

Marjaana Lehtonen

**NEUROVERKKOJEN KIELITEKNOLOGISISTA
SOVELLUKSISTA ERITYISESTI PUHETEKNOLOGIAN
ALUEELLA**

Tietojärjestelmätieteen
kandidaatintutkielma

25.4.2003

Jyväskylän yliopisto
Tietojenkäsittelytieteiden laitos
Jyväskylä

Tiivistelmä

Lehtonen Tiina Marjaana
Neuroverkkojen käyttö puheentunnistuksessa
Jyväskylä, Jyväskylän Yliopisto, 2003
XxxSIVUMÄÄRÄ (johdanto-yhteenveto)
Kandidaatintutkielma

Tässä tutkielmassa tarkastellaan neuroverkkoja ja erityisesti niiden soveltamista puheentunnistukseen.

AVAINSANAT: kieliteknologia, neuroverkko, puheentunnistus

Sisällysluettelo

1. Johdanto.....	1
1.1 Kieliteknologian ja puheentunnistuksen suhteesta.....	1
1.2 Kieliteknologian ja verkkoteorian suhteesta.....	2
2. Puheentunnistus	3
2.1 Puheentunnistuksen peruseriaate.....	3
2.2 Puheentutkimuksen nykytila ja tavoitteet.....	4
2.3 Puheentunnistuksen haasteet.....	4
2. Neuroverkot	5
3.1 Neuroverkon yleinen rakenne ja toimintaperiaate	5
3.2 SOM-kartat	6
3.3 MLP-verkot.....	7
3.4 Neuroverkkojen käyttöön liittyviä epävarmuustekijöitä	8
3. Neuroverkkojen soveltaminen puheentunnistuksessa.....	9
3.2 HMM-mallien käyttö puheentunnistukseen tarkoitetuissa neuroverkoissa.....	9
4. Yhteenveto.....	12
Lähteet.....	13

1. Johdanto

Tämän tutkielman tarkoituksena on kuvata neuroverkkojen käyttöä modernissa kieliteknologiassa, erityisesti puheentunnistuksessa. Tutkielmassa selvitetään neuroverkkojen yleinen perusrakenne ja toimintaperiaate sekä tarkastellaan lähemmin kahta neuroverkkotyyppiä, joita on sovellettu laajasti puheentunnistuksessa ja myös joillakin muilla kieliteknologian osa-alueilla. Luvussa 2 kuvataan puheentunnistuksen peruseriaate ja tutkimusmotiivit sekä luonnehditaan saavutettuja ja tulevaisuudessa tavoiteltavia tuloksia ja tutkimukseen liittyviä haasteita. Luvussa 3 kuvataan neuroverkkoja yleisellä tasolla sekä käsitellään esimerkkeinä SOM-karttoja ja MLP-verkkoja. Tämän lisäksi kartoitetaan neuroverkkojen käyttöön liittyviä epävarmuustekijöitä. Luvussa 4 kuvataan neuroverkkolähestymistapaan pohjautuvaa puheentunnistusta ja esitellään menetelmiä, joilla neuroverkkoja voidaan sovittaa vastaamaan puheentunnistuksen tarpeita. Tutkielman päättää luvussa 5 esitetty yhteenveto, jossa kootaan edellisten lukujen ydinasiat.

Tutkielma pohjautuu kirjallisuuteen.

1.1 Kieliteknologian ja puheentunnistuksen suhteesta

Kieliteknologia (engl. *language technology*) on useita tutkimusaloja käsittävä kattotermi, josta on esitetty hieman toisistaan poikkeavia tulkintoja. Kieliteknologia on kielitieteen ja muiden tieteiden sulautumisesta syntynyt verrattain tuore tieteenala, eikä sitä voi selkeästi rajata perinteisiin tieteisiin. Erityisesti aikaisemmin synonyymisesti kieliteknologia-termin kanssa on käytetty myös termiä tietokone-lingvistiikka (engl. *computational linguistics*). Luonnollisten kielten käsittelyllä (engl. *natural language processing*), joka on osa kieliteknologiaa, tarkoitetaan tekniikoita ja teorioita, jotka liittyvät tietokoneiden ja -järjestelmien sekä luonnollisten kielten, siis ihmisten välisessä vuorovaikutuksessa käytettävien kielten, yhteensovittamiseen. Tavallisimmin termillä luonnollisten kielten käsittely viitataan erityisesti kirjoitetun kielen käsittelyyn, siis käytännössä tiedostosta luettavaan tai näppäimistön kautta syötettyyn tekstiin. Perinteisiä osa-alueita ovat lauseenjäsennys sekä tiedon varastointiin ja tallennukseen liittyvät teorit ja sovellukset. Kunnianhimoisimpia tavoitteita ovat luonnolliseen kieleen pohjautuvat käyttöliittymät sekä tekstin merkityksen ja vuorovaikutuksen huomioon ottavat teorit ja sovellukset. (Ks. esim. Lehnert & kumpp. 1982).

Puheentunnistus on merkittävä kieliteknologian osa-alue, mutta sitä ei tyypillisesti lueta luonnollisten kielten käsittelyn alalajiksi. Karkeasti jaotellen voidaankin ajatella, että puhe muunnetaan puheentunnistusmenetelmillä tekstimuotoon, ja vasta tämä tekstimuotoon muunnettu kielidata toimii syöteenä varsinaisille luonnollisen kielen käsittelysovelluksille. (ks esim. Lehnert & kumpp. 1982).

