния им еквивалент - мозъка. По принцип с невронни мрежи може да бъде изчислена всяка изчислима функция, т.е. с невронни мрежи може да се прави всичко, което се прави със стандартните цифрови компютри. За разлика от алгоритмичния подход обаче, невронните мрежи използват принципа на паралелна обработка на информацията и поради това те могат да бъдат използвани успешно при задачи за обработване на данни (сигнали, изображения и др.) в реално време, когато се цели постигане на по-голямо бързодействие.

Невронните мрежи работят на принципа на черната кутия, като се използва информация от разглеждания процес или явление, с цел обучаване при настройване на теглатата на мрежата. По тази причина те са удобно средство за моделиране на процеси и явления, за които може да бъде събрана много входно-изходна информация (от измерване и др.), но трудно могат да бъдат описани аналитично.

Невронните мрежи са ефективни при решаване задачи за класификация и апроксимация. При наличие на достатъчно данни и компютърни ресурси при обучението, почти всяко изображение между две векторни пространства може да бъде апроксимирано с желана точност посредством многослойна невронна мрежа с еднопосочно предаване на сигнала. Невронните мрежи намират широко приложение в почти всички области на науката и живота. Те се използват в:

- инженерната дейност за обработка на сигнали и изображения, за филтрация, при разпознаване на образи, при моделиране, идентификация и управление на сложни многосвързани обекти и др.;
- приложната математика за решаване на различни видове задачи като апроксимация, оптимизация с или без ограничения, решаване на различни видове уравнения и системи уравнения и др.;

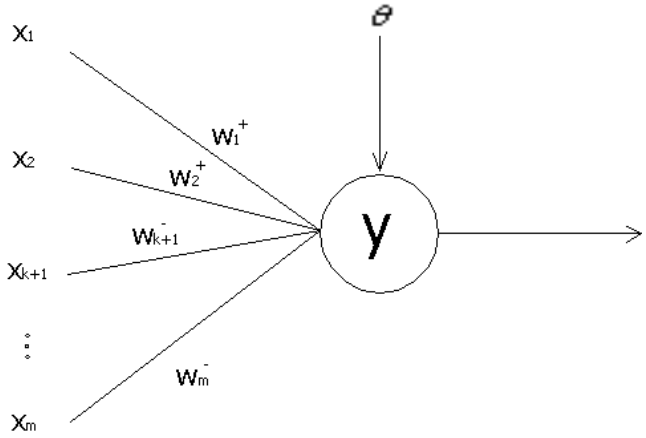
- статистиката за нелинейна регресия и класификация, прогнозиране на времеви редове и др.;
- когнитивните науки за описание и моделиране на мисленето и съзнанието;
- икономиката и финансите за изграждане на сложни икономически модели, за моделиране и прогнозиране на финансови пазари и много други. [50]

Видове невронни мрежи

Еднослойни невронни мрежи с прави връзки

1. *Неврон на Маккалък-Пит*

Невронът на Маккалък-Пит, предложен през 1943 г., е първото използване на невронна мрежа за изчисления (фиг. 13). Той се характеризира с изчислителна универсалност, тъй като всяка логическа функция може да бъда изчислена с мрежа от неврони на Маккалък-Пит. Същевременно всяка крайна последователност от дискретни действия може да бъде стимулирана с рекурентни невронни мрежи от такива неврони.



Фиг. 13. Неврон на Маккалък-Пит

При този вид невронна мрежа теглата са фиксирани, т.е. не се настройват с обучение. Входните сигнали $x_1, x_2, \dots x_m$ и изходните сигнали y на невроните са бинарни - $x_i \in \{0,1\}$, $i = 1, \dots, m$ и $y \in \{0,1\}$. Входните сигнали биват два вида – възбудителни и забранителни. Теглата, свързани със възбудителните входни сигнали $x_1, x_2, \dots x_k$ са положителни, а всеки ненулев входен сигнал на някой от забранителните входове $x_{k+1}, x_{k+2}, \dots x_m$ генерира нулев изходен сигнал. Изходен сигнал *1* се получава, когато сумата от теглата на възбудителните входни сигнали е по-голяма от предварително зададен праг θ и няма сигнал на никой от забранителните входове [50]:

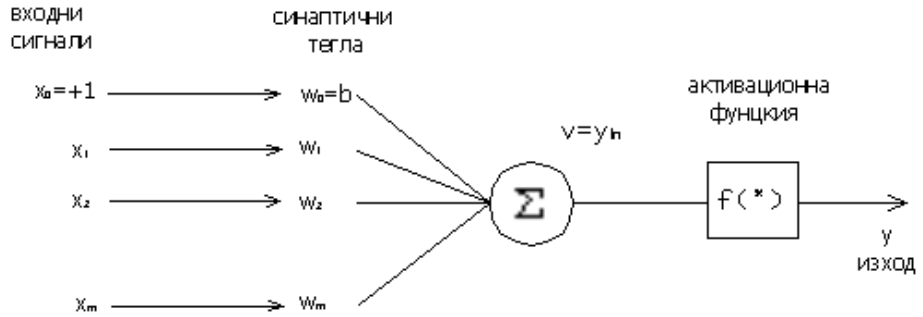
$$y = \begin{cases} y = 1, & \text{когато } \sum_{j=1}^{k} w_j^+ x_j \geq \theta \quad \text{или} \quad x_i = 0, \qquad i = k+1, \dots, m \\ 0, & \text{когато } \sum_{j=1}^{k} w_j^+ x_j < \theta \quad \text{или} \quad x_i = 1, \qquad i = k+1, \dots, m \end{cases}$$

