

CART-algoritmin hyödyntäminen liikkuvien työkoneiden suorituskykyindikaattoreiden estimoinnissa

Teemu Väyrynen

Tampereen teknillinen yliopisto, systeemitekniikan laitos
Korkeakoulunkatu 10, 33720, Tampere, Suomi
Puh. 050 377 5196, teemu.vayrynen@tut.fi

Suvi Paananen

Tampereen teknillinen yliopisto, systeemitekniikan laitos
Korkeakoulunkatu 10, 33720, Tampere, Suomi
suvi.paananen@tut.fi

Eero Anttila

Tampereen teknillinen yliopisto, systeemitekniikan laitos
Korkeakoulunkatu 10, 33720, Tampere, Suomi
eero.anttila@tut.fi

Matti Vilkkö

Tampereen teknillinen yliopisto, systeemitekniikan laitos
Korkeakoulunkatu 10, 33720, Tampere, Suomi
Puh. 040 833 2830, matti.vilkko@tut.fi

AVAINSANAT suorituskykyindikaattori, estimointi, liikkuva työkone, CART, regressiopuu

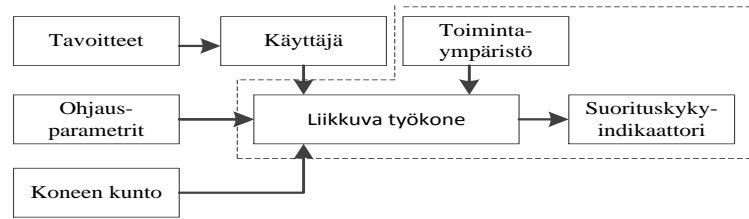
Tiivistelmä

Viime vuosina liikkuvien työkoneiden suorituskyvyn analysoinnista sekä optimoinnista on tullut entistä tärkeämpää. Analyysi- ja optimointitoimenpiteiden onnistumisen kannalta tärkeimpiä yksittäisiä tekijöitä ovat työkoneiden käyttäjien taitotaso sekä koneen suorituskykyä kuvaavien indikaattoreiden tarkkuus ja luotettavuus. Tässä työssä esitellään CART-algoritmia hyödyntävä data-analyysimenetelmä liikkuvien työkoneiden suorituskyvyn estimointiin. Esitellyn menetelmän tärkeimpiä etuja ovat täysin mittausdataan pohjautuva mallinnus sekä mahdollisuus muodostaa suhteellisia suorituskykyindikaattoreita. Suhteellisilla suorituskykyindikaattoreilla tarkoitetaan suorituskykyarvoja, jotka on liitetty tiettyyn liikkuvan työkoneen toimintaympäristön kuvaukseen matemaattisesti. Menetelmän avulla yksittäisen työkoneen suorituskykyä voidaan verrata laajaan referenssiaineistoon. Täten suorituskykyindikaattoreista tulee luotettavampia ja informatiivisempia työkoneiden käyttäjille. Tämän työn luvussa tulokset esitellään menetelmän soveltamista liikkuvista työkoneista kerättyyn mittausdataan.

1 JOHDANTO

Viime vuosina liikkuvien työkoneiden suorituskyvyn analysoinnista sekä optimoinnista on tullut entistä tärkeämpää [1–2]. Erityisesti liikkuvien työkoneiden käyttäjät ovat alkaneet kiinnittää enenevässä määrin huomiota koneidensa tuottavuuteen ja tehokkuuteen. Suorituskykyindikaattoreihin voidaan kiinnittää enemmän huomiota, koska liikkuvien työkoneiden toimintavarmuus on parantunut merkittävästi viime vuosien aikana.