Seuraava osa-aluejako noudattaa Helsingin yliopiston kieliteknologiaopetuksessa käytettäviä, professori Koskenniemen muotoilemia linjauksia. Tämän mallin mukaan kieliteknologia voidaan jakaa kahdeksaan pääosa-alueeseen:

1. tiedon haku ja hallinta
2. puheteknologia
3. luonnollisen kielen käyttöliittymä

4. neuraalilaskenta
5. tietokoneavusteinen kielenoppiminen
6. kielenkääntäjän apuvälineet
7. kirjoittajan apuvälineet
8. kielen kuvaamisen mallit.

Edellä lueteltujen osa-alueiden rajat ovat liukuvia, ja erityisesti menetelmätasolla tietyn tutkimusalueen piirissä kehitettyjä menetelmiä voidaan usein yleistää muihin alueisiin soveltuviksi. Tässä tutkielmassa painopiste on puheentunnistuksessa sovellettavissa neuroverkoissa. Kieliteknologian osa-aluejakoa ajatellen tutkielman aihe koskee varsinaisen puheteknologian ohella myös neuraalilaskentaa (tarkasteltavien menetelmien vuoksi), luonnollisen kielen käyttöliittymää (koska merkittävä osa puheentunnistustutkimuksesta keskittyy ääniohjattujen järjestelmien kehittämiseen) ja kielen kuvaamisen malleihin (menetelmät hyödyntävät kuvausmalleja).

1.2 Kieliteknologian ja verkkoteorian suhteesta

Perinteisen kielitieteen ja verkkoteorian liitto ei ole lainkaan uusi ilmiö, esimerkiksi Ziererin vuonna 1970 julkaistussa teoksessa *The Theory of Graphs in Linguistics* esitellään useita tapoja, joilla kielitieteellisessä tutkimuksessa voidaan hyödyntää verkkoteorian menetelmiä.

Kielen mallintamisessa on käytetty jo 60-luvulta alkaen paljon erilaisia verkkomalleja, varsinkin automaatteja. Pohjoismaissa on perinteisesti painotettu konekäännöksen tutkimusta ja korpusanalyysimenetelmiä. 1970-luvulta alkaen painopiste on ollut teoreettisessa tietokonelingvistiikassa. Näissä kaikissa on verkkopohjaisilla menetelmillä ollut vankka jalansija. 80-luvulla Kimmo Koskenniemi julkaisi TWOL-kaksitasomorfologian, joka on kieliriippumaton tietokonepohjainen morfologiasovellus. Koskenniemen malli on sovellus äärellisestä automaatista. Mallia on sovellettu suomen kielen lisäksi useisiin muihinkin kieliin. (Hovdhaugen & kumpp. 2000: 537–540).

Modernin kieliteknologian ja verkkoteorian välinen suhde määritellään eri yhteyksissä eri tavoin, esimerkiksi johdannon alussa esitetyn Koskenniemen luokittelun mukaan neuraalilaskenta katsotaan jo sinällään kieliteknologian osa-alueeksi.

2. Puheentunnistus

Johdannossa esitetyn jaottelun mukaisesti puheteknologia on yksi kieliteknologian osa-alueista. Puheteknologia on tieteenala, joka keskittyy puheenkäsittelyyn ja siinä käytettyjen menetelmien kehittämiseen. Puheentunnistus on merkittävä osa puheteknologista tutkimuskenttää. Seuraavat puheenkäsittelyn ja puheentunnistuksen määritelmät pohjautuvat Jaakohuhtan (2001: 500) esittämiin määritelmiin. Tietoteknisessä kontekstissa puheenkäsittelyllä tarkoitetaan pääosin joko puheen tunnistamista tai puhesignaalin tuottamista. Puheen tunnistaminen on puheen muuttamista sanoiksi ja lauseiksi. Puhesignaalin tuottaminen on päinvastainen prosessi, jossa kirjallisesta esitysmuodosta muunnetaan puhesignaali. Puheentunnistus on puheenkäsittelyn menetelmä, jossa puhuttu viesti muunnetaan tietoteknisin keinoin joko kirjoitetuksi tekstiksi tai komennoiksi, joita järjestelmä ymmärtää. Jälkimmäinen voidaan nähdä myös osana luonnollisen kielen käyttöliittymien tutkimusalaa.

Tässä luvussa kuvataan puheentunnistuksen peruseriaate ja tutkimusmotiivit sekä luonnehditaan saavutettuja ja tulevaisuudessa tavoiteltavia tuloksia ja tutkimukseen liittyviä haasteita.

2.1 Puheentunnistuksen peruseriaate

Puheentunnistuksen peruseriaa on, että järjestelmä tulkaa äänimuotoisen syötteen luonnollisen kielen ilmaukseksi, käytännössä kirjaimiksi. Mahdollisten syöteeänteiden määrä on rajaton, mutta vain tietyt äänteet tai äänneyhdistelmät ovat laillisia. Järjestelmän on joko hylättävä syötesignaali tai tulkittava se joksikin 'lailliseksi' merkiksi tai merkin osaksi. Puheentunnistus voidaan jakaa kolmeksi prosessiksi, jotka ovat puhesignaalin muuntaminen digitaaliseen muotoon, signaalin ominaisuuksien analysointi sekä signaalin tunnistaminen vertailemalla ominaisuuksia. Vertailussa ja etsinnässä käytetään malleja kielen ominaisuuksista ja puheen yksiköistä. (Salmela 2000a.)