За правилната работа на неврона трябва да е изпълнено условието:

$$\sum_{j=1}^{k} w_j^+ - w_j^- < \theta, \forall i = k+1, \dots m$$

2. Адаптивен линеен елемент

Адаптивният линеен елемент (*ADAptive LINear Element=ADALINE)* е предложен през 1962 г. от Бернард Уидроу. Той представлява единичен неврон с линейна активационна функция и настройващи се тегла (фиг.14). Алгоритъмът на Уидроу-Хоф за настройка на теглата използва метода на най-малките квадрати и е пример за използване на обучение при настройката. Той е в основа за различни видове адаптивна филтрация, използвана при обработка на сигнали.



Фиг. 14 Линейна активационна функция

За обучението на мрежите от тип *ADALINE* е необходимо да се разполага с обучаваща извадка с обем p, представляваща набори от входни сигнали и желан изход на мрежата за всяки от тях. При подаване на всеки *k*-ти комплект входни сигнали се определя изходът от невронната мрежа и грешката между него и еталонния изход при зададената комбинация на входа. Целта на обучението е теглата $w = [w_1\ w_2\ ...\ w_m]^T$да бъдат настроени така, че да бъде минимизирана грешката $\sum_{i=1}^{p} e_i$ за цялата обучаваща съвкупност. В различни варианти на обучаващите алгоритми промяната на теглата се прави след всеки компонент от обучаващата серия или след всяка епоха. Във всички случаи промяната на теглата е пропорционална на съответния входен сигнал, грешката и параметър η, наречен скорост на обучение.

Мрежите *ADALINE* принадлежат към общия клас алгоритми, наречени адаптивни линейни филтри. Те намират приложение в много области - за проектиране на изравняващи филтри при високоскоростни модеми, за адаптивни ехокомпенсатори при телефонни разговори на големи разстояния и сателитни комуникации, както и за предсказване на сигнали. *ADALINE* се прилагат и в медицината - за подтискане на шума от биене на сърцето на майката при ембрионална електрокардиография.

3. Перцептрон

Перцептронът е предложен през 1957 г. Розенблат. За разлика от адаптивния линеен елемент перцептронът е нелинеен неврон. Той използва релейна активационна функция (фиг. 15).

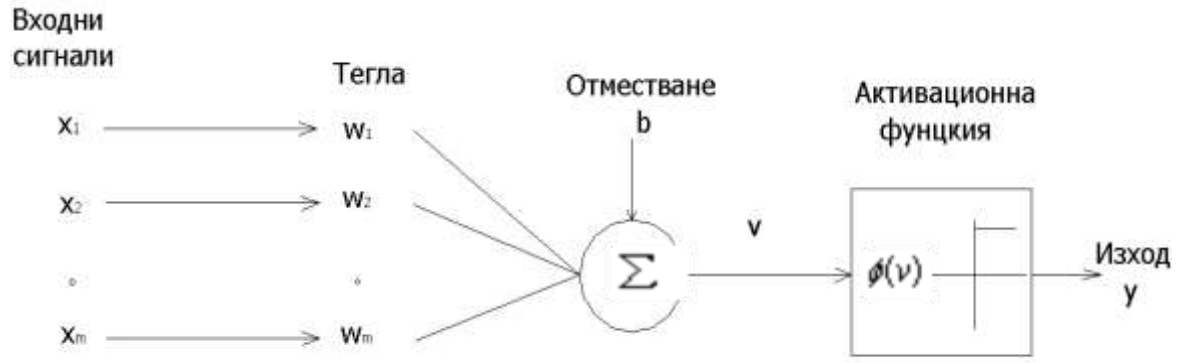
Изходът на суматора се изчислява като:

$$v = \sum_{j=1}^{m} w_j x_j + b = \sum_{j=0}^{m} w_j x_j$$

където $w_0 = b$, а x_0 е сигнал с постоянна стойност 1. Изходният сигнал на перцептрона се определя с формулата:

$$y = \phi(v) = \begin{cases} 1, & \text{ако } v \geq \theta \\ 0, & \text{ако } v < \theta \end{cases}$$

където θ е праг на активация на перцептрона.



Фиг. 15 Релейна активационна функция

При перцептрона се използва обучение с учител. При този тип обучение критерият за обучението се задава с комплект примерни входно-изходни последователности за желаната работа на мрежата от типа

$$\{p_1,\ d_1\},\{p_2,\ d_2\},\ldots,\{p_Q,\ d_Q\}$$

където p_k се подават на входа на мрежата, а d_k е съответната правилна стойност на изхода на мрежата. След сравняване на d_k с получения резултат се стартира алгоритъм за обучение, посредством който се настройват теглата и отместванията на мрежата така, че получаваната стойност на изхода на мрежата да се доближава до желаната.
Обучението се извтршва по следното правило:

- ако $d = 0$ и $\phi(w^T x) = 1$, то $w_{\_new} = w_{\_old} - x$,
- ако $d = 1$ и $\phi(w^T x) = 0$, то $w_{\_new} = w_{\_old} + x$,
- ако $d = \phi(w^T x) = 1$, то $w_{\_new} = w_{\_old}$,

където с $w_{\_new}$ се означава новата, а с $w_{\_old}$ - старата стойност на тегловия коефициент от преходната интерация; с x се означава съответната стойност на p_k, подадена на входа и с d съответното d_k Функцията $\phi(w^T x)$ приема в случая стойност *0* или *1*, т.е.