Liikkuvien työkoneiden suorituskykyindikaattorien arvojen tulkintaan vaikuttaa moni asia. Tällaisia asioita ovat esimerkiksi työn tavoitteet, työkoneen ohjausjärjestelmän parametrit, toimintaympäristö, työkuorma, käyttäjän taitotaso sekä työkoneen mekaanisten osajärjestelmien kunto. Valitettavasti vain kahta ensin mainittua voidaan tarkastella luotettavasti ja reaaliaikaisesti. Moderni liikkuva työkone on kompleksinen kokonaisuus, mistä johtuen jokainen kone on hieman yksilöllinen mekaanisten ominaisuuksiensa osalta. Koneiden valmistustekniikoiden kehittyessä yksittäisten koneiden väliset erot ovat kuitenkin pienentyneet, mikä helpottaa data-analyysimenetelmien käyttöä laajaa konekantaan hyödyntäen. Kuvassa 1 on esitetty liikkuvan työkoneen suorituskykyindikaattoreihin vaikuttavia tekijöitä sekä katkoviivan sisällä rajattuna tämän työn aihealue.



Kuva 1. Liikkuvan työkonteen suorituskykyindikaattoreihin vaikuttavat tekijät.

Moderni liikkuva työkone koostuu useista integroiduista osajärjestelmistä /3/. Joitakin näistä voidaan säätää työkonteen automaatiojärjestelmän parametreilla. Riippuen työkonteen kompleksisuudesta, automaatiojärjestelmä voi mahdollistaa käyttäjälle jopa satojen eri parametrien muokkauksen omien mieltymyksiensä sekä toimintaympäristön perusteella. Näillä ohjausparametreilla on merkittävä vaikutus työkonella saavutettavaan suorituskykyyn.

Perinteisesti ohjausparametrien viritys on suoritettu kokeneiden käyttäjien ja konevalmistajien asiantuntijoiden kehittämällä heuristisilla säännöillä. Liikkuvan työkonteen viritäminen on kuitenkin erittäin haastava ja aikaa vievä operaatio kokemattomille käyttäjille. /4/ Jotta koneen ohjausparametrien viritäminen voidaan suorittaa tarkasti, toiminnasta tarvitaan palautetta. Moderneissa liikkuvissa työkonneissa on yleensä runsaasti antureita, joilta kerätty mittaustieto voidaan esittää käyttäjälle koneen käyttöliittymän kautta. Tämän mittaustiedon tulkinta jätetään kuitenkin nykyisellään usein yksin käyttäjän vastuulle.

Suorituskyvyn analysointi ja siten myös liikkuvien työkonneiden ohjausparametrien viritystapa vaihtelee nykyisin merkittävästi käyttäjien välillä. Suorituskykyindikaattorien virheellinen tulkinta voi johtaa helposti väärin toimenpiteisiin. Tällainen tilanne voi syntyä, jos käyttäjä ei osaa yhdistää suorituskykyyn vaikuttavia muuttujia saavuttamaansa suorituskykyarvoon.

Aiempi tutkimustyö suorituskykyindikaattoreiden estimoinnissa kehittyneiden data-analysikeinojen avulla on tuottanut lupaavia tuloksia. Useita eri menetelmiä on esitetty ratkaisuksi kyseiseen ongelmaan erilaisissa sovelluskohteissa. Menetelmät vaihtelevat yksinkertaisesta mittausarvojen visualisoinnista aina monimutkaisiin ja kokonaisvaltaisiin estimointimenetelmiin. Esimerkkinä yksinkertaisesta suorituskyvyn estimoinnista on auton keskimääräinen polttoainekulutus. Erilaisia koneoppimisen menetelmiä on sovellettu myös metsäkoneisiin /1–2/, maatalouskoneisiin /5–9/ ja teollisuusprosesseihin /10–12/. Lähteet 1 ja 2 ovat tämän työn kannalta tärkeintä tausta-aineistoa, koska ne käsittelevät myös suhteellisen suorituskyvyn käsitettä, vaikka toteutus onkin erilainen.

Tämän työn tutkimusongelma keskittyy liikkuvien työkonneiden suorituskyvyn arviointiin eri toimintaympäristöissä. Tutkimusongelmasta johdettuna tämän työn tutkimuskysymys on: "Miten data-analysointia voidaan hyödyntää liikkuvien työkonneiden suorituskykyindikaattoreiden estimoinnissa?" Tämän työn tavoitteena on esittää yksi data-analysimenetelmä ratkaisuksi määritettyyn tutkimuskysymykseen.