Puheentunnistuksessa syötteenä olevasta puhesignaalista erotetaan kontekstisidonnaiset äänteet (allofonit). Puhesignaali muunnetaan äänteiksi, käyttämällä välimuotona malleja, jotka muodostetaan signaalin ominaisuuksista perusteella. Koska puhesignaali sisältää suuria vaihteluita, se kannattaa yleensä hyödyllistä normalisoida, jolloin malliin voidaan sisällyttää olennaisimmat ominaisuudet. (Salmela 2000a.)

Puheentunnistusjärjestelmää varten puhesignaali siis digitalisoidaan. Tämä digitalisoitu signaali ositetaan peräkkäisiksi jaksoiksi, joiden kesto on tyypillisesti 10 tai 20 millisekuntia. Kun puhesignaali on jaksotettu, kustakin jaksosta seulotaan tunnistuksen kannalta olennaiset elementit. Käytännössä tämä tarkoittaa, että järjestelmä pyrkii erottelemaan vaihtelu- ja häiriöelementit ja mahdollisuuksien mukaan poistamaan ne. Signaalista olisi erotettavissa rajattomasti ominaisuuksia, mutta tunnistuksessa käytetään yleensä noin 20 ominaisuuden perusteella muodostettu ominaisuusvektoria. (Salmela 2000a.)

Ominaisuusvektorin avulla tunnistettavaa jaksoa voidaan verrata äänne- ja sanastomalleihin ja hakea niiden perusteella todennäköisintä tulkintaa. Tässä vaiheessa hyödynnetään erittäin yleisesti HMM-malleja. (Salmela 2000a.)

2.2 Puheentunnistuksen nykytila ja tavoitteet

Puheentunnistuksen arvellaan olevan lähitulevaisuuden merkittävimpiä osa-alueita kieliteknologiassa. Salmelan (1997: 1) mukaan puheentunnistuksen keskeinen tavoite on laajentaa mahdollisuuksia käyttää erilaisia järjestelmiä äänikomennoilla. Puheentunnistuksen merkittävyys johtuukin ensisijaisesti matkapuhelinten ja muiden mobiililaitteiden yleistymisestä ja pyrkimyksistä kehittää näihin laitteisiin ääniohjattuja sovelluksia. Toinen merkittävä alue on pyrkimys kehittää äänipohjaisia toimistosovelluksia, esimerkiksi Wordin puheella toimiva vastine.

Puheentunnistuksen tutkimus on tällä hetkellä erittäin vilkasta. Erityisesti näytetään panostavan erillään kehitettyjen mallien pohjalta muodostettaviin hybridimalleihin, joissa on tavoitteena osittaa puheentunnistuksen vaiheet ja valjastaa paras menetelmä kuhunkin vaiheeseen.

Nykyiset puheentunnistussovellukset edellyttävät tavallisesti, että syötteitä rajataan erittäin voimakkaasti. Rajoitukset koskevat tavallisesti joko äännetasoa tai sanastotasoa. Äännetasolla syöteavaruutta voidaan kaventaa rajoittamalla käyttäjien määrää (yhdele käyttäjälle tarkoitetut sovellukset). Tämän lisäksi voidaan edellyttää, että käyttäjä ääntää sanat yksittäin. Sanastotasolla syötteitä voidaan rajata määrittämällä sovellus koskemaan vain tiettyä erityisalaa, esimerkiksi suomen kielen lukusanoja välillä 0–9. (Salmela 1997.)

Ensimmäiset puheentunnistussovellukset kehitettiin 1950-luvulla, mutta kaupallisesti toimivaa tasoa ei varsinaisesti ollu saavutettu (Salmela 2000b: 72). Kaupallisissa tuotteissa on kuitenkin jo joitakin puheentunnistusominaisuuksia, esimerkiksi markkinoilla olevissa matkapuhelimissa on äänikomentopohjaisia hakemisto- ja numerovalintatoimintoja.

Tulevaisuuden tavoitteissa on kaksi päälinjaa: mobiililaitteiden käytön helpottaminen äänikomentopohjaisilla käyttöliittymillä ja puheen tekstimuotoon muuttamisen automatisointi. Matkapuhelimien taloudellisen merkityksen vuoksi puheentunnistukseen halutaan antaa resursseja, ja selviä edistysaskeleita on otettu viime vuosina. Vaikuttaa, että yleisesti uskotaan kehityksen olevan lähitulevaisuudessa nopeaa. Kukaan ei kuitenkaan uskalla ennustaa, milloin tuotantoon merkittävästi vaikuttavia uudistuksia voidaan lanseerata.

Toimistosovelluksiin haaveiltava, näppäilyn sijaan puheen perusteella toimiva tekstinkäsittelyjärjestelmä on vielä hyvin kaukana toteutuksesta. Tällä hetkellä markkinoilla olevissa sovelluksissa ei ole puheentunnistus pohjaista käyttöliittymää tekstin perussyöttöä varten.

