

Aleksander Lempinen

**Ulotteisuuden pienentäminen pääkomponenttianalyysillä
liikeanalyysissä**

Tietotekniikan kandidaatintutkielma

7. huhtikuuta 2019

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Aleksander Lempinen

Yhteystiedot: aleksander.a.lempinen@student.jyu.fi

Ohjaaja: Sanna Mönkölä

Työn nimi: Ulotteisuuden pienentäminen pääkomponenttianalyysillä liikeanalyysissä

Title in English: Dimensionality reduction with principal component analysis in gait analysis

Työ: Kandidaatintutkielma

Sivumäärä: 20+0

Tiivistelmä: Liikeanalyysissä tuotetaan paljon korkeaulotteista mittausdataa, jonka käsitteilyyn tarvitaan usean muuttujan menetelmiä. Suuret datamäärät johtavat myös siihen, että menetelmät tarvitsevat enemmän laskentatehoa. Ohjaamattomaan oppimiseen kuuluva ulotteisuuden pienentämisen menetelmä pääkomponenttianalyysi on laajasti käytössä liikeanalyysissä. Tässä tutkielmassa käsitellään pääkomponenttianalyysin hyödyntämistä kliinisessä liikeanalyysitutkimuksessa.

Avainsanat: pca, pääkomponenttianalyysi, koneoppiminen, liikeanalyysi

Abstract: Human locomotion research or gait analysis measurements produce large amounts of high-dimensional data, which requires a multivariate approach. Large amounts of data also require more computational resources. Principal component analysis which is an unsupervised dimensionality reduction method is in widespread use in gait analysis. This thesis addresses applications of principal component analysis in clinical gait analysis.

Keywords: pca, principal component analysis, machine learning, gait analysis, human locomotion

Kuviot

Kuvio 1. Alkuperäinen data ja data pääkomponenttien suhteen..... 7

Sisältö

1	JOHDANTO	1
2	LIIKEANALYYSI	2
3	KONEOPPIMINEN	4
4	PÄÄKOMPONENTTIANALYYSI	6
5	ESIKÄSITTELY	8
6	PÄÄKOMPOENENTTIEN TULKINTA	11
7	YHTEENVETO.....	13
	LÄHTEET	14

1 Johdanto

Olentojen liikettä on pohdittu sivilisaation alusta lähtien, mutta ensimmäisten joukossa asioita tarkemmin pohti Aristoteles antiikin Kreikassa omissa teoksissaan (Farquharson 2000). Biomekaniikkaa, eli eliöiden rakenteen tai liikkeen tarkastelua mekaniikan avulla, tutki tieteellisesti ensimmäisenä Galileon oppilas Borelli (1681) teoksessaan *De motu animalium*. Biomekaniikka ja liikeanalyysi ovat tasaisesti kehittyneet tieteenalana, mutta viime vuosina teknologian kehityksen myötä biomekaniikkaa on tutkittu todella paljon aikaisempaan tutkimusmäärään verrattuna (Muro-De-La-Herran, Garcia-Zapirain ja Mendez-Zorrilla 2014). Esimerkiksi Muro-De-La-Herran, Garcia-Zapirain ja Mendez-Zorrilla (2014) löysivät jopa 3400 tieteellistä julkaisua, joiden otsikossa esiintyy sana “gait“ eli liike vuosina 2012 ja 2013. Liikeanalyysi biomekaniikan osa-alueena on kehittynyt objektiiviseksi ja hyödylliseksi työkaluksi lääketieteessä ja liikuntatieteissä (Baker ym. 2016).

Perinteiset tilastolliset menetelmät on kehitetty pienelle joukolle piirteitä ja suurelle joukolle havaintoja, mutta nykyään näkee usein suuren joukon piirteitä ja pienen joukon havaintoja (Johnstone ja Titterington 2009). Ulotteisuuden pienentäminen eli piirteiden joukon pienentäminen on tärkeä datan esikäsittelyvaihe monelle koneoppimisen menetelmälle, mahdollistaa korkeaulotteisen datan visualisointia eksploraatiivisessa analyysissä tai piilomuuttujien sitomiseen ilmiöön (Ringnér 2008; Johnstone ja Titterington 2009). Eräs suosittu hyväksi todettu ulotteisuuden pienentämisen menetelmä on PCA eli pääkomponenttianalyysi (Pearson 1901; Hotelling 1933).

Tässä kirjallisuuskatsauksessa tutkitaan, miten pääkomponenttianalyysia on hyödynnetty ulotteisuuden pienentämisessä liikeanalyysissä. Luvussa 2 käsitellään liikeanalyysia ja liikeanalyysidataa yleisellä tasolla. Luvussa 3 käsitellään ohjattua ja ohjaamatonta oppimista ja ulotteisuuden pienentämistä yleisellä tasolla. Luvussa 4 käsitellään pääkomponenttianalyysia tarkemmin. Luvussa 5 käsitellään pääkomponenttianalyysia liikeanalyysissä esikäsittelymenetelmänä ja luvussa 6 pääkomponenttien tulkintaa. Luvussa 7 on yhteenveto.