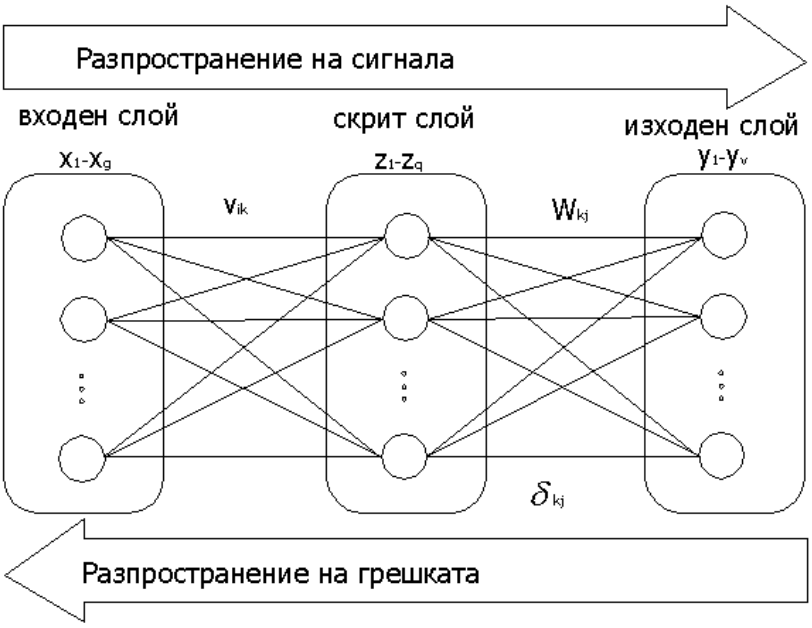
$$y = \phi(v) = \begin{cases} 1, & v \geq \theta \\ 0, & v < \theta \end{cases}.$$

Ако бъде въведена грешка $e = d - \phi(w^T x)$, правилото за обучение може да бъде представено в следня компактен вид: $w_{\_new} = w_{\_old} + e.x$.
Правилото може да се обобщи за един слой от *K* на брой перцептрони, използващи едни и същи входни сигнали $x = [x_1\ x_2\ \ldots\ x_m]^T$. В този случай формулите за настройка на вектора с теглата w_s и отместване b_s, свързани с s-тия перцептрон са равни на: $w_{s\ \_new} = w_{s\ \_old} + e_s.x,\ b_{s\ \_new} = b_{s\ \_old} + e_s$, където $e_s = d_s - \phi(w_s^T x + b_s)$.

Мнгослойни невронни мрежи с прави връзки

1. Многослойни невронни мрежи с обратно разпространение на грешката

На практика най-често използваните многослойни мрежи са двуслойни, защото е доказано, че при достатъчно неврони в междинния слой, всяка двуслойна мрежа може да реализира функциите на мрежа с повече слоеве. Архитектурата на двуслойна невронна мрежа с еднопосочно предаване на сигнала, използваща метода на обратно разпространение на грешката е показана на фигура 16.
Всички алгоритми за обучение на многослойните невронни мрежи се основават на алгоритъма с обратно разпространение на грешката, който е известен още като обобщено делта правило. В основата си той представлява градиентен метод, целящ минимизиране на грешката на изхода на мрежата.



Фиг. 16. Невронна мрежа с еднопосочно предаване на сигнала, използваща метода на обратно разпространение на грешката

Обучението на многослойните невронните мрежи преминава през три етапа.
По време на първия, наречен *право разпространение*, всеки входен неврон получава входен сигнал и го изпраща на всеки от невроните от скрития слой. Невроните в скрития слой формират своя изходен сигнал и го изпращат на невроните от изходния слой. В края на процеса изходните неврони изчисляват своите изходни сигнали и по този начин формират реакцията на мрежата на подаденото входно въздействие.
На следващия етап от обучаващия процес, реакцията на всеки изходен неврон се сравнява с желаната реакция с цел определяне на грешката. На нейна основа се определя коефициентът δ_k за всеки неврон в изходния слой. Коефициентът δ_k се използва, за да се разпредели грешката от изходния слой към предходния. По-късно тя се използва за преизчисляване на теглата между скрития и изходния слой. По подобен начин се изчислява и δ_j за скрития слой. Не е необходимо разпространението на грешката назад към входните неврони, но тя се използва за обновяване на стойността на тегловните коефициенти между входа и скрития слой.
В следващия трети етап, след като всички коефициенти δ бъдат определени, теглата на всички слоеве на мрежата се обновяват едновременно.

Активационната функция на мрежата с обратно разпространение на грешката трябва да има няколко важни свойства – да бъде непрекъсната, диференцуема и монотонно растяща. Друго условие е функцията да има крайни максимум и минимум, които да се достигат асимптотично. Най-често използваната функция, отговаряща на горните условия, е сигмоидалната.
Многослойните мрежи, използващи обучение с обратно разпространение на грешката имат няколко съществени предимства:

- могат да моделират всяка непрекъсната функция;
- нуждаят се само от подходяща обучаваща извадка;
- не изискват предварително познаване на проблемите и затова са приложими за решаване на лошо структурирани задачи;
- устойчиви са на шум и липса на данни от обучаващата извадка.