2.3 Puheentunnistuksen haasteet

Puheentunnistustehtävän haasteellisuutta lisääviä tekijöitä ovat vaihtelu ja kohina. Vaihtelulla tarkoitetaan puhujien välisiä ääntämiseroja ja kohinalla ylimääräisiä ääniä. (Salmela 2000b: 1–2.) Vaihtelua aiheuttavat esimerkiksi murre-erot, puheviat ja ei-äidinkielisyys. Kohina voi olla ympäristöperäistä, kuten tuulen humina tai kaikuminen, tai puhujaperäistä, kuten maiskaukukset, huokaukset tai yskiminen. Ääntämistavan lisäksi puhujilla on myös persoona- ja tilannekohtaista vaihtelua puheen temmossa ja äänenvoimakkuudessa.

3. Neuroverkot

Neuroverkko (engl. [*artificial*] *neural network*, suom. myös hermoverkko) on tietokoneohjelma, joka voi ns. oppia analysoimaan ja luokittelemaan tietokokonaisuuksia, esimerkiksi tunnistaa puhuttuja tai käsinkirjoitettuja sanoja. Neuroverkko on muihin menetelmiin verrattuna tehokkaampi, jos lopputuloksen ja mittaussuureiden väliset riippuvuudet ovat epälineaarisia. (Jaakohuhta 2001: 366.)

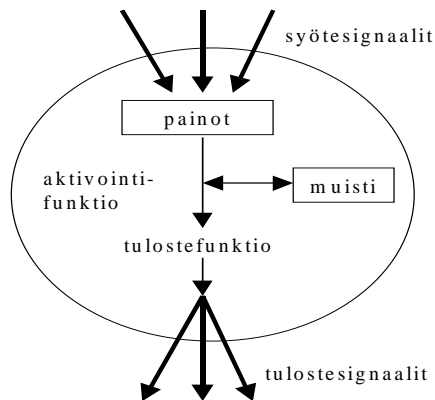
Tikkalan mukaan neuroverkko on mekanismi, joka kokoaa tietoa ympäristöstään ilman, että jokainen erityistapaus olisi ohjelmoitava erikseen. Koska neuroverkot voivat soveltaa aikaisempia tietoja ja muodostaa niiden pohjalta yleistyksiä, verkot toimivat myös muuttuvissa oloissa. Joissakin tapauksissa neuroverkot voivat reagoida 'oikein', vaikka syötteessä olisi kohinaa tai syöte olisi virheellinen. (Tikkala 1997: 16).

Salmelan (2000b: 27) mukaan ensimmäiset matemaattiset neuroverkot ovat peräisin 1940-luvulta. Neuroverkoilla on alunperin haluttu mallintaa erityisesti aivojen toimintaa. Ihmisaivoissa on noin 10^{11} – 10^{12} solmua eli neuronina, jotka kukin ovat yhteydessä muihin neuroneihin noin 10^4 yhteyden kautta (Tikkala: 13). Nykyisellä laskentateholla ja käytettävissä olevilla menetelmillä aivotoimintaa ei pystytä mallintamaan neuroverkoilla realistisesti. Neuroverkkojen avulla voidaan kuitenkin esittää ja ainakin osittain ratkaista huomattavankin kompleksisia ongelmia. Käyttökelpoisuutensa ansiosta neuroverkkotutkimus on levinnyt laajasti eri tieteenaloille, myös kieliteknologian useille osa-alueille. Erityisesti puheteknologiassa on menestyksekkäästi hyödynnetty neuroverkkoja. Niiden avulla voidaan vähentää puhujakohtaisesta vaihtelusta, kuten murteesta tai puheviasta, tai häiriöistä, kuten tuulesta tai yskähdyksestä, johtuvia tunnistusvirheitä.

Tässä luvussa kuvataan neuroverkon yleinen toimintaperiaate ja esitellään kahden neuroverkkotyypin, SOM-karttojen ja MLP-verkkojen, peruspiirteet. SOM-karttoihin ja MLP-verkkoihin pohjautuvat menetelmät ovat erittäin yleisiä niin kieliteknologiassa laajemmin tarkasteltuna kuin erityisesti puheentunnistuksessa.

3.1 Neuroverkon yleinen rakenne ja toimintaperiaate

Neuroverkot koostuvat solmuista, joita neuroverkkojen tapauksessa nimitetään tavallisimmin neuroneiksi. Neuronit on tyypillisesti järjestetty kerroksittain. Neuronien välillä on yhteyksiä, jotka voivat yhdistää samassa kerroksessa tai eri kerroksissa olevia neuroneja. Kullakin yhteydellä on positiivinen tai negatiivinen painotus. Kullakin neuronilla on puolestaan paikallinen muisti, sisäinen prosessi ja tietty laukaisukynnys. Yhteyksien kautta neuroni saa edeltäjiltään syötteitä, jotka se prosessoi paikallisen muistin ja sisäisen prosessinsa avulla. Mikäli laukaisukynnys ylittyy, neuroni välittää syötteen seuraajilleen. Tämä tai nämä puolestaan prosessoivat syötteen sisäisen prosessinsa mukaisesti. Alimman kerroksen neuronien lähettämä tuloste on samalla koko verkon tuloste. (Tikkala 1997: 16-18).