2 Liikeanalyysi

Liikeanalyysi tarkoittaa ihmisen liikkeen systemaattista mittausta ja tutkimista (Davis III ym. 1991). Vaughin (1999, luku 1) mukaan ihmisen liike alkaa keskushermostossa ja kulkee ääreishermoston kautta lihaksiin, mikä aiheuttaa nivelissä voimia ja momenttia, raajojen siirtymää ja maareaktivoimien muodostumisen. Liikeanalyysi tieteenä sai alkunsa 1600-luvulla, mutta moderni liikeanalyysi on lähtöisin 40-luvulta (Sutherland 2001).

Kliininen liikeanalyysi on hyödyllinen yksilöityä hoitoa varten esimerkiksi tuki- ja liikuntaelinsairauksien diagnosoinnissa tai hoidon suunnittelussa, mikä mahdollistaa paremmat tulokset, kuin pelkästään silmämääräistä ja subjektiivista arviointia käyttämällä (Chambers ja Sutherland 2002). Liikeanalyysia käytetään myös muihin liikkumiseen vaikuttaviin sairauksiin kuten CP-vamman tai Parkinsonin taudin tutkimiseen, millä yritetään ymmärtää sairautta paremmin tai kehittää uusia diagnosointi tai seulontamenetelmiä (Gage 1993). Liikeanalyysia hyödynnetään myös media-alalla, kuten elokuvissa tai peleissä animaatiohahmojen liikkeen parantamiseen. Urheilussa liikeanalyysia voidaan käyttää valmennukseen tai tautien ehkäisemiseen ja liikeanalyysia voidaan hyödyntää myös biometriikassssa turvallisuusalalla.

Liikeanalyysi perustuu sensoreista saatuun dataan. Sensorit voidaan jakaa puettaviin järjestelmiin ja ei-puettaviin järjestelmiin (Muro-De-La-Herran, Garcia-Zapirain ja Mendez-Zorrilla 2014). Liikeanalyysissa hyödynnetään mm. infrapunakameroita, 3D kameroita, lämpökameroita, lattiasensoreita, kiihtyvyyssantureita, goniometrejä, elektrodeja tai markkereita (Muro-De-La-Herran, Garcia-Zapirain ja Mendez-Zorrilla 2014). Raakadatasta tyypillisesti lasketaan useita parametreja kuten maareaktivoimia, kadenssia tai askeleen pituutta (Muro-De-La-Herran, Garcia-Zapirain ja Mendez-Zorrilla 2014). Erilaisia sensoreita voidaan myös käyttää yhdessä eli samanaikaisesti voidaan mitata lihassähkökäyrää elektrodeilla, maareaktivoimia voimalevyllä ja kehon asentoa infrapunakameroilla, mikä tuottaa suuria määriä moniulotteista dataa.

Fukuchi (2017) ehdottaa, että liikeanalyysin monimutkaiseen dataan tarvitaan robustimpaa analyysia, kuin aikaisemmin käytetyt yhden muuttujan tilastolliset menetelmät. Hän huomauttaa, että esimerkiksi tukivektorikoneen luokittelijaan kouluttamiseen yleensä tarvitaan

vain osa piirteistä liikeanalyysikontekstissa (Fukuchi ym. 2011). Liikeanalyysissa usein käytetään ennakkoon laskettuja piirteitä, jotka Phinyomarkin (2018) mukaan perustuvat lähinnä subjektiiviseen mielipiteeseen tai ennakkotietoon ja iso osa datasta jää hyödyntämättä. Hänen mukaan modernista datatieteestä tuttu ulotteisuuden pienentämisen piirteiden valinta ja piirteiden irrottaminen voidaan hyödyntää liikeanalyysissa. Liikeanalyysi raakadatassa voi olla jopa tuhansia muuttujia (Phinyomark ym. 2018).

Ulotteisuuden pienentäminen on siis välttämätöntä liikeanalyysin suurien datamäärien käsittelyssä. Ulotteisuuden pienentäminen karsii turhan informaation ja pienentää datan kokoa, nopeuttaa laskentaa ja joissakin tapauksissa parantaa tuloksia (Fukuchi ym. 2011). Koneellinen ulotteisuuden pienentäminen voidaan jakaa piirteiden valintaan eli turhien piirteiden karsintaan ja piirteiden irrottamiseen eli uusien piirteiden luontiin. Pääkomponenttianalyysi on yleinen liikeanalyysissa käytetty ohjaamattoman piirteiden irrottamisen menetelmä.