Esiteltävässä ratkaisumallissa hyödynnetään datan esikäsittelyä ja CART-algoritmia. CART (engl. Classification and Regression Trees) luokitellaan koneoppimismenetelmäksi, ja sen ovat kehittäneet Breiman ym. /13–14/. Tässä työssä CART-algoritmia käytetään muodostamaan malleja suorituskykyindikaattoreiden ja toimintaympäristöön liittyvien mittausmuuttujien välille. Esiteltävä menetelmä mahdollistaa suorituskykyindikaattoreiden arvioinnin suhteellisina eli liitettyinä tiettyihin toimintaympäristöihin. Kontekstiedon liittäminen suorituskykyindikaattoreihin lisää tulosten luotettavuutta ja helpottaa saavutettujen suorituskykyarvojen tulkintaa.

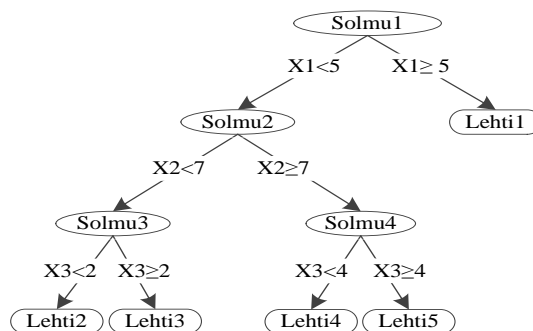
Työn luku 2 sisältää aihepiirin taustatutkimusta. Luvussa 3 kuvataan lyhyesti työssä käytettävät menetelmät. Luvussa 4 esitellään työssä käytetty data-aineisto ja tulokset. Työstä tehdyt johtopäätökset ja jatkotutkimusehdotukset esitellään luvussa 5.

2 MENETELMÄT

Tässä työssä esiteltävässä data-analysimenetelmässä yhdistetään datan esikäsittelymenetelmiä /15/ ja CART-algoritmi. CART rakentaa joko luokittelu- tai regressiopuun käytettävän data-aineiston perusteella. Luokittelupuu muodostetaan, mikäli selitettävä muuttuja voi saada vain rajallisen määrän ennalta määrättyjä ja diskreettejä arvoja. Jos taas selitettävä muuttuja saa jatkuvia reaalilukuarvoja, CART muodostaa regressiopuun. Tässä työssä suorituskykyindikaattoreita mallinnetaan liikkuvan työkonteen toimintaympäristöä kuvaavien muuttujien avulla. Muuttujien välisiä vuorovaikutuksia kuvaamaan muodostetaan regressiopuu, jonka avulla voidaan ennustaa tyypillisiä suorituskykyindikaattorien arvoja tietyissä toimintaympäristöissä. /13/

CART muodostaa regressiopuun jakamalla data-aineistoa alijoukkoihin eli solmuihin, jotka yhdistetään toisiinsa haaroilla. Regressiopuun alimpia solmuja nimitetään lehdiksi, joissa jokaisessa on omankaltaisensa toimintaympäristö ja sitä vastaava keskimääräinen suorituskyskyarvo. CART-menetelmän soveltamiseksi data-aineisto tulee jakaa kahteen osaan: opetusdataan sekä validointidataan. Ensin mainittua käytetään regressiopuun muodostamiseen ja jälkimmäistä mallin ennustuskysyvyn arviointiin. /13/

Regressiopuun rakentaminen aloitetaan juurisolmusta, joka sisältää kaiken data-aineiston. Tämän jälkeen suoritetaan binaarinen jako data-aineistolle yhden selittävän muuttujan avulla siten, että muodostettavien alisolmujen sovitevirhe minimoituu. Sama binaarinen jako suoritetaan tämän jälkeen rekursiivisesti jokaiselle alisolmulle. Alla olevassa kuvassa 2 on esimerkki regressiopuun rakenteesta, ja sen muodostuminen juurisolmusta (Solmu1) aina lehtiin asti. Jokaisen solmun jälkeen voidaan havaita ehtorakenne muuttujan $X_1 \dots X_3$ suhteen, jolla data-aineiston arvot jaetaan joko oikeaan tai vasempaan alisolmuun. /13/



Kuva 2. Esimerkki regressiopuun rakenteesta.