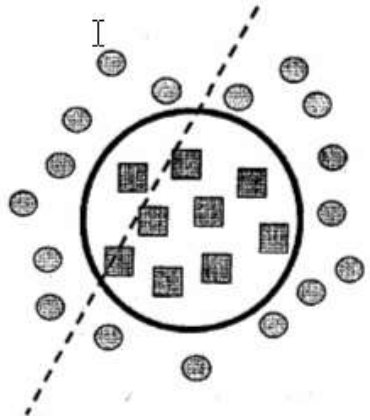
Наред с това този вид невронни мрежи притежават и някои недостатъци:

- обучението може да изисква голям обем ресурси;
- при обучението на мрежата е възможно достигане на локален минимум, който да е различен от глобалния минимум.

За преодоляване на последния проблем са предложени редица модификации на базовия обучаващ алгоритъм. Един ефективен подход е ново стартиране на обучението с променени начални тегла и/или разместване на обучаващите комплекти в извадката.

Многослойни невронни мрежи с радиални базисни функции

Идеята при въвеждане на тези мрежи е свързана с използването на линейни разделящи правила при класификация на образи. За двумерния пример от фигура 17, двата вида образи са разделими с нелинейна крива (показаната окръжност), но са линейно неразделими, т.е. не съществува права която да ги разделя.



Фиг. 17. Двумерен пример

Невронните мрежи, способни да се справят с този проблем, са мрежи с радиална базисна функция (фиг. 18). Те са двуслойни мрежи, чиито неврони в скрития слой имат нелинейна активационна фунцкия, а невроните в изходния слой – линейна активационна функция.

Трите основни параметъра, които влияят на качеството на мрежата с радиални базисни функции, са:

- центрове на радиално-базисните функции.
- радиусите на радиално-базисните функции.
- теглата на връзките в изходния слой.

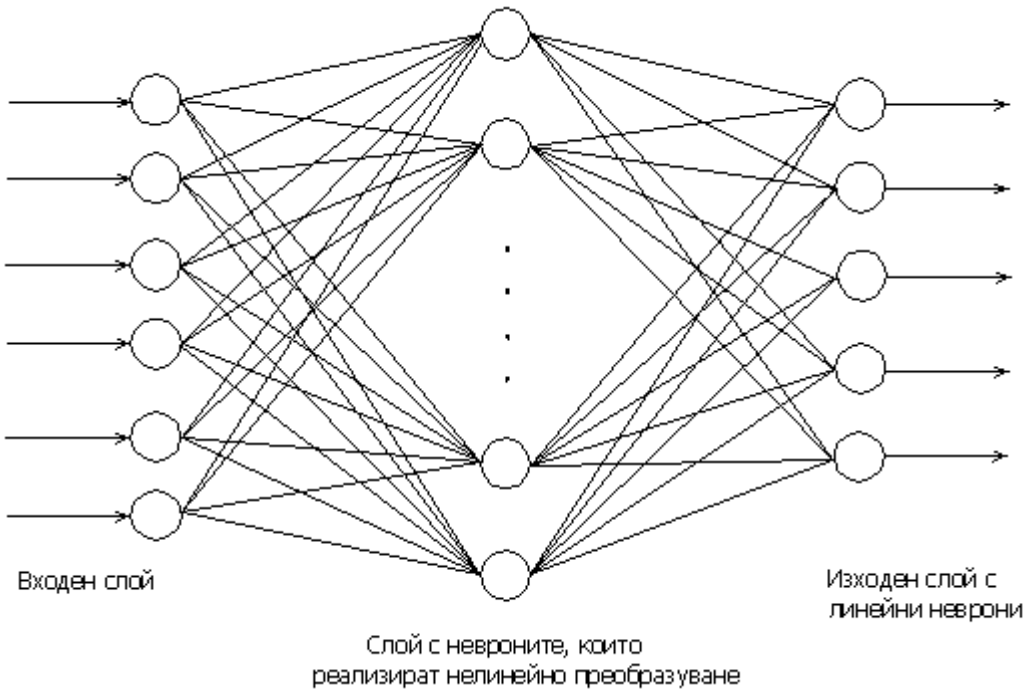
При обучение на тези мрежи обикновено центровете и радиусите на радиално-базисните функции се определят без учител, а за теглата се използва обучение с учител.

Радиусите на функциите се изчилсляват по формулата:

$$\sigma = \frac{d_{max}}{\sqrt{2m}},$$

където d_{max} е максималното разстояние между два неврона, а m е броят на невроните в скрития слой.

