

Tartu Ülikool
Loodus- ja tehnoloogiateaduskond
Ökoloogia ja Maateaduste instituut
Geograafia osakond

Magistritöö inimgeograafias

**Inimese tegevuskohtade leidmine nutitelefonipõhiste
käitumisandmestike alusel**

Kaisa Vent

Juhendaja: prof Rein Ahas

Kaitsmisele lubatud:
Juhendaja:
Osakonna juhataja:

Tartu 2014

SISUKORD

Sissejuhatus.....	3
1. Teoreetilised lähtekohad	5
1.1 Inimese ruumiline käitumine.....	5
1.2 Ruumilise käitumise ajalised dimensioonid.....	6
1.3 Variatsioonid ruumilises käitumises	7
1.4 IKT levik ja ruumilise käitumise uued suunad.....	9
1.5 Tegevuskohad ja nende uurimine nutitelefoniga	12
1.5.1 Tegevuskohtade tuvastamise algoritmid GPS andmetes.....	12
1.5.2 Tegevuskohtade semantiliste aspektide leidmine.....	15
2. Andmed	17
2.1. YouSense andmete kogumine ja salvestamine.....	17
2.1.1 Asukohainfo kogumine	19
2.1.2 Telefonikasutusinfo kogumine.....	19
2.2 Tegevuskohtade annoteerimine.....	20
3. Tegevuskohtade tuvastamine	22
3.1 Tegevuskohad YouSense andmestikus	22
3.2 Tegevuskohtade tuvastamine naabruskonna tiheduse alusel.....	23
3.2.1 Parameetrite Eps ja MinPts valik.....	23
3.2.2 DBSCANi rakendamine	26
3.3. Tegevuskohtade tuvastamine “aukude” meetodil	27
4. Külastustunnuste lisamine tegevuskohtadele ja statistilise analüüsi meetodika	30
4.1 Külastuse alguse ja kestuse arvutamine	30
4.2 Mobiilikasutustunnuste sidumine külastustega.....	31
4.3 Külastustunnuste analüüs	32
5. Tulemused	33
5.1 Tegevuskohtade tuvastamine	33
5.2 Tegevuskohtade külastustunnused.....	35
5.2.1 Külastuspäeva tüüp kui iseloomulik tunnus	35
5.2.2 Külastuse algusaeg kui iseloomulik tunnus	37
5.2.3 Külastuse kestus kui iseloomulik tunnus	39
5.2.4 Mobiilikasutus kui iseloomulik tunnus.....	43
6. Arutelu ja järeldused	48
6.1 Tegevuskohtade tuvastamine	48
6.2 Külastustunnused ja nende iseloomulikkus tegevuskoha kirjeldamisel.....	50
Kokkuvõte	54
Summary	55
Tänuavaldused	56
Kasutatud kirjandus	57
Lisad	63

Sissejuhatus

Ruumilise käitumise ja selle erinevate aspektide käsitlemine pikemate perioodide vältel, kasutades traditsioonilisi meetodeid nagu reisipäevikud, on kallis ja aeganõudev (Nitsche et al. 2013). Seetõttu on alternatiivide leidmine sellistele meetoditele olnud päevakorras ajast, mil GPS tehnoloogiad muutusid odavamaks ning kättesaadavaks (Wolf et al. 2001). Info- ja kommunikatsioonitehnoloogiate (IKT), sh mobiilsete tehnoloogiate, laiaulatuslik levik ühiskonnas on teinud võimalikuks ruumilise käitumise uurimise kaasata uusi andmeid, mis võivad olla kogutud erinevatest sotsiaalvõrgustikest nagu Twitter (Birkin & Malleson 2012) ja Foursquare (Noulas et al. 2011) või pärinevad inimese igapäevaelu jälgivatest sensoritest (Petrenko et al. 2013). Nutitelefon on näide võrdlemisi odavast ning kättesaadavast tehnoloogiast (Higuera de Furtos & Castro 2014), mis sisaldab mitmeid sensoreid, võimaldades lisaks asukohainfole koguda teavet inimese käitumise ja teda ümbritseva keskkonna kohta (Do & Gatica-Perez 2013). Sellised mitmest allikast pärinevat informatsiooni sisaldavad andmed võimaldavad paremini mõista inimeste ruumilise käitumisega seonduvaid erinevaid tahke (Loseto et al. 2012).