Regressiopuun jakoa jatketaan, kunnes solmujen jakaminen ei enää pienennä sovitevirheen arvoa, tai jokin käyttäjän määrittämistä lopetusehdoista saavutetaan. CART-menetelmä käyttää binaarijaon tekemiseen eri jakofunktiota riippuen siitä, muodostetaanko luokittelu- vai regressiopuu. Regressiopuussa käytettävä jakofunktio on pienimmän neliösumman menetelmä, jolla valitaan selittävä muuttuja ja sille jakoarvo, joka minimoi sovitevirheen. Jollei lopetusehdoja puun muodostamiselle ole määritetty, menetelmä jakaa data-aineistoa, kunnes jokainen lehti sisältää vain samoja suorituskyskyarvoja. Tästä muodostuu erittäin monimutkainen ja opetusdata-aineistoon ylisovitettu regressiopuu. Puun kasvava monimutkaisuus ei ole tae puun selityskysyvyn lisääntymisestä validointidatalle. /13/

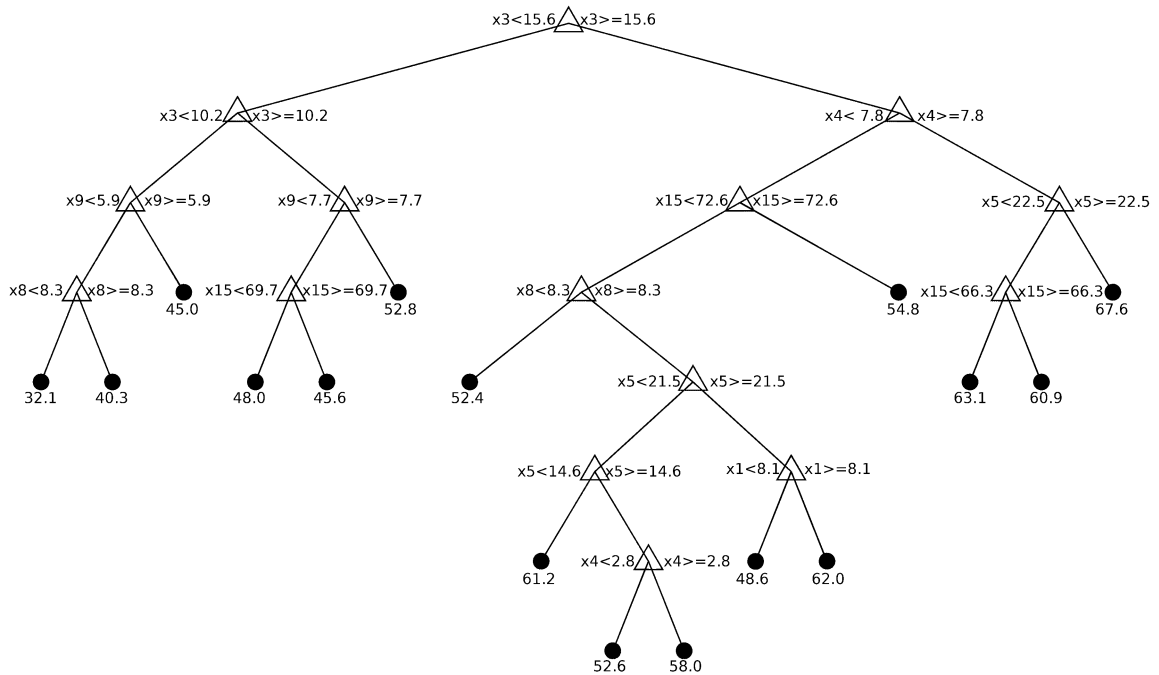
CART-menetelmää käytettäessä noudatetaan yleensä toimintatapaa, jossa ensin luodaan monimutkainen puurakenne. Tämän jälkeen monimutkaista puuta yksinkertaistetaan karsimalla alimpia lehtiä ja haaroja pois. Käytännön sovelluskohteissa tehdään aina kompromissi mallin monimutkaisuuden ja ennustustarkkuuden välillä. Data-analyysin tavoitteet ja sovelluskohteen erityispiirteet määrittävät, miten näitä ominaisuuksia painotetaan lopullista puurakennetta muodostettaessa. /13/