Предимствата на мрежите с радиално базисни функции пред стандартните мрежи, използващи алгоритъм с обратно разпространение на грешката, са в това, че те имат по-бързо обучение и по-сигурно достигане на глобаленя минимум на грешката. Недостатък на тези мрежи е, че много често изискват по-голям брой неврони в скития слой. [50]



Фиг. 18. Мрежа с радиални базисни функции

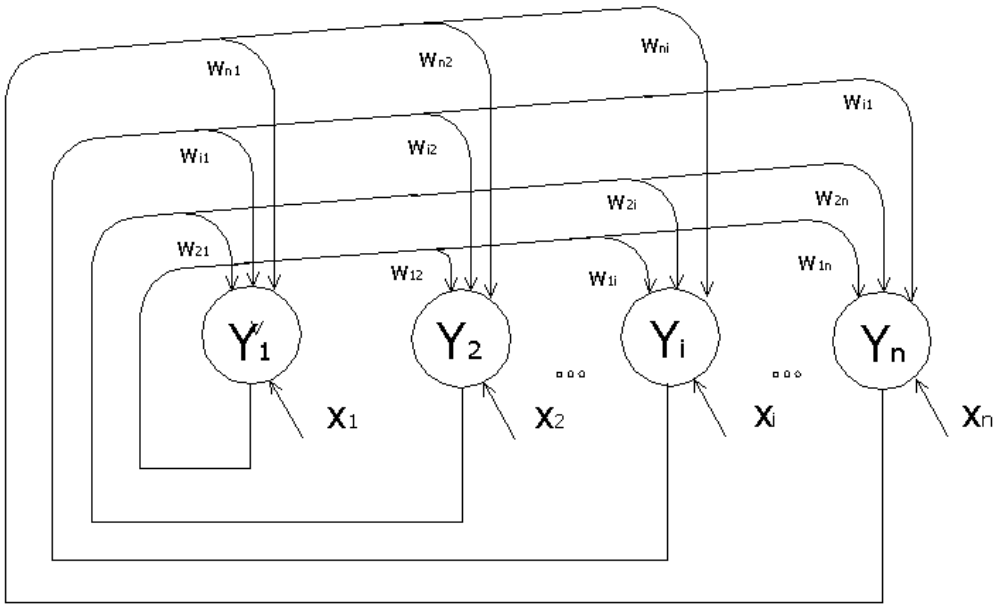
Рекурентни невронни мрежи

Невронните мрежи, в които информацията се пренася само в една посока – последователно от входния, през скритите слоеве (ако има такива) до изходния слой се наричат невронни мрежи с прави връзки. Съществуват мрежи, при които информация се движи и в обратна посока – към текущия или към предходни слоеве на мрежата. Тези мрежи се наричат *рекурентни мрежи*. Когато на една рекурентна мрежа се подаде входен вектор, не се получава директно изходен вектор след краен брой стъпки, а се предизвиква кръгов процес или резонанс на нервната активност в мрежата. В резултат на това едни и същи слоеве се активират многократно. Ако мрежата резонира известно време и след това се установи в устойчиво състояние, тя се нарича собственоустойчива. Целта на тренирането на рекурентните мрежи е да се намерят такива тегла на връзките, които да гарантират собствена устойчивост на мрежата при стабилизирани стойности на изхода, съответстващи с еталонни, зададени за обучаващите входни серии. [49]
Съществуват различни архитектури на рекурентни невронин мрежи. Едни от най-разпространените са предложените от Хопфийлд, Елман и Джордан.

1. *Невронни мрежи на Хопфийлд*

Невронните мрежи на Хопфийлд са еднослойни мрежи, за които всеки неврон е свързан с всички останали. Те биват два вида: дискретни и непрекъснати и могат да бъдат използвани не само като асоциативни памети, а също така за решаване на сложни комбинаторни задачи.
Архитектурата на дискретна мрежа на Хопфийлд е показана на фигура 19. Теглата в мрежата са съответно $w_{ij} = w_{ji}$ и $w_{ii} = 0$, т.е. матрицата с теглата на мрежата е симетрична с нулеви елементи по главния диагонал. Входните сигнали $x_1\ x_2 \ldots\ x_n$ не са задължителни. В някои случаи входни сигнали може да има, а в други случаи може да няма, като входове за невронната мрежа са началните състояния на изходните сигнали на невроните, а изходи са изходните сигнали на невроните след съответния брой

интерации. В тези случаи изходните сигнали на невроните, които определят състоянието на мрежата на всяка интерация, служат съответно за вход и изход на мрежата. Активационните функции на невроните са релейни, като в зависимост от данните (бинарни или биполярни) стойностите им са в $[0\,;1]$ или $[-1\,;1]$.



Фиг. 19. Невронни мрежи на Hopfield

Матрицата с теглата на мрежата се изчислсява като се използва правилото на Хеб, като диагоналните ѝ елементи се нулират.
При биполярни данни:

$$w_{ij} = \sum_{p} S_i(p)S_j(p), \qquad i \neq j$$
$$w_{ii} = 0$$

а при бинарни данни:

