

Neuroverkkojen peruskäsitteet

Teemu Sarapisto

Kypsyysnäyte
HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

Helsinki, 14. huhtikuuta 2018

Keinotekoisissa neuroverkoissa kulminoituu vuosikymmenten tutkimus funktioiden approksimoinnin menetelmiin. Neuroverkot ovat viimevuosina mahdollistaneet uusien sellaisten tehtävien automatisoinnin, joita on pidetty vain ihmisille mahdollisina. Kaikkivoipia ne eivät kuitenkaan ole, vaan soveltuvat erityisesti tehtäviin, jotka voidaan tulkita jonkin syötevektorin muuntamiseksi ulostulovektoriksi. Tällainen tehtävä on esimerkiksi kuvien sisällön tunnistaminen, jossa kuvan pikseleiden väriarvojen voidaan tulkita olevan syötevektori, ja ulostulovektorin olevan jokaista tunnistettavaa asiaa kohden todennäköisyys sille, että kyseinen asia löytyy syötteenä annetusta kuvasta.

Neuroverkot koostuvat kerroksittain toisiinsa yhdistetyistä laskentayksiköistä joita kutsutaan biologisten esikuviansa mukaisesti neuroneiksi. Tietys- sä kerroksessa olevat neuronit saavat syötteenään niitä edeltävän kerroksen neuroneilta ulostuloarvoja ja vastaavasti tämän kerroksen neuroneiden ulostuloarvot toimivat niitä seuraavan kerroksen syöteinä. Neuronin on tallennettu kullekin syötepoluistaan jokin painotusarvo, joka määrittelee kuinka voimakkaasti kyseistä polkua pitkin tuleva syöte vaikuttaa tämän neuronin ulostuloarvoon. Painotetut syötteet summataan ja syötetään aktivaatiofunktioille, jonka ulostuloarvo toimii koko neuronin yksittäisenä ulostuloarvona. Aktivaatiofunktiona toimii yleensä jokin helposti derivoitavissa oleva funktio, kuten esimerkiksi usein käytetty sigmoidinen funktio.

Ensimmäinen neuroverkon kerros on niin kutsuttu syötekerros, johon neuroverkon syöte koodataan. Tämän jälkeen seuraa yksi tai useampia piilokerroksia joissa varsinainen muunnos syötevektorista ulostulovektoriksi tapahtuu. Viimeisenä kerroksena toimii ulostulokerros, josta neuroverkon laskennan tulos on luettavissa. Käytännössä verkkorakenne on kuitenkin vain ajattelua helpottava abstraktio, sillä neuroverkot ovat todellisuudessa tietokonesovelluksissa vain matriiseja joihin verkon kriittiset arvot tallennetaan.

Neuroverkon rakenteen valitsemisen lisäksi se täytyy harjoittaa jollakin harjoitusaineistolla. Harjoitettaessa neuroverkon tuottaman approksimaation hyvyttä mitataan yleensä virhefunktioilla. Harjoitusvaiheessa neuroverkolle syötetään jokin harjoitusaineiston yksikkö ja koska toivottu tulos syötteelle tunnetaan, neuroverkon syötteesä perusteella tuottaman tuloksen poikkeama halutusta tuloksesta voidaan mitata virhefunktioilla. Koska neuroverkon tuottamat arvot riippuvat harjoitusvaiheessa enää vain sen neuroneiden syötteiden painotuksista, myös virhefunktion arvo riippuu niistä. Neuroverkon harjoittamisen voidaan siis ajatella olevan tämän virhefunktion arvon minimoimista painotusarvoja säätämällä. Tällä hetkellä yleisimmin käytössä oleva menetelmä virhefunktion minimointiin on niin kutsuttu gradienttime- netelmä, jossa painotuksien osittaisderivaatat virhefunktion suhteen toimivat virhefunktion gradientin komponentteina.

Painotuksien osittaisderivaattojen selvittämiseen virhefunktion suhteen käytetään takaisinvirtausalgoritmia. Takaisinvirtausalgoritmissa verkon tul-

kitaan olevan yksi suuri yhdistetty funktio, jolloin virhefunktion derivaatta tietyssä kohtaa verkkoa ja siten tietyn painotuksen kohdalla, on selvitettävissä hyödyntäen derivaatan ketjusääntöä.

Yksi suurimmista haasteista neuroverkkojen harjoittamisessa on neuroverkon liika erikoistuminen tiettyyn harjoitusaineistoon. Tällöin neuroverkko saa harjoitusaineistosta peräisin oleville syötteille toivotusti virhefunktios- ta pieniä arvoja, mutta toimii harjoitusaineiston ulkopuolisten syötteiden kanssa huonosti. Tällöin neuroverkko on epäonnistunut yleistämään harjoitusaineiston piirteitä toivotusti. Ongelman ratkaisuun on kehitetty lukuisia menetelmiä, joista esimerkkinä laskentayksiköiden pudotus, jossa yksittäisiä neuroneita poistetaan satunnaisesti käytöstä estämään tiettyä neuronien erikoistumasta johonkin harjoitusaineiston ominaisuuteen liiaksi.

Onnistuneesti valituilla verkkorakenteilla ja hyvälaatuisella ja suurella harjoitusaineistolla on mahdollista kehittää neuroverkkoja jotka ovat erittäin hyviä löytämään harjoitusaineistostaan yleistettävissä olevia piirteitä. Jotkin näistä piirteistä ovat usein jopa neuroverkot luoneille ihmisille vaikeita täysin ymmärtää. Neuroverkkojen yllättävän hyvän yleistämiskyvyn ansiosta onkin onnistuttu kehittämään neuroverkkoja hyödyntäviä sovelluksia, jotka suoriutuvat ihmistäkin paremmin tehtävissä joita on pidetty ennen mahdottomina tehdä koneellisesti. Neuroverkot ovat yksi viime vuosien tutkituimmista koneoppimisen menetelmistä, ja uutta kehitystä alueella tapahtuu jatkuvasti. Uusia sovelluskohteita joihin neuroverkot soveltuvat löydetään jatkuvasti, sovelluskohteiden sisältäessä kaikkea sairauksien diagnosoinnista videopelien tekoälyihin.