3 Koneoppiminen

Historiallisesti järjestelmiin on koodattu tietoa formaalisti käsin kuten esimerkiksi shakin säännöt, mutta epäformaaleissa ja vähemmän abstrakteissa käyttötapauksissa tiedon esittäminen formaalisti tietokonetta varten on hankalaa ja työlästä (Goodfellow, Bengio ja Courville 2016, luku 1). Ratkaisuksi on noussut koneoppiminen, jolla tarkoitetaan kykyä irrottaa tietoa raakadatasta löytämällä hahmoja ja rakenteita (Goodfellow, Bengio ja Courville 2016, luku 1). Tilastotieteessä perinteisesti ollaan keskitytty tietomalleihin eli ymmärtämään dataa tuottavaa luonnollista ilmiötä ja tämän jälkeen mallintamaan sitä helposti tulkittavilla malleilla, mutta koneoppimisessa dataa tuottava luonnollinen ilmiö todetaan olevan liian monimutkainen ja jätetään mustaksi laatikoksi eli tuntemattomaksi ja mallit pyritään saamaan mahdollisimman suorituskkyisiksi, vaikka mallit myös jäisivät mustaksi laatikoksi (Breiman ym. 2001). Koneoppimisessa olennaisena onkin mallien tarkka validointi erillisellä testidatalla ja mallien suorituskkyyn vertailu keskenään, eikä mallien ymmärrettävyys tai ilmiön selittäminen (Shmueli ym. 2010).

Koneoppiminen voidaan jakaa ohjattuun oppimiseen ja ohjaamattomaan oppimiseen (Bishop 2006; Friedman, Hastie ja Tibshirani 2001). Näiden kahden lisäksi joskus puhutaan myös vahvistusoppimisesta (Bishop 2006; Sutton ja Barto 2011). Vahvistusoppiminen on ihmisille intuitiivista ja tarkoittaa tilanteiden yhdistämistä tekoihin niin, että maksimoidaan mahdollisesti myöhemmin tulevaa palkintoa kokeilemalla ja oppimalla virheistä (Sutton ja Barto 2011, luku 1). Tämä yhdistettynä geneettisiin algoritmeihin onkin varsin yleinen tapa esittää koneoppimista helposti ymmärrettävällä tavalla suuremmalle yleisölle esimerkiksi YouTube -videoissa.

Ohjatussa oppimisessa pyritään ennustamaan jokin ennalta valittu vaste eli riippuva muuttuja käyttämällä piirteitä eli riippumattomia muuttujia oppimalla sääntöjä opetusdatasta (Friedman, Hastie ja Tibshirani 2001, luku 2). Tavoite on siis oppia jokin funktio $f(\mathbf{x})$, jolla voidaan ennustaa vaste y piirteiden \mathbf{x} avulla riittävän tarkasti. Opetusdatan sekoittaminen voi johtaa täysin erilaiseen malliin ja yhtä hyvin ennustavia malleja voi olla useampiakin, mikä takia validointi ja vertailu on erittäin tärkeää eli mallien suorituskky arvioidaan testidatalla, jota malli ei ole aikaisemmin nähnyt (Breiman ym. 2001). Mallin ylisovittaminen,

jolla tarkoitetaan sitä, että mallin hyvä suorituskyky opetusdatalla ei yleisty riippumattomalle opetusdatalle, havaitaan juurikin huolellisella validoinnilla (Shmueli ym. 2010). Ohjattu oppiminen vaatii opetusdatan, jossa on vaste saatavilla ja tämä ei ole aina käytännöllistä.

Ohjaamattomassa oppimisessä vastetta ei ole (Bishop 2006, luku 1) eli tarkoitus on irroittaa datassa piilevät rakenteet ja hahmot ilman opetusdatan tarjoamia esimerkkejä. Toisin kuin ohjatussa oppimisessä, ohjaamattoman oppimisen suorituskyvyn tai onnistumisen arviointi on hankalaa ja käytännössä vaatii mielipiteeseen perustuvia oletuksia ja päätöksiä (Friedman, Hastie ja Tibshirani 2001, luku 14). Klusterianalyysissä keskenään samankaltainen data ryhmitellään omiksi joukoiksi eli klustereihin jonkinlaisen erilaisuuden mitan eli etäisyyden avulla (Friedman, Hastie ja Tibshirani 2001, luku 14). Assosiaatiosääntöanalyysi perustuu yhdessä esiintyvien arvojen löytämiseen ja se tunnetaan myös "kauppakassi"analyysinä, jossa siis etsitään yhdessä ostettuja tuotteita (Friedman, Hastie ja Tibshirani 2001, luku 14). Ulotteisuuden pienentäminen tarkoittaa piirteiden määrän pienentämistä niin, että suurin osa informaatiosta alkuperäisestä datasta säilyy (Haykin ym. 2009, luku 1)

Ulotteisuuden pienentämistä voidaan haluta tehdä laskennan nopeuttamiseksi tai esikäsitelymenetelmänä muista syistä (Bishop 2006). Jotkut algoritmit käytännössä vaativat esikäsitelyä tai niihin on sisäänrakennettu jokin ulotteisuutta pienentävä komponentti (Friedman, Hastie ja Tibshirani 2001, luku 1). Esimerkiksi harjanneregressio voi hyödyntää pääkomponenttianalyysia, jotta turhia piirteitä on vähemmän (Friedman, Hastie ja Tibshirani 2001, luku 3).