$$w_{ij} = \sum_{p} (2S_i(p) - 1)(S_i(p) - 1), \qquad i \neq j$$
$$w_{ii} = 0$$

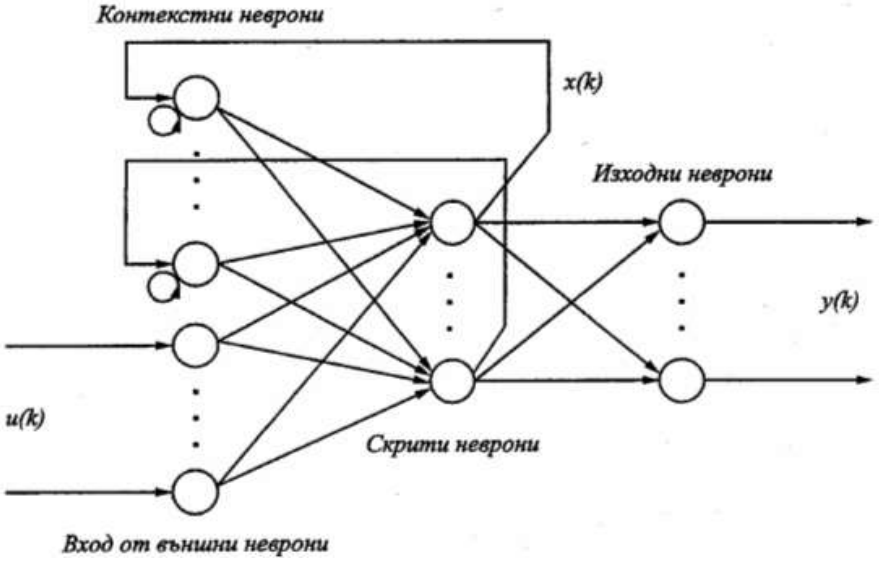
като в случая при конструиране на матрицата с теглата всеки бинарен вектор се преобразува в биполярен.
Най-важният момент при реализацията на интерациите е асинхронната промяна на изходните сигнали на невроните. Това означава, че на всяка стъпка се променя изходният сигнал само за един неврон, като този неврон се избира случайно. Вероятността за това произволен невроните да бъде избран на дадена стъпка е равна. Получената нова стойност на изхода на неврона се използва при промяната на изходния сигнал за следващите неврони. Този подход гарантира сходимостта на промяната на изходните сигнали на невроните на мрежата. [50]

2. Невронни мрежи на Елман

На фигура 20 е показана архитектурата на мрежата на Елман. Тя е трислойна. Първият слой се състои от две различни групи неврони – външни входни неврони и вътрешни входни неврони, наричани още контекстни възли. На входовете на контекстните възли постъпват изходите от неврони, разположени в скрития слой. Вътрешните входни

неврони се наричат и неврони на паметта, тъй като съхраняват стойността на изхода на неврон от скрития слой в предходен времеви такт. Вторият слой е скрит слой. На входовете на неговите неврони постъпват изходите на външните и вътрешните входни неврони.

Теглата на рекурентните връзки не се променят, като в общия случай се приемат за равни на *1*. На настройване подлежат само теглата на правите връзки. При начална инициализация на мрежата на тях се присвояват произволни малки начални стойности, а на изходите от контекстните неврони – стойност *0,5* при използване на сигмоидна функция на активация.

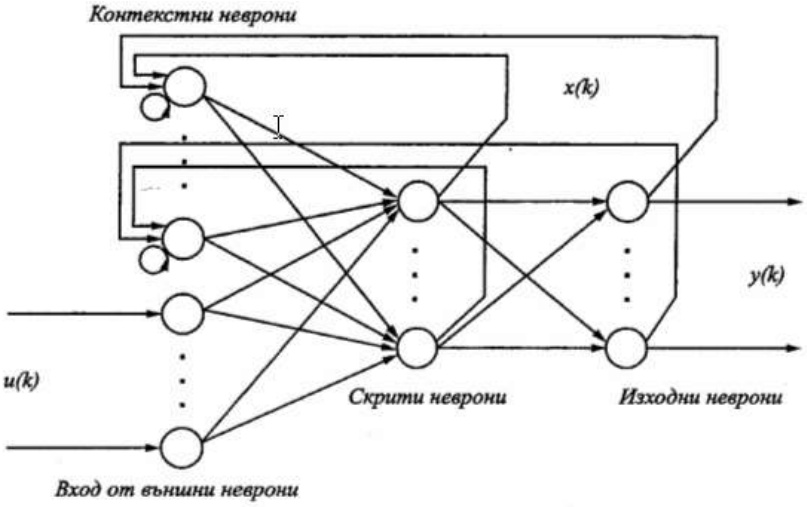


Фиг. 20 Невронни мрежи на Елман

3. Невронни мрежи на Джордан

Мрежата на Джордан (фиг. 21.) също е трислойна и с автообратни връзки при контекстните неврони, но обратните връзки са от изходния към контекстния слой. Наличието на обратни връзки и от скрития слой, подобно на мрежата на Елман позволяват чрез нея да се си моделират произволни динамични системи.

Мрежите на Джордан могат да бъдат тренирани чрез стандартния алгоритъм с обратно разпространение на грешката, като стойностите на теглата на обратните връзки са фиксирани. [51]