KUVA 1 Neuronin sisäinen toiminta (Tikkalaa 1997: 17 soveltaen).

Tikkalan väitöskirjassa esitellään kolme neuroverkkomallia, jotka on kaikki tarkoitettu simuloimaan suomen kielen sanojen oppimista. Malli 1 on diskreetti sarjamalli, malli 2 interaktiivinen aktivointimalli ja malli 3 takaisinsyöttömalli. Mallit 1 ja 2 simuloivat yksittäisten sanojen tuottamista normaalin ja poikkeavan mallin mukaisesti. Mallit 1 ja 2 eivät ole oppivia. Malli 1 simuloi vain perusmuotojen tuottamista, mallilla 2 voidaan simuloida myös taivutusmuotojen tuottamista. Malli 3 on oppiva. Myös tämä malli simuloi yksittäisten suomen kielen sanojen tuottamista, mutta lapsen oppimisprosessin mukaisesti. (Tikkala 1997: 14).

Tikkalan käyttämistä neuroverkoista kaksi ei ole oppivia, mutta sinänsä oppivuutta pidetään yhtenä neuroverkkojen tärkeimmistä ominaisuuksista. Verkkoja voidaan opettaa useilla eri tavoilla, ja osa verkoista oppii itsestään. Oppimisvaiheessa verkolle annetaan palautetta ratkaisun oikeellisuudesta ja säädellään verkon painotuksia siten, että oikean ratkaisun todennäköisyys kasvaa. (Salmela 1997)

3.2 SOM-kartat

SOM-kartat (engl. *self-organizing map*, suom. myös itseorganisoituva kartta) ovat suomalaisen Teuvo Kohosen 1980-luvulla kehittämä ohjaamattoman oppimisen neuroverkkoarkkitehtuuri. Kartan syötteet ovat tavallisimmin vektoreita. Malli järjestää syötteet niin, että toisiaan muistuttavat kohteet ovat lähekkäin. Kartta mallintaa tietojen tai havaintojen joukkoa. Mallin taustalla olevan matemaattisen kaavan perusteella kartta järjestäytyy itsestään vertaamalla havaintojoukon jäseniä. (Jaakohuhta 2001: 496, Somervuo 2000: 11–12.)

KUVA X SOM-kartan arkkitehtuuri (Salmela 2002: 30, mukaeltu)

SOM-kartan taustalla oleva perusalgoritmi on kaksivaiheinen ja sisällöltään seuraava (Somervuo 2000: 9):

1. Hae kartasta kohde^{*}, joka vastaa parhaiten syötettä käytettävän vertailuperusteen ehtojen mukaisesti. (^{*}Tämä kohde on ns. BMU eli best matching unit.)
2. Päivitä valitun kohteen malli ja valitun kohteen naapurien mallit.

SOM-karttojen merkittävänä etuna pidetään niiden mukailtavuutta. Tyypillisimmillään SOM-kartta koostuu kaksikulotteisista suorakaiteen tai kuusikulmion muotoisista kerroksista, joissa neuronit on järjestetty säännölliselle etäisyydelle toisistaan. Kuten edellä mainittiin, SOM-kartta on tyypillisesti tarkoitettu vektorimuotoisen tiedon käsittelyyn. Sekä rakenteen että tiedon esitysmuodon suhteen voidaan kuitenkin valita myös muita vaihtoehtoja. Rakennetta voidaan muokata sovellusalueen mukaan. Eräässä sovelluksessa on päädytty mm. visuaalisuuden vuoksi kahdeksanulotteiseen Hammingin kuutioon, jossa on 256 neuronia. (Somervuo 2000: 11–12.) Luvussa 4 esitellään ei-vektoripohjaista tietoa käsittelevä SOM-karttasovellus, joka soveltuu merkkijonojen käsittelyyn.

Kuten neuroverkkoja yleisesti, myös SOM-karttoja on sovellettu laajasti eri tieteenalojen tarpeisiin. Kangas ja Kaski (1998) ovat koonneet hakutiedot yli kolmestatuhannesta SOM-karttatutkimuksesta. Näiden joukossa kieliteknologian eri osa-alueet ovat vahvasti edustettuina. Referenssikokoelman perusteella kieliteknologisen tutkimuksen piirissä suosituimpia SOM-sovellusalueita (vuoteen 1998 mennessä) ovat tiedonhaku, puheentunnistus ja käsinkirjoitettujen merkkien tunnistus. Kiinnostava, mutta tässä tutkielmassa tarkempaa valotusta vaille jäävä seikka, on aasialaisiin kieliin keskittyvän kieliteknologian vahva edustus.

3.3 MLP-verkot

MLP-verkot (engl. *multi-layered perceptron [network]*, suom. myös monikerroksiset perceptron-verkot) ovat neuroverkkoja, joissa on vähintään kolme kerrosta eteenpäinsyöttävää kerrosta. MLP-verkko koostuu syötekerroksesta, yhdestä tai useasta piilokerroksesta ja tulostekerroksesta. Mikäli mallissa ei ole lainkaan piilokerroksia, sitä kutsutaan perceptroniksi. (Salmela 1997, Salmela 2002.)