4 Pääkomponenttianalyysi

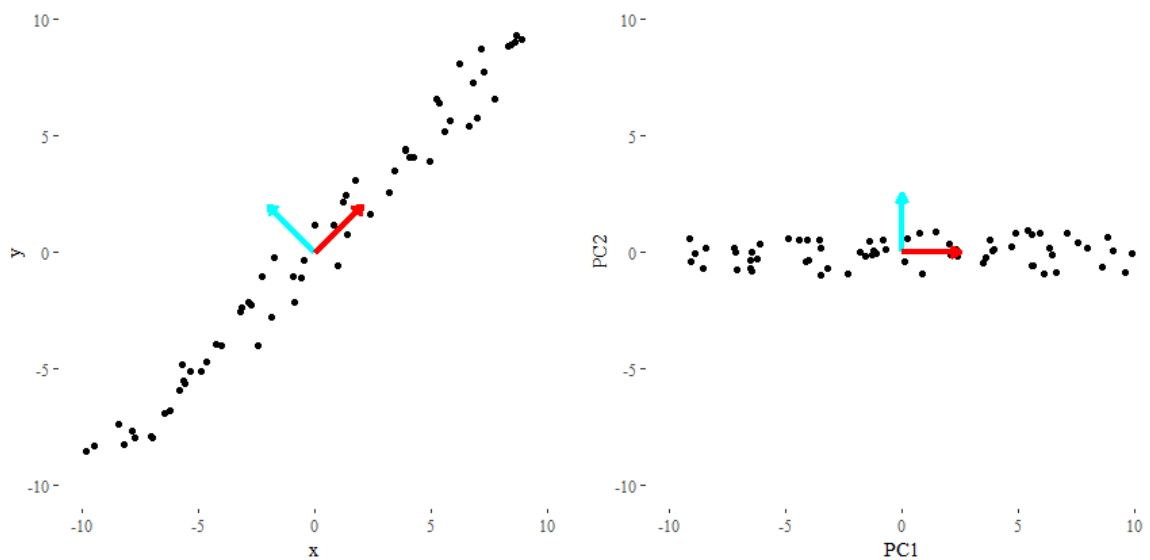
Pääkomponenttianalyysi (PCA) on yleinen menetelmä muun muassa koneoppimisessa, tilastotieteessä, fysiikassa, insinööritieteissä, biologiassa ja lääketieteessä. Pääavoite PCA:n hyödyntämisessä on yleensä ulotteisuuden pienentäminen esikäsittelynä tai esimerkiksi visualisointia varten, mutta sovelluksesta riippuen voidaan myös tulkita tuloksia suoraan. Ulotteisuuden pienentäminen esikäsittelyssä helpottaa laskentaa ja yksinkertaistaa mallit pienemmän piirteiden määrän ansioista. Jos ulottuvuuksia on kolme tai vähemmän, datan visualisointi on mahdollista.

Jolliffen (2011, luku 1) mukaan PCA ja sen tyyppisiä menetelmiä on kehitetty itsenäisesti eri aloilla. Viime vuosisadan alussa PCA on kehitetty esimerkiksi biostatistiikassa (Pearson 1901) ja kasvatopsykologiassa (Hotelling 1933) jopa ennen tietokoneita. Paperilla laskeminen usealla muuttujalla oli Pearsonin (1901) mukaan käytännössä liian hankalaa ja PCA:ta hyödynnettiin vasta tietokoneiden yleistymisen myötä. Hotellingin ajatus lähti liikkeelle faktorianalyysistä ja päättyi siihen, että voi olla olemassa pienempi määrä toisistaan riippumattomia muuttujia, joista alkuperäisen datan muuttujat koostuvat ja ihmistieteellisestä kirjallisuudesta poiketen päätti kutsua nämä riippumattomat muuttujat komponenteiksi eikä fakto-reiksi. Faktorianalyysi usein sekoitetaankin PCA:han erityisesti ihmistieteissä (Jolliffe 2011; Suhr 2005; Fabrigar ym. 1999, luku 1). Faktorianalyysissä pyritään löytämään datan sisäinen rakenne piilomuuttujien avulla, kun taas pääkomponenttianalyysi ei liity datan sisäiseen rakenteeseen tai piilomuuttujiin millään tavalla (Fabrigar ym. 1999). Pääkomponentteja lähtökohtaisesti ei voi tulkita, mutta käytännössä tätä silti tapahtuu erilaisista syistä (Jolliffe 2011; Suhr 2005).