3 TULOKSET

CART-menetelmän soveltumista liikkuvien työkonoiden suorituskysyvyn suhteelliseen estimointiin arvioitiin koejärjestelyn avulla. Työssä käytetty data-aineisto on peräisin globaalien liikkuvia työkonetta valmistavan yrityksen tietokannoista. Data-aineisto rajattiin sekä maakohtaisesti että konetyyppikohtaisesti, jotta data olisi vertailukelpoista. Data-aineisto sisälsi mittausdataa vaihtelevissa olosuhteissa toimineilta liikkuvilta työkonelilta. Kolmetoista samankaltaista konetta valittiin data-aineistoksi, jolla CART-menetelmää testattiin. Mallin muodostamista varten valittiin 17 muuttujaa kuvaamaan toimintaympäristöä ja yksi suorituskyskyindikaattori kuvaamaan koneen tehokkuutta.

Data-aineistolle suoritettiin datan esikäsittelyä ennen CART-menetelmän hyödyntämistä. Aluksi data-aineisto koottiin yhdistämällä mittaustietoa eri tietokannoista. Tämän jälkeen data-aineistolle suoritettiin puuttuvien arvojen käsittely, virheellisen datan poisto sekä normalisointi. Data jaettiin lopuksi opetusdataan (75 %) ja validointidataan (25 %). Datapisteiden määrä opetusdatassa oli 233200.

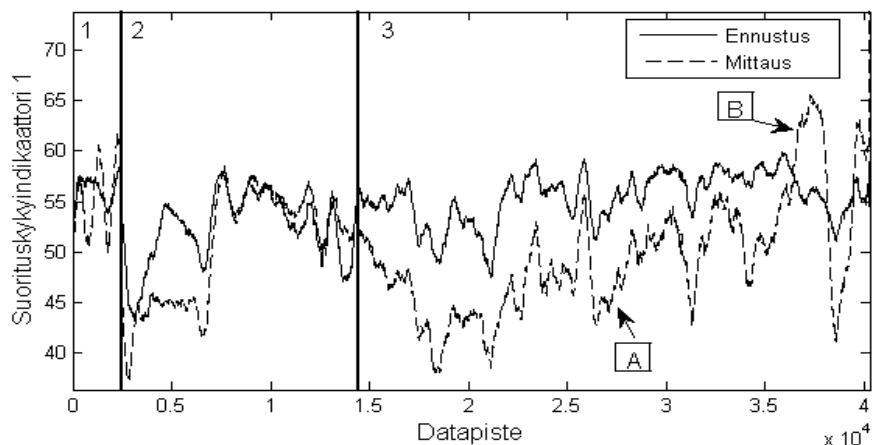
Regressiopuu muodostettiin esikäsittelyn opetusdatan perusteella MATLAB-ohjelmistolla CART-menetelmää hyödyntäen. Aluksi muodostettu monimutkainen regressiopuu karsittiin kuvassa 3 esitettyyn muotoon. Puun karsinta esitettyyn muotoon suoritettiin Breiman ym. esittämän menetelmän avulla /13/. Regressiopuulle valittu rakenne on kompromissi ennustuskysyvyn ja kompleksisuuden välillä.