MLP-verkon kerros koostuu neuroneista, jotka ovat laskennallisesti rinnakkaisia yksiköitä. Syötekerroksen neuronit eivät suorita matemaattisia operaatioita, vaan vain välittävät syötteen edelleen. Piilokerrosten neuronit vastaanottavat syötteen edellisen kerroksen neuronilta, suorittavat sisäisen prosessinsa mukaiset operaatiot, ja lähettävät tulosteen seuraavaan kerrokseen. MLP-verkossa yhteydet ovat aina edellisestä kerroksesta seuraavaan. Samaan kerrokseen kuuluvien neuronien välillä ei siis ole yhteyttä, eikä tietoa syötetä takaisin edellisiin kerroksiin (Salmela 2002.)

Salmelan (1997: 12) mukaan MLP-verkkojen käytössä merkittävä etu on siinä, että datan suhteen ei tarvitse pystyä määrittämään ennen käyttöä tilastollisia oletuksia. Verkkoa voidaan opettaa diskriminoivasti eli niin, että oikean lopputuloksen tuottavan prosessin todennäköisyyttä kasvatetaan ja väärin lopputuloksiin johtavien todennäköisyyttä pienennetään.

Verkkomallit ovat usein melko monimutkaisia, ja siksi niiden käsittely on laskennallisesti vaativaa. Yksi mahdollinen tapa alentaa laskennallista monimutkaisuutta on muuttaa MLP-verkon käsittelyä koko verkon yksittäisten neuronien käsittelystä kerroksittaiseksi (Kärkkäinen 2000, Kärkkäinen & Heikkola 2002).

3.4 Neuroverkkojen käyttöön liittyviä epävarmuustekijöitä

Neuroverkkojen käyttöön liittyy tiettyjä epävarmuustekijöitä.

Jotta MLP-verkolla saadaan luotettavia tuloksia, verkon aloituspainotukset on määritettävä huolellisesti. Tähän virittämiseksi kutsuttuun vaiheeseen ei ole tällä hetkellä olemassa varmuudella hyvän tuloksen tuottavia menetelmiä. Usein virittäminen on käsityötä ja siten aikaavievää. (Salmela 2000b: 67–68.)

Neuroverkkolähestymistapaa on kritisoitu myös hajautetusta tiedonesitystavasta, jota ei pidetä ihanteellisena monimutkaisten hierarkkisten rakenteiden kuvaamisessa. Arvostelijoiden mukaan neuroverkkomallien suosio perustuukin lähinnä niitä varten kehitettyjen laskennallisten menetelmien tehokkuuteen verrattuna kilpaileviin malleihin. Erityisesti symbolipohjaisten mallien suosijat arvioivat, että neuroverkot syrjäytyvät, kun kilpailevien mallien laskennallinen suorituskyky kehittyy. (Tikkala 1997: 34.)

Neuroverkkoja ei pidetä erityisen tehokkaina malleina temporaalisten ilmiöiden käsittelyyn. Tämä on koettu puutteeksi varsinkin puheentunnistusmenetelmien kehittämisessä, koska puheen mallintamisessa ajallinen ulottuvuus on pystyttävä ottamaan huomioon. Koska temporaalisuuden käsittelyyn on olemassa muita menetelmiä, on kehitetty erilaisia hybridimalleja, joissa neuroverkkoihin on liitetty apualgoritmeja. (Haykin 1994: 228.)

Somervuo (2000: 14) toteaa erityisesti SOM-verkoista, että vaikka niiden kaksivaiheinen perusalgoritmi onkin sekä logiikaltaan että toteutuksellisesti erittäin yksinkertainen, on verkkojen toiminnan analysoiminen kuitenkin vaikeaa.

4. Neuroverkkojen soveltaminen puheentunnistuksessa

Kun tulkitsemme puhetta, joudumme ratkaisemaan äännetasolla jatkuvasti uusia tilanteita. Opimme yhdistämään kuulemiamme signaaleja riittävän epätarkasti ja käytämme tulkinassa apuna kokemuksen perusteella muodostettua todennäköisyyskäsitystä. Esimerkiksi *Kari* on sangen yleinen miehen etunimi Suomessa. *Kare* puolestaan on suomalainen miehen etunimi sekkin, mutta paljon harvinaisempi kuin *Kari*. Niinpä kuulija usein luottaakin korviensa sijaan aikaisempiin kokemuksiinsa ja luulee kuulevansa nimen *Kari* silloinkin, kun todellisuudessa kuulee sanan *Kare*. Tarvittaessa kuulijat toki oppivat erottamaan nämä nimet. Toisaalta tutunkin sanan tunnistaminen edellyttää älykästä arvaamista. Esimerkiksi r-äänteen puhujakohtainen vaihtelu on huomattavaa, ja jonkin puhujan ääntämistyyllillä *Kale* ja *Kare* saattavat kuulostaa identtisiltä. Meille tämä kuuntelemiseen liittyvä prosessointi on itsestäänselvää, mutta puheentunnistusjärjestelmää laadittaessa on määritettävä, miten järjestelmä tunnistaa ja hyväksyy tai hylkää äänteet, miten se muodostaa äänneistä sanoja ja lauseita sekä miten järjestelmä saadaan tunnistamaan vaihtelu (Salmela 2000b).