PCA:n perusidea Jolliffen (2011, luku 1) mukaan on pienentää datan ulotteisuutta luomalla uudet muuttujat, jotka säilyttävät mahdollisimman paljon vaihtelua, eivät korreloi keskenään ja järjestynä muutama ensimmäinen pääkomponentti sisältää suurimman osan alkuperäisen datan vaihtelusta. Ringner (2008) vastaavasti kuvailee pääkomponentit suuntina, joissa vaihtelu on datassa suurinta. Kuviossa 1 sekä x , että y sisältävät yhtä paljon informaatiota, eikä kumpaakaan voi jättää pois. Ne kuitenkin korreloivat keskenään. Luomalla uusi pääkomponentti PC1 suurimman vaihtelun suuntaisesti (punainen nuoli) ja siihen kohtisuoraan

pääkomponentti PC2 (vaaleansininen nuoli) saadaan uudet muuttujat PC1 ja PC2, jotka eivät korreloi keskenään. Valitsemalla vain ensimmäiset pääkomponentit voidaan pienentää datan ulotteisuutta, koska jälkimmäiset pääkomponentit sisältävät vähemmän informaatiota. Nykyään PCA on tyypillisesti toteutettu pääakselihajotelman avulla (Jolliffe 2011, liite A.1).

Kuvio 1. Alkuperäinen data ja data pääkomponenttien suhteen
Pääkomponenttien geometrinen tulkinta



5 Esikäsittely

Monien menetelmien algoritminen kompleksisuus on riippuvainen piirteiden määrästä. Ulotteisuuden pienentäminen onkin usein tarpeellista laskennan helpottamiseksi erityisesti, jos kyse on aikasarjadatasta, kuten tuhansia näytteitä sisältävästä signaalista. Seuraavaksi tarkastellaan liikeanalyysin tutkimuksia, joissa pääkomponenttianalyysia on käytetty esikäsittelymenetelmänä.

Siljeppevic ja muut (2018) pyrkivät luokittelemaan liikkeeseen vaikuttavia häiriöitä pelkästään maareaktiivoimadatan avulla. Heidän mielestä silmämääräinen arviointi tai videon avulla arviointi on liian subjektiivista ja 3D-liikeanalyysi vaatii liikaa resursseja ja aikaa päivittäisessä kliinisessä käytössä. Potilaat jaettiin ryhmiin diagnoosien perusteella ja kontrolliryhmänä toimivat terveet koehenkilöt. Koehenkilöt kävelivät voimalevyjen päältä, mikä tuotti viisi eri analogista signaalia. Voimalevyn analogiset signaalit käsiteltiin digitaalisiksi ja normalisoitiin.

Digitaaliset signaalit käsiteltiin pääkomponenttianalyysilla, koska pääkomponenttianalyysi kykeni ottamaan koko signaalin ja muuttamaan sen piirteiksi (Slijeppevic ym. 2018). Tämän lisäksi he piirteyttivät signaalit käsin laskemalla aikaisempaan tutkimukseen perustuvia parametreja raakasignaalista. Käsin lasketut piirteet, pääkomponenttianalyysilla saadut piirteet ja näiden yhdistelmää käytettiin syötteenä eri luokittelijoille. Tuloksena oli, että parhaat tulokset saatiin käsittelemällä raakasignaali pääkomponenttianalyysilla ja lisäämällä siihen muutama käsin laskettu piirre. Pääkomponenttianalyysilla esikäsitellyllä datalla ilman käsin laskettuja piirteitä luokittelijan suorituskyky oli pahimmassa tapauksessa 5 % huonompi ja parhaassa tapauksessa 0.1 % parempi.

Pääkomponenttianalyysia voidaan hyödyntää puhtaasti ulotteisuuden pienentämisen menetelmänä, mutta tällöin vaihtoehtoisia menetelmiä on paljon. Joskus pääkomponenttianalyysin ominaisuuksilla on analyysin kannalta merkitystä ja samoja ominaisuuksia ei löydy muilta menetelmiltä.

McGrath, Fineman ja Stirling (2018) pyrkivät hajoittamaan reiden ja säären inertiamittausyksiköiden suhteellisen kulmanopeuden polven rotaatioakseliin niin, että inertiamittausyksiköi-

tä ei tarvinnut kalibroida tai sijoitella kovin tarkasti. Koehenkilöiden reisiin ja sääriin kiinnitettiin inertiamittausyksiköt ja he nousivat tuoilta ylös ja kävelivät muutaman askeleen eteenpäin. Inertiamittausyksiköt sisälsivät kiihtyvyyssanturin, gyroskoopin ja magnetometrin. Raakadata ensin käsiteltiin laitevalmistajan Kalman-suodattimella.