Фиг. 21. Невронни мрежи на Джордан

V. Заключение

Развитието на технологиите за гласово разпознаване започва през 60-те години на XX век с малък набор единични думи (10-100), които биват разпознавани на базата на акустично-фонетичните свойства на речевите звуци. Ключовите технологии, развили се през този ранен период са филтър анализи, методи за нормализация по време и е сожено началото на сложните методологии за динамично програмиране. През 70-те години се осъществява разпознаването на средно голяма лексика (100-1000 думи), като се използват шаблонно базирани методи за разпознаване на мостри. Основните технологии през този период са модели за разпознаване на шаблони, въведението на LPC методите за спектрално представяне, методи за създаване на клъстер от шаблони за независими от говорещия разпознаватели, и въведението на методите за динамично програмиране за решаване на свързани задачи за разпознаване на думи. През 80-те години се работи със системи за гласово разпознаване, използващи голяма лексика (>1000 думи), основавайки се на статистически методи, с голям набор от мрежи за справяне с езиковите структури. Ключовите технологии през този период са скритите модели на Марков и стохастичния езиков модел, които заедно са в основата на мощни нови методи за ефективно и високопроизводително решение на всяка задача, свързана с непрекъснато речево разпознаване. През 90-те години се въвеждат системи за лексика с неограничени езикови модели и ограничените по задачи синтактични модели за непрекъснато речево разпознаване и разбиране. Ключовите технологии по това време са методите за стохастично разбиране на езика, статистическо обучение на акустични и езикови модели, и въведението в трансформаторните системи от крайни състояния , като и методите за тяхното детерминиране и минимизиране за ефикасно внедряване в големите системи за разпознаване на говор.

В следващите десетилетия се въвеждат огромни от лексикална гледна точка системи с пълни семантични модели, интегрирани с речево-говорни синтезирани системи и много-модални входове (клавиатури, мишки и т.н.). Тези системи осъществяват речеви диалог с набор от входни и изходни модалности за лесна употреба и гъвкавост при работа с разнообразни среди, където речта може да не е най-подходящото средство тъй както други входно-изходни модалности. По време на този период се зараждат високоефективни естествени конкативни системи за синтезиране на речта, употребата на машинното обучение за подобряването на разбирането на реч и диалога, и се въвеждат смесено-инициативните диалогови системи, които предоставят потребителски контрол при необходимост.

Основно предизвикателството все още е задачата за проектиране на машина, която реално да функционира като интелигентно човешко същество. Постиженията до днешна дата са само началото и ще са нужни още много изследвания, докато при машините бъде постигната производителност в реално време, която съперничи на тази на човека.

References

[1] Davis K., R. Biddulph, S. Balashek, "Automatic Recognition of Spoken Digits," *The Journal of the Acoustical Society of America,* vol. 24, no. 6, pp. 627-642, 1952.

[2] Olson H., H. Belar, “Phonetic Typewriter,” *The Journal of the Acoustical Society of America,* vol. 28, no. 6, pp. 1072-1081, 1956.

[3] Forgie J., C. Forgie, “Results Obtained from a Vowel Recognition Computer,” *The Journal of the Acoustical Society of America,* vol. 31, no. 11, pp. 1480-1489, 1959.

[4] Suzuki J., K. Nakata , “Recognition of Japanese Vowels—Preliminary to the Recognition of Speech,” *J. Radio Res. Lab,* vol. 37, no. 8, pp. 193-212, 1961.

[5] Sakai T., S. Doshita, “The Phonetic Typewriter,” *The Journal of the Acoustical Society of America,* vol. 33, no. 11, 1961.

[6] Nagata K., Y. Kato, S. Chiba, „Spoken Digit Recognizer for Japanese Language,“ *NEC Res. Develop,* № 6, 1963.

[7] Denes P., “The Design and Operation of the Mechanical Speech Recognizer at University College London,” *British Institution of Radio Engineers,* vol. 19, no. 4, pp. 211-229, 1959.

[8] Martin T., A. Nelson, X. Zadell, “Speech Recognition by Feature Abstraction,” Tech. Report AL-TDR-64-176, Air Force Avionics Lab, 1964.

[9] Vintsyuk T., “Speech Discrimination by Dynamic Programming,” *Kibernetika,* vol. 4, no. 2, pp. 81-88, 1968.

[10] Sakoe H., S. Chiba, “Dynamic Programming Algorithm Quantization for Spoken Word,” *Speech and Signal Proc.,* vol. 26, no. 1, pp. 43-49, 1978.

[11] Viterbi A., “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm,” *IEEE Trans. Informaiton Theory,* vol. 13, pp. 260-269, 1967.

[12] Atal B., S. Hanauer, “Speech Analysis and Synthesis by Linear Prediction of the Speech Wave,” *J. Acoust. Soc. Am.,* vol. 50, no. 2, pp. 637-655, 1971.

[13] Itakura F., S. Saito, “A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies,” *Electronics and Communications in Japan,* vol. 53, pp. 36-43, 1970.

[14] Itajura F., “Minimum Prediction Residual Principle Applied to Speech Recognition,” *IEEE Trans. Acoustics, Speech and Signal Proc,* vol. 23, pp. 57-72, 1975.

[15] Rabiner L., S. Levinson, A. Rosenberg, J. Wilpon, “Speaker Independent Recognition of Isolated Words Using Clustering Techniques,” *IEEE Trans. Acoustics, Speech and Signal Proc.,* vol. 27, pp. 336-349, 1979.

[16] Lowerre B., “The HARPY Speech Understanding System,” *Trends in Speech Recognition, Speech Science Publications, 1986, reprinted in Readings in Speech Recognition,* pp. 576-586, 1990.

[17] Mohri M., “Finite-State Transducers in Language and Speech Processing,” *Computational Linguistics,* vol. 23, no. 2, pp. 269-312, 1997.

[18] Klatt D., “Review of the DARPA Speech Understanding Project (1),” *J. Acoust. Soc. Am.,* vol. 62, pp. 1345-1366, 1977.

[19] Jelinek F., R. Bahl, R. Mercer, “Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech,” *IEEE Trans. On Information Theory,* vol. 21, pp. 250-256, 1975.