4.1 Puheentunnistukseen viritetyt hybridimallit

4.2 HMM-mallien käyttö puheentunnistukseen tarkoitetuissa neuroverkoissa

Puheentunnistuksen tulos on aina hieman epävarma. Tuloksia on pyritty parantamaan todennäköisyysmentelmillä, esimerkiksi HMM-malleilla (*hidden Markov model*, suomeksi esiintyvät ainakin termit Markovin piilomalli, piilotettu Markov-malli ja piiloinen Markov-malli), joita käytetään erittäin yleisesti puheentunnistussovelluksissa. HMM-malli perustuu kaksinkertaiseen stokastiseen prosessiin, joka sisältää ns. piilotetun prosessin. Tätä piilotettua prosessia ei voida tarkastella suoraan, vaan sen toimintaa on analysoitava toisen stokastisen prosessin generoiman symbolisarjan (niin kutsuttujen havainnoitavien kohteiden) kautta. (Haykin 1994, Salmela 1997.)

Sanat kuvataan kuvataan HMM-malleilla, joille on määritetty äärellinen kesto ja jotka etenevät vasemmalta oikealle. Mallit perustuvat joko yksittäisiin foneemeihin eli äänneisiin tai kokonaisiin sanoihin. Äännepohjaisissa malleissa kutakin äännettä varten on oma HMM-mallinsa, ja sanat muodostetaan konkatenoimalla äännepohjaisia HMM-malleja. Sanapohjaisissa järjestelmissä kullekin sanalle on oma erillinen HMM-mallinsa. Mallien tilojen määrä riippuu äänneiden määrästä. (Salmela 1997: 21).

Salmelan (1997: 21) mukaan sanapohjaisia malleja pidetään parempina kuin äännepohjaisia malleja. Kuitenkin käytännössä lähes kaikki sovellukset perustuvat toistaiseksi äännepohjaisiin malleihin,

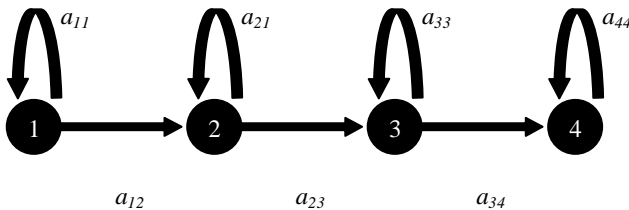
koska sanakohtaisten HMM-mallien määrittäminen laajaan sanastoon on liian työlästä. Äänepohjaiset mallit voidaan luoda vähemmällä parametreillä ja opettamisvaiheessa riittää merkittävästi rajatumpi data kuin sanapohjaisissa malleissa. Sanapohjaisia malleja käytetään joissakin rajatuissa sovelluksissa, esimerkiksi matkapuhelimen numerovalinnassa. Kun numerot luetaan yksittäin, syöteavaruus koostuu lukusanoista 0–9. Tällöin tarvitaan siis kymmenen HMM-mallia.

Seuraavassa HMM-mallin ominaisuudet on kuvattu Haykinin (1994: 226–229) esityksen mukaisesti.

- HMM-mallilla on aina äärellinen määrä mahdollisia tiloja.
- HMM-mallin tiloja vastaavien havainnoitavien kohteiden joukko on äärellinen. Nämä havainnoitavat kohteet vastaavat mallinnettavan järjestelmän kohteita.
- HMM-malliin kuuluville siirtymille on määritetty todennäköisyydet (todennäköisyysjakauma). Siirtymän todennäköisyys ilmentää todennäköisyyttä, jolla järjestelmä tietystä tilasta toiseen siirtymän mukaisesti.
- Havainnoitaville kohteille on määritetty todennäköisyydet todennäköisyydet (todennäköisyysjakauma). Järjestelmä tuottaa aina siirtymän jälkeen saavutettua tilaa vastaavan symbolin. Kutakin tilaa vastaa yksi symboli riippumatta siirtymää edeltäneistä tapahtumista. Jos mallissa on K tilaa, niin myös todennäköisyysjakaumia on K kappaletta.
- Mallille on määritetty aloitushetkeä kuvaava tilatodennäköisyysjakauma.

Puheentunnistuksessa HMM-mallia käytettäessä puhe tulkitaan äärelliseksi automaatiksi (Haykin 1994: 227).

Alla KUVA 3 on kuvattu yksinkertainen HMM-malli, jossa on neljä tilaa. Malli etenee vasemmalta oikealle, josta englanninkielinen nimitys left-right-malli. Esimerkkimalli tunnetaan myös Bakis-mallina. Yleensä mallissa ajatellaan ajan kulkevan vasemmalta oikealle. HMM-mallien etuna pidetään yleisesti niiden kykyä ilmentää ajallisuutta ja käsitellä puhetta ajallisuus huomioon ottaen.



KUVA 3 Nelitilainen HMM-malli (Haykin 1994: 227)

Kuvan siirtymille on määritetty todennäköisyydet a_{ij} seuraavasti:

$$a_{11} + a_{12} = 1$$

$$a_{22} + a_{23} = 1$$

$$a_{33} + a_{34} = 1$$

$$a_{44} = 1$$

4.3 Symbolijonojen käsittelyyn tarkoitettu SOM-sovellus

Kohonen ja Salmela (2000) esittelevät SOM-karttojen sovelluksen, jossa data ei ole vektorimuodossa, vaan symbolijonoina. Tällä mallilla voidaan luokitella merkkijonoja puheentunnistuksessa. Kuvattavassa tapauksessa mallilla luokiteltiin 1 760 puhunnosta, jotka sisälsivät 20 puhujan lausumina 22 suomenkielistä komentosanaa.