Mallintamalla polven rotaatiota yhden akselin saranana aikaisempaan lääketieteelliseen tutkimukseen perustuen pystyttiin approksimoimaan polven akselia pääkomponenttianalyysin ensimmäisen pääkomponentin avulla (McGrath, Fineman ja Stirling 2018). Heidän mukaansa pääkomponenttianalyysi vaatii vähemmän laskentaa edellisiin polvea approksimoiviin menetelmiin verrattuna. Tämä on mahdollista, koska polven rotaatioakselin suurin vaihtelu tapahtuu jalan ojennuksessa tai loitonuksessa ja polven muut rotaatioakselit ovat siihen nähden ortogonaalisia, mikä johtaa ensimmäisen pääkomponentin olevan juurikin polven rotaatioakseli ojennuksessa ja loitonuksessa. Tästä McGrath, Fineman ja Stirling (McGrath, Fineman ja Stirling 2018) laskivat liikeanalyysin kannalta oleellisia arvoja ilman kalliita kamerajärjestelmiä tai inertiamittausyksiköiden kalibrointia.

Raakadatan lisäksi voidaan käsitellä myös raakadatasta laskettuja piirteitä. Tällöin ei yleensä puhuta tuhansista piirteistä havaintoa kohti esimerkiksi signaalin muodossa vaan kymmenistä piirteistä, joilla on jonkinlainen merkitys.

Amene ja muut (2019) pyrkivät löytämään pes planovalgus -tyyppisestä lattajalasta kärsivien CP-vammaisten keskuudesta aliryhmiä. Heidän mielestä näiden aliryhmien löytäminen tulee parantamaan potilaiden hoitoa. Potilaiden jalkoihin kiinnitettiin heijastavia markkereita ja heidän kävelyä kuvattiin 14 kameran Vicon-järjestelmällä. Jaloista otettiin myös röntgenkuvat, joiden avulla kalibroitiin markkereiden sijainnit suhteessa luihin.

Aikaisempaan lääketieteelliseen tutkimukseen perustuen raakadatasta laskettiin 34 liikeanalyysin kannalta olennaista muuttujaa, jotka toimivat normalisoinnin jälkeen syötteenä korrelaatiomatriisiin perustuvalla pääkomponenttianalyysille (Amene ym. 2019). Eri piirteille asetettiin kriteerejä ja suoritettiin piirteiden valinta, koska haluttiin säilyttää yhteys pääkomponenttien ja piirteiden välillä. Pääkomponenteista ensimmäiset 7 käytettiin syötteenä klusterianalyysia varten. Tämän jälkeen pääkomponentteja yritettiin tulkita eri klustereille.

Esikäsitteilymenetelmänä pääkomponenttianalyysin ensimmäisiä pääkomponentteja voidaan

hyödyntää sisääntulona muihin menetelmiin uhraamatta liikaa suorituskykyä. Myös muita pääkomponenttianalyysin ominaisuuksia voidaan hyödyntää tai jopa yrittää tulkita pääkomponentteja.

6 Pääkomponenttien tulkinta

Vaikka pääkomponenttianalyysia on lähtökohtaisesti vaikea tulkita, pääkomponenttien tulkintaa tehdään silti erityisesti faktorianalyysiin yhteydessä. Seuraavaksi tarkastellaan liikeanalyysin tutkimuksia, joissa pääkomponentteja on onnistuneesti tulkittu.

Kobayashi ja Ogata (2018) pyrkivät löytämään yhteyksiä vanhusten kävelykykyä arvioivaa Two-step -testin pisteiden ja vanhusten liikkeestä löytyvien hahmojen välillä kehittääkseen ohjeistusta ehkäisemään ongelmia liikkumisessa. Heidän mukaansa vanhusten kaatuminen on suuri terveystarve ja Two-Step -testin tulokset viittaavat kaatumisriskiin, mutta testin yhteyttä liikkeen hahmoihin ei tunneta. Koehenkilöihin kiinnitettiin heijastavia marketteja ja heitä kuvattiin Vicon kamerajärjestelmällä. Tämän lisäksi potilaat kävelivät voimalevyjen päältä.

Voimalevyjen raakasignaaleista laskettiin käsin 101 piirrettä ja kamerajärjestelmän ohjelmistolla laskettiin nivelten kulmia ja momenteja ja muita liikeanalyysin kannalta oleellisia piirteitä (Kobayashi ja Ogata 2018). Nämä lasketut piirteet syötettiin pääkomponenttianalyysiin ja saatuja pääkomponentteja verrattiin piirteisiin Pearsonin tulomomenttikorrelaatiokertoimen avulla. He toteavat, että ainoastaan toinen pääkomponentti korreloi Two-Step -testin pisteiden kanssa.

Tämänkaltaisen suoraviivainen tapa löytää helposti tulkittavia pääkomponentteja on yleensä hankalaa, jos alkuperäisiä piirteitä on paljon (Jolliffe ja Cadima 2016). Pääkomponenttien tulkinnan helpottamiseksi tyypillisesti käytetään faktorianalyysistä tuttua pääkomponenttien rotaatiota (Jolliffe 2011; Jolliffe ja Cadima 2016).