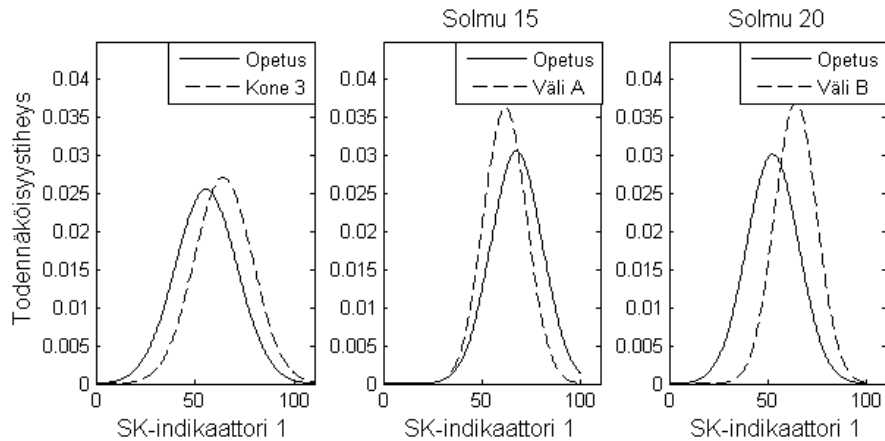
Kuva 3. Data-aineistosta CART-menetelmällä muodostettu regressiopuu.

Kuva 3 regressiopuun perusteella tärkeimmät suorituskykyindikaattoria selittävät toimintaympäristömuuttujat ovat x_3 , x_4 , x_5 , x_9 ja x_{15} . Vertailtaessa regressiopuun avulla identifioituja tärkeimpiä muuttujia datan keränneeltä taholta saadun asiantuntijätietämyksen kanssa, tulokset ovat yhtenevät. Kuten olettaa voi, malli ei jakanut data-aineistoa sellaisten muuttujien perusteella, joilla ei ole asiantuntijätietämyksen perusteella suurta vaikutusta tarkasteltavaan suorituskykyindikaattoriin. Puurakenteen lehdistä nähdään jokaiseen eriteltyyn toimintaympäristöön liittyvät keskimääräiset suorituskykyindikaattorin arvot. Kuten puun opetukseen käytetystä data-aineistosta huomataan, eri toimintaympäristöissä saavutetut keskimääräiset suorituskykyindikaattoreiden arvot vaihtelevat suuresti.

Kuvassa 4 on esitetty validointidatassa olevien kolmen koneen ennustettu ja mitattu suorituskykyindikaattorin arvo. Ennuste on saatu etsimällä jokaiselle validointidatan pisteelle vastaava lehti regressiopuusta ja ottamalla siitä talteen tyypillinen suorituskykyindikaattorin arvo. Kuva 4 yläreunassa olevat numerot 1, 2 ja 3 esittävät validointidatassa olevien koneiden tunnistetta. Kuvasta 4 voidaan havaita, miten validointidatan koneet ovat suoriutuneet verrattuna opetusdatan koneiden keskimääräiseen suorituskykyyn. Tarve suorituskyvyn arviointiin eri toimintaympäristöissä on selvästi havaittavissa Kuva 4 perusteella, sillä suorituskyvyn arvoa 55 voidaan pitää toimintaympäristön perusteella joko huonona, keskinkertaisena tai hyvänä.



Kuva 4. Regressiopuun muodostama ennuste ja mitatut suorituskykyindikaattorin arvot validointidatalle.



Kuva 5. Suorituskykyindikaattorin normaalijakauman sovitukset koneen 3 datalle ja opetusdatalle.