Järjest enam nähakse põhjust uurida inimese ruumilist käitumist pikemate perioodide vältel (Järv et al. 2014), kuna muutuv elurütm ja potentsiaalne IKT mõju on kujundamas ümber inimese reisikäitumist (Line et al. 2011). Traditsioonilistel meetoditel põhinevad uuringud on aga sageli respondendile koormavad ning võivad sisaldada “respondendi väsimusest” tulenevaid vigu (Golob & Meurs 1986). Mitmed uuringud on seetõttu tähelepanu koondanud automaatselt kogutavatele andmetele ning nende töötlemisele ilma otsese respondendipoolse sisendita. Näited hõlmavad transpordivahendi automaatset tuvastamist (Bolbol et al. 2012) või tegevuskohale semantiliste aspektide lisamist (Liu et al. 2013). Selliste klassifikatsioonimudelite koostamine põhineb sageli aga tunnuste tunnetuslikul valikul, vaatlemata lähemalt nende statistilist tausta (Bolbol et al. 2012).

Käesoleva töö raames vaadeldakse Androidi-põhise YouSense rakenduse poolt kogutud käitumisandmestikku, mis lisaks asukohainfole (GPS) sisaldab teavet ka inimese telefonikasutuse kohta. Töö eesmärk on luua selliste andmetega töötamiseks metodoloogiline alus ning hinnata statistiliste vahenditega kogutud andmestiku kasutuspotentsiaali tegevuskoha klassifikatsioonimudelites. Töös on sõnastatud järgmised alameesmärgid:

1. Kujundada meetod tegevuskohtade tuvastamiseks YouSense asukohaandmestikust.
2. Luua vahendid külastustunnuste tuvastamiseks ja nende sidumiseks tegevuskohtadega.
3. Pilootuuringu raames testida 18 inimese 6 kuu andmetel tegevuskohtade leidmise meetodikat ja tuvastada külastustunnuste tegevuskoha-põhiste erinevuste esinemine.

Töö teoreetiline raamistik põhineb üheltpoolt reisikäitumise uuringutel (Buliung et al. 2008, Kang et al. 2010, Kenyon & Lyons 2007, Lenz & Nobis 2007), võimaldades hinnata inimese igapäevaste tegevuskohtadega seotud olulisi külastustunnuseid. Teisalt tugineb töö ka uuringutele, mis käsitlevad GPS ja nutitelefonide andmete töötlemist ja kasutamist (Nurmi 2009, Liu et al. 2013, Nitsche et al. 2013, Stopher et al. 2006, Wan & Lin 2013). Töö metoodilises osas luuakse alused tegevuskohtade tuvastamiseks ja külastustunnuste leidmiseks YouSense andmestikust. Loodud metoodilisi võtteid kasutatakse 18 inimese 6 kuu tegevuskohtade ja nendega seotud külastustunnuste tuvastamiseks, hindamaks pilootuuringu raames vaadeldud külastustunnuste sobilikkust klassifikatsioonimudelitesse kaasamisel. Külastustunnustena käsitletakse külastuste kestust, algusaega ning nädalapäeva tüüpi. Samuti vaadeldakse külastustunnustena mobiilikasutusega seonduvaid näitajaid nagu ekraani sisselülitamise ja lahtilukustamise sagedus ning aktiivse mobiilikasutuse kestus tegevuskohas.

Tööl on neli uudset aspekti: 1) vaadeldav kuue kuu asukohainfo on kogutud inimeste, mitte transpordivahendite liikumise tulemusena; 2) külastustunnuste tegevuskohapõhiseid erinevusi vaadeldakse pikema perioodi vältel; 3) külastustunnuste variatiivsust hinnatakse kõikide peamiste tegevuskohtade lõikes; 4) mobiilikasutuslike tunnuste variatsioone hinnatakse esmakordselt erinevate tegevuskohtade lõikes.

Töö metoodiliste eesmärkide edukal täitmisel võiksid tulemused olla kasulikud edasiste YouSense andmeid kasutavate uuringute jaoks. Töö sisulise poole tulemused võimaldavad aga kujundada alused YouSense andmete kasutamiseks tegevuskoha semantiliste aspektide automaatseks tuvastamiseks.