Kikkert ja muut (Kikkert ym. 2017) pyrkivät parantamaan kaatumisriskimallien ennustekykä poistamalla turhia piirteitä ja lisäämällä piirteitä potilaiden kognitiivisista kyvyistä. Heidän mukaan 30 % yli 65-vuotiaista vanhuksista kaatuu vähintään kerran vuodessa, mikä voi johtaa jopa kuolemaan ja kaatumisien ehkäisy on tärkeää. Potilaiden kognitiivisia kykyjä testattiin ja pisteytettiin, minkä jälkeen potilaiden alaselkään kiinnitettiin kiihtyvyyssantureita ja he kävelivät 160 metrin matkan, jonka aikana kerättiin dataa.

Kerätystä raakadatasta laskettiin käsin 11 liikeanalyysin kannalta oleellista piirrettä ja tämä käsiteltiin pääkomponenttianalyysillä yhdistettynä Varimax-rotatioon ja Kaiser-normalisointiin (Kikkert ym. 2017). Tämän tyyppinen rotaatio on tyypillisesti faktorianalyysissä käytetty menetelmä ja pääkomponentteihin sovellettuna se voi joskus helpottaa pääkomponenttien tulkintaa (Jolliffe 2011, luku 7). Tämän jälkeen Kobayashi ja Ogata (2017) tulkitsivat, että ensimmäinen pääkomponentti liittyy kävelyn nopeuteen, askelsyklin aikaan ja kehon kiihtyvyyteen eli toisinsanoen kävelytahtiin. Toinen pääkomponentti liittyi liikkeen vaihteluun ja kolmas pääkomponentti koordinaatioon. Nämä kolme pääkomponenttia syötettiin regressiomalleihin ja näiden suorituskykyä arvioitiin ja analysoitiin eteenpäin.

Pääkomponenttianalyysin tuottamia pääkomponentteja voidaan siis tietyissä tapauksessa tulkita, jos huomataan pääkomponentin liittyvän johonkin ilmiöön. Pääkomponenttianalyysistä kehittyikin paljon eri versioita ja osa näistä on huomattavasti helpommin tulkittavissa (Jolliffe ja Cadima 2016).

7 Yhteenveto

Pääkomponenttianalyysia on laajasti käytetty koneoppimisen menetelmä liikeanalyysissä siitä huolimatta, että se on vanha ja yksinkertainen menetelmä. Pääkomponenttianalyysia on hyödynnetty sekä puhtaasti ulotteisuuden pienentämisessä laskennan helpottamiseksi, että on hyödynnetty pääkomponenttianalyysin muita ominaisuuksia. Se on osoittautunut hyödylliseksi riippumatta siitä, että käytetäänkö sitä raakaan signaalidataan vai raakadatasta käsin laskettuihin piirteisiin. Yhdistämällä pääkomponenttianalyysi faktorianalyysiin pystyttiin pääkomponentteja myös tulkitsemaan.

Jatkotutkimusta varten on selkeä tarve kartoittaa myös muita ulotteisuuden pienentämisen menetelmiä ja erityisesti vertailla eri menetelmillä saatuja tuloksia. Liikeanalyysi on kuitenkin suhteellisen nuori tiedeala ja muilla aloilla laajasti käytettyjä menetelmiä ei välttämättä ole vielä sovellettu laajasti nimenomaan liikeanalyysitutkimuksessa.

Lähteet

- Amene, Juliet, Joseph J Krzak, Karen M Kruger, Logan Killen, Adam Graf, Haluk Altıok, Peter A Smith ja Gerald F Harris. 2019. “Kinematic foot types in youth with pes planovalgus secondary to cerebral palsy”. *Gait & posture* 68:430–436. doi:10.1016/j.gaitpost.2018.12.026.
- Baker, Richard, Alberto Esquenazi, Maria Grazia Benedetti ja Kaat Desloovere. 2016. “Gait analysis: clinical facts.” *European journal of physical and rehabilitation medicine* 52 (4): 560–574.
- Bishop, Christopher M. 2006. *Pattern recognition and machine learning*. springer.
- Borelli, Giovanni Alfonso. 1681. *De motu animalium*. Romae: Bernabò.
- Breiman, Leo, ym. 2001. “Statistical modeling: The two cultures (with comments and a rejoinder by the author)”. *Statistical science* 16 (3): 199–231.
- Chambers, Henry G, ja David H Sutherland. 2002. “A practical guide to gait analysis”. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* 10 (3): 222–231. doi:10.5435/00124635-200205000-00009.
- Davis III, Roy B, Sylvia Ounpuu, Dennis Tyburski ja James R Gage. 1991. “A gait analysis data collection and reduction technique”. *Human movement science* 10 (5): 575–587. doi:10.1016/0167-9457(91)90046-Z.
- Fabrigar, Leandre R, Duane T Wegener, Robert C MacCallum ja Erin J Strahan. 1999. “Evaluating the use of exploratory factor analysis in psychological research.” *Psychological methods* 4 (3): 272. doi:10.1037/1082-989X.4.3.272.
- Farquharson, ASL. 2000. *On the gait of animals by Aristotle*. http://classics.mit.edu/Aristotle/gait_anim.html.
- Friedman, Jerome, Trevor Hastie ja Robert Tibshirani. 2001. *The elements of statistical learning*. Nide 1. 10. Springer series in statistics New York.