Kuvassa 5 esitetään normaalijakaumia sovitettuina opetusdataan ja koneen 3 suorituskykyindikaattorin mittauksiin eri tilanteissa. Kuvan 5 vasemmanpuoleisessa osiossa esitetään koko opetusdatan suorituskykyindikaattorin arvojen tiheysfunktio verrattuna Kuvan 4 kolmannen koneen suorituskykyindikaattorin tiheysfunktioon. Kuvassa 5 esitetään myös opetusdatan suorituskyvyn tiheysfunktiot verrattuna kolmannen koneen mittauksiin puun solmuissa 15 ja 20. Kuvan 5 perusteella voidaan todeta, että koneen 3 suorituskyky vaihtelee ympäristöstä riippuen. Suorituskyky on solmussa 15 huonompi ja solmussa 20 parempi verrattuna samankaltaisissa ympäristöissä työskennelleisiin opetusdatan koneisiin. Siten pelkän toimintaympäristöstä riippumattoman suorituskyvyn tarkasteleminen ei anna tarpeeksi informaatiota käyttäjälle koneen toiminnan parantamiseksi.

Opetusdatan kattaessa suuren joukon samankaltaisia liikkuvia työkoneita, käyttäjä voi arvioida koneensa suorituskykyä suuremmalla tarkkuudella ja luotettavuudella. Esitetty menetelmä mahdollistaa tilastotiedon hyödyntämisen päätöksentekoprosessissa. Syyt suorituskykyindikaattorin arvojen samankaltaisuuksiin ja eroavaisuuksiin voivat olla erinäisiä, mutta esitetyn menetelmän avulla eri tilanteet voidaan huomata. Eroja voitaneen selittää esimerkiksi työn tavoitteilla kuten tehokkuus- ja tuottavuusvaatimuksilla, jotka koneelle on asetettu. Tietoa työn tavoitteista ei ole ollut suoraan saatavissa tässä testissä.

4 JOHTOPÄÄTÖKSET JA JATKOTUTKIMUSEHDOTUKSET

Tässä työssä esiteltiin luokittelu- ja regressiopuu tilastollisen data-analyysin menetelmänä liikkuvien työkoneiden suorituskyvyn suhteelliseen arviointiin. Työn perusteella voidaan osoittaa, että CART-menetelmä on käyttökelpoinen menetelmä suhteellisten suorituskykyindikaattorien arviointiin. Menetelmä tarjoaa potentiaalisen mahdollisuuden tarkempaan ja luotettavampaan suhteellisen suorituskyvyn estimointiin.

Ehdotettu menetelmä on täysin datapohjainen, sillä erilaiset toimintaympäristöjen mittaukset ja niitä vastaavat suorituskykyindikaattorien arvot saadaan oikeilta työkoneilta. Tämä on yksi menetelmän eduista, sillä dataan perustuvassa menetelmässä voidaan hyödyntää käyttäjien asiantuntevuudesta, kokeiluista ja virheistä seuranneita suorituskykyjä eri toimintaympäristöissä. Menetelmän etuna on myös selittävien muuttujien automaattinen löytyminen ilman tarkkaa tietoa muuttujien vuorovaikutuksista. Näin ollen menetelmää voidaan käyttää myös tärkeiden muuttujien automaattiseen etsimiseen.

Tässä työssä keskityttiin rakentamaan regressiopuu vain työn kuormitukseen ja toimintaympäristöön viittaavien parametrien perusteella. Puun rakentamisessa voidaan kuitenkin jatkossa käyttää myös muista muuttujista saatua dataa. Yksi kiinnostava muuttujaryhmä on koneen ohjausparametrit, joiden asetuksilla tiedetään olevan vaikutusta suorituskykyindikaattoreihin. Kuten erilaiset toimintaympäristöt myös ohjausparametrien viritykset asettavat rajoja suorituskykyindikaattoreiden mahdollisille arvoille. Siten ohjausparametrien lisääminen ongelmaan lisää tarvetta suhteellisen suorituskyvyn vertailulle. Ohjausparametrien lisäämisellä voidaan saada puusta myös tarkempia ennustuksia suorituskyvyn arvoille.