[20] Shannon C., “A Mathematical Theory of Communication,” *Bell System Technical Journal,* vol. 27, pp. 379-423, 623-656, 1948.

[21] Juang B., S. Levinson, M. Sondhi, “Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains,” *IEEE Trans. Information Theory,* vol. 32, no. 2, pp. 307-309, 1986.

[22] Juang B., “Maximum Likelihood Estimation for Mixture Multivariate Stochastic Observations of Markov Chains,” *AT&T Tech. J,* vol. 64, no. 6, pp. 1235-1249, 1985.

[23] Lee C., L. Rabiner, R. Pieraccini, J. Wilpon, “Acoustic modeling for large vocabulary speech recognition,” *Computer Speech & Language,* pp. 1237-1265, 1990.

[24] Wilpon J., L. Rabiner, C. Lee , E. Goldman, “Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models,” *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, no. 11, 1990.

[25] Ferguson J., “Hidden Markov Analysis: An Introduction in Hidden Markov Models for Speech,” Institute for Defense Analyses, Princeton, 1980.

[26] Levinson S., L. Rabiner, M. Sondhi, “An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition,” *Bell Syst. Tech. J.*, vol. 62, no. 4, pp. 1035-1074, 1983.

[27] Rabiner L., B. Juang, “Statistical Methods for the Recognition and Understanding of Speech,” Encyclopedia of Language and Linguistics, 2004.

[28] Baum L., “An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes,” *Inequalities*, vol. 3, pp. 1-8, 1972.

[29] Theodoridis S., K. Koutroumbas, “Pattern Recognition: Second Edition,” Elsevier Academic Press, 2003.

[30] Baum L., “An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes,” *Inequalities*, vol. 3, pp. 1-8, 1972.

[31] Poritz A., “Linear Predictive Hidden Markov Models and the Speech Signal,” in Proc. ICASSP-82, Paris, 1982.

[32] Liporace L., “Maximum Likelihood Estimation for Multivariate Observations of Markov Sources,” *IEEE Trans. on Information Theory*, vol. 28, no. 5, pp. 729-734, 1982.

[33] Juang B. , S. Levinson, M. Sondhi, “Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains,” *IEEE Trans. Information Theory*, vol. 32, no. 2, pp. 307-309, 1986.

[34] Juang B., “Maximum Likelihood Estimation for Mixture Multivariate Stochastic Observations of Markov Chains,” *AT&T Tech. J.*, vol. 64, no. 6, pp. 1235-1249, 1985.

[35] Mohri M., “Finite-State Transducers in Language and Speech Processing,” *Computational Linguistics*, vol. 23, no. 2, pp. 269-312, 1992.

[36] McCullough W., W. Pitts, “A Logical Calculus of Ideas Immanent in Nervous Activity,” *Bull. Math Biophysics*, vol. 5, pp. 115-133, 1943.

[37] Lippmann R., Review of Neural Networks for Speech Recognition, Readings in Speech Recognition, 1990.

[38] Juang B., C. Lee, W. Chou, “Minimum Classification Error Rate Methods for Speech Recognition,” *IEEE Trans. Speech & Audio Processing, T-SA*, vol. 5, no. 3, pp. 257-265, 1997.

[39] Vapnik V., Statistical Learning Theory, John Wiley and Sons, 1998.

[40] Lee K., Large-vocabulary Speaker-independent Continuous Speech Recognition: The Sphinx System, Ph.D. Thesis, Carnegie Mellon University, 1988.

[41] Schwartz R., C. Barry , Y. Chow, etc., „The BBN BYBLOS Continuous Speech Recognition System,“ in Proc. of the Speech and Natural Language Workshop, Philadelphia, 1989.

[42] Murveit H., M. Cohen , P. Price , etc., „SRI's DECIPHER System,“ in proceedings of the Speech and Natural Language Workshop, 1989, Philadelphia.

[43] Young S., „the HTKBook,“ http://htk.eng.cam.ac.uk/.

[44] Glass J., E. Weinstein, „SpeechBuilder: Facilitating Spoken Dialogue System Development,“ 7th European Conf. on Speech Communication and Technology, Aalborg Denmark, 2001.

[45] Zue V., “Jupiter: A Telephone-Based Conversational Interface for Weather Information,” *IEEE Trans. On Speech and Audio Processing,* vol. X, pp. 100-112, 2000.

[46] Gorin A., B. Parker, R. Sachs, J. Wilpon, “How May I Help You?,” 1996.

[47] Kishorjit N., R. Vidya, Y. Nirmal, B. Sivaji, “Manipuri Morpheme Identification,” in *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing,* Mumbai, 2012.

[48] Huang X., A. Acero, H. Hon, Spoken Language processing – A Guide to Theory, Algorithms and System Development, Prentice Hall PTR, 2001, pp. 375-407.

[49] Потемкин В., В. Медведев, Нейронные сети. MATLAB 6, Диалог-МИФИ, 2002.

[50] Младенов В., С. Йорданова, Размито управление и невронни мрежи, София, 2006.

[51] Тренчев И., П. Миланов, Н. Пенчева, И. Мирчев, Невронни мрежи, Благоевград: ЮЗУ "Неофит Рилски", 2010.

[52] Георгиева П., Генетични размити ситеми, Бургас: Полиграф, 2016.