- Fukuchi, Reginaldo K, Bjoern M Eskofier, Marcos Duarte ja Reed Ferber. 2011. "Support vector machines for detecting age-related changes in running kinematics". *Journal of Biomechanics* 44 (3): 540–542. doi:10.1016/j.jbiomech.2010.09.031.
- Fukuchi, Reginaldo K, Claudiane A Fukuchi ja Marcos Duarte. 2017. "A public dataset of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics". *PeerJ* 5:e3298. doi:10.7717/peerj.3298.
- Gage, James R. 1993. "Gait analysis. An essential tool in the treatment of cerebral palsy." *Clinical orthopaedics and related research*, numero 288: 126–134.
- Goodfellow, Ian, Yoshua Bengio ja Aaron Courville. 2016. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Haykin, Simon S, Simon S Haykin, Simon S Haykin, Kanada Elektroingenieur ja Simon S Haykin. 2009. *Neural networks and learning machines*. Nide 3. Pearson Upper Saddle River.
- Hotelling, Harold. 1933. "Analysis of a complex of statistical variables into principal components." *Journal of educational psychology* 24 (6): 417.
- Johnstone, Iain M, ja D Michael Titterton. 2009. *Statistical challenges of high-dimensional data*. doi:10.1098/rsta.2009.0159.
- Jolliffe, Ian. 2011. *Principal component analysis*. Springer.
- Jolliffe, Ian T, ja Jorge Cadima. 2016. "Principal component analysis: a review and recent developments". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2065): 20150202. doi:10.1098/rsta.2015.0202.
- Kikkert, Lisette HJ, Maartje H De Groot, Jos P van Campen, Jos H Beijnen, Tibor Hortobágyi, Nicolas Vuillerme ja Claudine CJ Lamoth. 2017. "Gait dynamics to optimize fall risk assessment in geriatric patients admitted to an outpatient diagnostic clinic". *PloS one* 12 (6): e0178615. doi:10.1371/journal.pone.0178615.
- Kobayashi, Yoshiyuki, ja Toru Ogata. 2018. "Association between the gait pattern characteristics of older people and their two-step test scores". *BMC geriatrics* 18 (1): 101. doi:10.1186/s12877-018-0784-5.

- McGrath, Timothy, Richard Fineman ja Leia Stirling. 2018. “An Auto-Calibrating Knee Flexion-Extension Axis Estimator Using Principal Component Analysis with Inertial Sensors”. *Sensors* 18 (6): 1882. doi:10.3390/s18061882.
- Muro-De-La-Herran, Alvaro, Begonya Garcia-Zapirain ja Amaia Mendez-Zorrilla. 2014. “Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications”. *Sensors* 14 (2): 3362–3394. doi:10.3390/s140203362.
- Pearson, Karl. 1901. “LIII. On lines and planes of closest fit to systems of points in space”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11): 559–572.
- Phinyomark, Angkoon, Giovanni Petri, Esther Ibáñez-Marcelo, Sean T Osis ja Reed Ferber. 2018. “Analysis of big data in gait biomechanics: Current trends and future directions”. *Journal of medical and biological engineering* 38 (2): 244–260. doi:10.1007/s40846-017-0297-2.
- Ringnér, Markus. 2008. “What is principal component analysis?” *Nature biotechnology* 26 (3): 303. doi:10.1038/nbt0308-303.
- Shmueli, Galit, ym. 2010. “To explain or to predict?” *Statistical science* 25 (3): 289–310.
- Slijepcevic, Djordje, Matthias Zeppelzauer, Anna-Maria Gorgas, Caterine Schwab, Michael Schüller, Arnold Baca, Christian Breiteneder ja Brian Horsak. 2018. “Automatic Classification of Functional Gait Disorders”. *IEEE journal of biomedical and health informatics* 22 (5): 1653–1661. doi:10.1109/JBHI.2017.2785682.
- Suhr, Diana D. 2005. “Principal component analysis vs. exploratory factor analysis (paper 203-30)”. Teoksessa *Proceedings of the thirtieth annual SAS® users group international conference*, 203:30.
- Sutherland, David H. 2001. “The evolution of clinical gait analysis part I: kinesiological EMG”. *Gait & posture* 14 (1): 61–70. doi:10.1016/S0966-6362(01)00100-X.
- Sutton, Richard S, ja Andrew G Barto. 2011. “Reinforcement learning: An introduction”.
- Vaughan, Christopher L, Brian L Davis, CO Jeremy ym. 1999. “Dynamics of human gait”.