Menetelmässä puhe muunnetaan foneemisekvensseiksi HMM-pohjaisella foneemitunnistimella, joka ei sisällä mitään kielioppirajoitteita. Tunnistin ei siis luokittele syötteitä kieliopillisesti hyväksyttäväksi tai hylätyiksi. SOM-verkko opetetaan käyttämään näitä sekvenssejä seuraavan opetusalgoritmin avulla:

1. Hae luetteloon kunkin karttaan kuuluvan kohteen i osalta ne mallijonot, joiden läheisin vertailujono yksikkö i on.
2. Valitse kullekin kartan kohteelle i uudeksi vertailujonoksi keskiarvo tai mediaani yksikön i vierekkäisten solmujen luetteloista kootusta unionista.

Malli on kehitetty puheentunnistukseen, mutta se soveltuu muihinkin tarkoituksiin. Sillä on esimerkiksi luokiteltu kemiallisia yhdisteitä kuvaavia merkkijonoja. Tämä SOM-kartan sovellus on erittäin helppo toteuttaa.

5. Yhteenveto

Mobiililaitteiden yleistymisen on motivoinut tehostamaan puheella ohjattujen käyttöliittymien kehittämistä ja sitä kautta nostanut puheentutkimuksen merkittäväksi tutkimusalaksi matkapuhelinteollisuudessa. Tämä on lisännyt puheentunnistustutkimuksen käytössä olevia resursseja ja potentiaalista taloudellista merkitystä viime vuosina huomattavasti. Puheentunnistuksessa tarvitaan uusia menetelmiä, ja neuroverkot sovelluksineen ovat osoittautuneet käyttökelpoisiksi välineiksi puheentunnistusjärjestelmien kehittämisessä.

Puheentunnistuksessa käytetään erityisesti ns. hybridiverkkoja, joissa neuroverkkoja on yhdistetty keskenään tai joissa on yhdistetty neuroverkkoja ja muita menetelmiä. Erittäin tavallinen malli on MLP-verkkoja, johon on liitetty HMM-malli.

Tässä tutkielmassa on kartoitettu neuroverkkomenetelmän soveltamista puheentunnistuksessa. Mahdollisia aiheeseen liittyviä jatkotutkimuskohteita ovat esimerkiksi morfologisten mallien hyödyntämismahdollisuudet pyrittäessä laajentamaan järjestelmän sallittujen sanojen määrää.

Lähteet

- Ďuriš, D. & Maňuch, J. 2001. On the computational complexity of infinite words. Turku centre for computing science, Technical report No 423.
- Harris, M. D. 1985. Introduction to natural language processing. Reston: Prentice-Hall
- Haykin, S. 1994. Neural networks: a comprehensive foundation. TARKASTA KUSTANNUSPAIKKA JA KUSTANTAJA
- Hovdhaugen, E., Karlsson, F., Henrikson, C. & Sigurd, B. 2000. The history of linguistics in the Nordic countries. Jyväskylä: Gummerus.
- Jonhson, R. 1983. Parsing with transition networks. Teoksessa King, M. (toim.). Parsing natural language. London: Academic Press, 59–75.
- Judd, J. 1991. Neural network design and the complexity of learning. London: Bradford.
- Kangas, J. & Kaski, S. 1998. 3 043 works that have been based on the Self-Organizing Map (SOM) method developed by Kohonen. Helsinki University of Technology. Department of computer science and engineering. Report A49.
- Kohonen, T. & Somervuo, P. 1998. Self-organizing maps of symbol strings. Neurocomputing, 21: 19–30.
- Kärkkäinen, T. 2000. MLP-network in a layer-wise form: derivations, consequences and applications to weight decay. University of Jyväskylä, Reports of mathematical information technology series C. Software engineering and computational intelligence No C 1/2000.
- Kärkkäinen, T. & Heikkola, E. 2002. Robust MLP. University of Jyväskylä, Reports of mathematical information technology series C. Software engineering and computational intelligence No C 1/2002.
- Lehnert, W. & Ringle, M. 1982 (toim.). Strategies for natural language processing. New Jersey: Lawrence Erlbaum Associates.
- Salmela, P. 1997. Spoken digit recognition with neural networks and hidden Markov models. Tampere University of Technology, Signal processing laboratory, Technical report 3-1997.
- Salmela, P. 2000a. Applying dynamic context into MLP/HMM speech recognition system. Tampere University of Technology, Digital and computer systems laboratory 1-2000.
- Salmela, P. 2000b. Neural networks in speech recognition. Tampere University of Technology, Publications 259.
- Salmela, P. 2002. Neural networks for speech recognition. Tampere University of Technology, Institute of digital and computer systems, Technical report 1-2002.

Somervuo, P. 2000. Self-organizing maps for signal and symbol sequences. Acta Polytechnica Scandinavica, Mathematics and computing series No. 107.

Tikkala, A. 1997. Neural networks and lexical processing: three models for Finnish. Kuopio University Publications C. Natural and Environmental Sciences 63. Kuopio: Kuopio University press.

Zierer, E. 1970. The theory of graphs in linguistics. Paris: Mouton.