Magistritöö koosneb kuuest osast. Esimeses, teoreetilises osas, antakse ülevaade inimese ruumilise käitumise uurimise erinevatest tahkudest, sealhulgas on teoreetiline osa juhiseks tegevuskoha külastustunnuste valikul. Teoreetilisele osale järgneb peatükk töös kasutatavatest andmetest. Kolmas ja neljas peatükk käsitlevad tegevuskohtade ja külastustunnuste tuvastamise protsessi, viies ja kuues osa sisaldavad töö tulemusi ning arutelu.

1. Teoreetilised lähtekohad

1.1 Inimese ruumiline käitumine

Inimese aeg-ruumilise käitumise uurimine kiirenes oluliselt 20. sajandi teisel poolel, mil bihevioristid ning ajageograafid sellele oma tähelepanu suunasid. Paljuski mõjutas tähelepanu koondumist inimeste tegevustele ja nendega seotud tegevuskohtadele kiire linnastumine ja transpordi areng. Üheks esimeseks, kes püüdis omavahel ühendada aja ja ruumi ühtsesse geograafilisse raamistikku, oli Torsten Hägerstrand (1952), kelle loodud ajageograafia on suures osas teoreetiliseks aluseks tänastele reisikäitumise uuringutele.

Vastavalt rõhuasetusele võib inimese ruumilise käitumise käsitlemise jagada kahte suuremasse rühma: tegeliku ruumilise käitumise ning potentsiaalse ruumilise käitumise käsitlused (Buliung et al. 2008). Horton ja Reynolds (1971) tutvustasid inimese tegeliku ruumilise käitumise uuringute valdkonnale tegevusruumi (*activity space*) ja tegutsemisruumi kontseptsiooni (*action space*), mis võimaldavad üheltpoolt hinnata tegevuskohti, millega inimesed vahetult kokku puutuvad, teisalt aga annavad võimaluse vaadelda ka ruumi seda osa, mille olemasolu kohta inimesel eksisteerivad teatud teadmised. Hägerstrandi ajageograafia (1952) laiendab aga potentsiaalse ruumikasutuse käsitlusi, kontseptualiseerides inimeste tegevusruumi eluraja (*life path*) mõistega. Elurada on aegruumiline trajektoor, mis saab alguse sünniga ning lõpeb surmaga ja selle kuju on määratletud erinevate piirangutega (*constraints*). Ajageograafia sõnastab kolme tüüpi piirangud, mis limiteerivad või võimaldavad teatud tegevuste sooritamist. Need on võimekuse (*capability*), ühendatavuse (*coupling*) ja võimu (*authority*) piirangud. Eelnimetatu käsitleb inimese bioloogilisi võimeid, võimalusi ning vajadusi tegevusi omavahel ühildada, samuti kultuuri ja ühiskonna poolt seatud piiranguid, mis limiteerivad ajas ja ruumis juurdepääsu teatud tegevustele ja tegevuskohtadele. Sel moel on aeg ja ruum teineteisest lahutamatud ning tegevuste sooritamine või mitte-sooritamine teatud kohas piiritletud kui potentsiaalne võimalus (Buliung et al. 2008). Seega oli Hägerstrand ajageograafia printsiipide loomisega täiendanud juba olemasolevat tegevuste-põhist käsitlust aeg-ruumi sidususe ja lahutamatus kontseptsiooniga.

Dijst (1999) on inimeste tegutsemisruumi (*action space*) jaganud kolmeks tinglikuks osaks, millest tajutav tegutsemisruum (*perceptual action space*) on kõige suurem, hõlmates seda osa, mida inimene on võimeline oma teadmistele tuginedes hoomama ning tegelik tegutsemisruum (*actual action space*) sisaldab endas kõiki neid tegevuskohti, millega inimesel on olnud vahetu kokkupuude. Võimaliku tegutsemisruumi (*potential action space*) moodustab aga ruumi see osa, mis on inimesele kättesaadav, lähtudes erinevatest ajalistest piirangutest, mis inimese igapäevaelus ilmnevad.

Golledge ja Stimson (1997) seevastu lähtuvad tegutsemisruumi käsitlemisel kahest aspektist: indiviidide liikumine ning suhtlemine ruumis (*communication over space*). Indiviidide liikumist puudutav tegutsemisruumi osa on vaadeldav igapäevase tegevusruumina (*activity space*), mis seob neid ümbritseva keskkonnaga. Tegevusruum ise koosneb aga kolmest osast: 1) liikumised kodus ja selle lähiümbruses; 2) liikumised regulaarselt külastatavatesse tegevuskohtadesse nagu töö ja pood; 3) liikumised viimatimainitud tegevuskohtades ja nende lähiümbruses. Golledge ja Stimson rõhutavad, et tegevusruumi ajalis-ruumilised aspektid on seotud valikuga osa võtta ühest või teisest tegevusest.