Esitetty menetelmä on kirjoittajien ensimmäinen ratkaisuehdotus kuvattuun ongelmaan. Lisätutkimusta tarvitaan menetelmän luotettavuuden arviointiin liikkuvien työkoneiden alalla. Mahdollisia jatkotutkimusaiheita ovat: ”Mikä algoritmi valitaan regressiopuun rakentamiseen?” /16/, ”Paljonko dataa luotettavan puun rakentamiseen tarvitaan?”, ”Miten ohjausparametrien lisääminen malliin vaikuttaa tuloksiin?”, ”Onko viantunnistus mahdollista regressiopuun avulla?” ja ”Mikä on tärkein syy ennustetusta suorituskyvystä poikkeamiselle?”.

Kiitokset

Tutkimusta on rahoitettu Tekesin (D2I – Data to Intelligence) ja Suomen Akatemian (HOPE – Human Operator Modelling And Performance Evaluation In Human-machine Interaction) tutkimusohjelmilla.

Kirjallisuusluettelo

1. Hölttä V., Repo M., Palmroth L., Putkonen, A.: Index-based performance assessment and condition monitoring of a mobile working machine, Proceedings of the 2005 ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Long Beach, California, USA, September 24-28, 2005, 8 p.
2. Hölttä V.: Plant performance evaluation in complex industrial applications, Doctoral Thesis, Helsinki, Finland, 2009, 108 p.
3. Kokkarinen J.: Koneellinen puunkorjuu, book, Metsäteho Oy, Vantaa, 2013, 111 p.
4. Palmroth L., Tervo K., Putkonen A.: Intelligent coaching of mobile working machine operators, Intelligent Engineering Systems, 2009. INES 2009. International Conference on, IEEE Press, April 2009, pp. 149-154.
5. Macmillan R. H.: The mechanics of tractor-implement performance: theory and worked examples: a textbook for students and engineers, R. H. Macmillan, Chapter 5, 2002, 17 p.
6. Jokiniemi T., Rossner H., Ahokas J.: Simple and cost effective method for fuel consumption measurements of agricultural machinery, Agronomy Research Biosystemn Engineering Special Issue, 10(2012), pp. 97-107.
7. Park S. H., Kim Y. J., Im D. H., Kim C. K.: An assessment of eco driving system for agricultural tractor, Journal of Agricultural Science and Technology, 1(2011), pp. 906-912.
8. Inns F. M.: Selection, testing and evaluation of agricultural machines and equipment, FAO Agricultural Services Bulletin, 115(1995), pp. 32-38.
9. Ismail Z. E., Abdel-Mageed A. E.: Workability and machinery performance for wheat harvesting, Misr Journal of Agricultural Engineering, 27(2010), pp. 90-103.
10. Lu J. G., Liu Y., Li X.: The decision tree application in agricultural development, Artificial Intelligence and Computational Intelligence, 72(2011), pp. 372-379.
11. Waheed T., Bonnell R. B., Prasher S. O., Paulet E.: Measuring performance in precision agriculture: CART – a decision tree approach, Agricultural Water Management, 84(2006), pp. 173-185.
12. Li M., Feng S., Sethi I. K., Luciw J., Wagner K.: Mining production data with neural network & CART, Data Mining, 2003. ICDM 2003. Third IEEE International Conference on Data Mining, Melbourne, Florida, USA, November 19-22, 2004, pp. 731-734.
13. Breiman L., Friedman J., Stone C. J., Olshen R. A., Classification and regression trees, book, Wadsworth and Brooks, Monterey, CA, 1984, 368 p.
14. Bishop C. M.: Pattern recognition and machine learning, book, Springer Science+Business Media, New York, Usa, 2006, 738 p.
15. Kotsiantis S. B., Kanellopoulos D., Pintelas P. E.: Data preprocessing for supervised learning, International Journal of Computer Science, 1(2006)2, pp. 111-117.
16. Rokach L., Maimon O.: Top-down induction of decision trees classifiers – a survey, IEEE Transaction on Systems, Man and Cybernetics, Part C: Applications and Reviews, 35(2005)4, pp.476-487.