1.2 Ruumilise käitumise ajalised dimensioonid

Ajadimensiooni käsitlemine on oluline osa inimeste ruumilise käitumise uurimisest (Pinjari & Bhat 2010). Bhat ja Koppelman (1999) rõhutavad aja kui entiteedi tähtsust, kuivõrd “inimesel on 24 tundi päevas, mida tuleb jagada erinevate tegevuste ja reisimise vahel sõltuvalt ajakavast, sotsiaal-demograafilistest, asukoha- ning konteksti-põhistest piirangutest”. Samuti on aeg olulisem printsiip tegevuste organiseerimisel kui seda on ruum (Kurani & Lee-Gosselin 1996). Seetõttu saab ruumilise käitumise aspektist oluliseks teadmine, kuidas ja millal inimesed erinevaid tegevusi ajas jaotavad (Meloni et al. 2004).

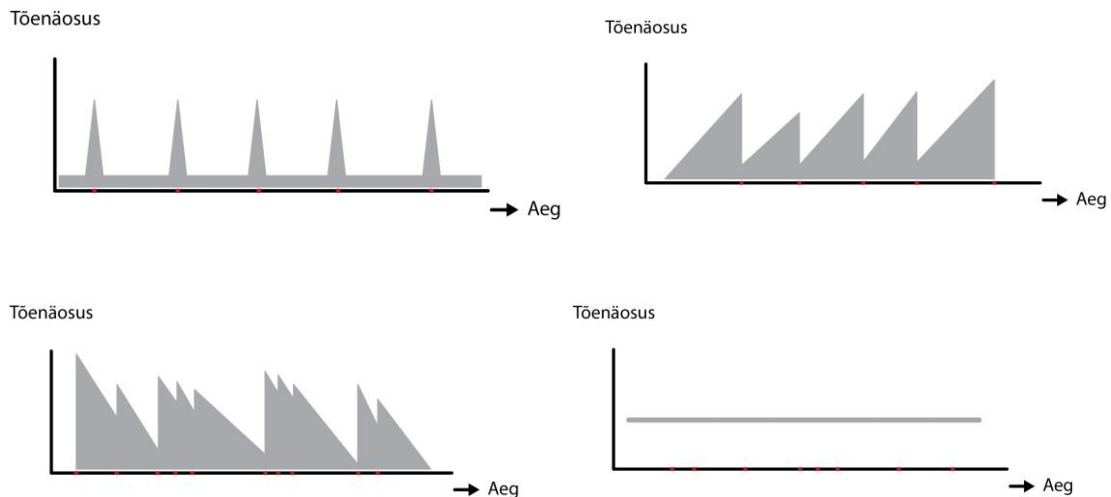
Ajadimensiooni uurimise võib jagada kaheks suuremaks haruks: aja jaotamine tegevuste vahel ehk ajakasutus (*time allocation*) ning tegevuste planeerimine ajalises raamistikus (Pinjari & Bhat 2010). See, kuidas indiviid oma tegevused ja reisirid planeerib, sõltub suuresti sotsiaalsetest rütmidest, linnakeskkonnast, olemasolevast transpordisüsteemist ning sotsiaal-demograafiast (Raux et al. 2011). Inimese valikuid aja optimaalse jaotamise osas vaadeldakse sageli jaotatuna kolmeks erinevaks tüübiks (Bhat & Koppelman 1993): elatusega seonduvad tegevused (*subsistence*), majapidamist puudutavad toimetused (*maintenance*) ja vaba aeg (*discretionary*). Bhat ja Koppelman (1993) püüavad siduda ajadimensiooni käsitlemise ruumiliste aspektidega läbi ajakasutuse analüüsi (*time-use analysis*) praktika. Ajakasutuse analüüsi käigus vaadeldakse tegevusi kategoriseerituna eelnimetatud gruppidesse, jagades need vastavalt kas kodus või kodust väljas toimuvateks. Samuti eristatakse argipäevi ning nädalavahetusi. Valik, kas tegevus toimub kodus või väljaspool kodu, on otseselt seotud ruumilise käitumisega ning on vajalik teadmine transpordiplaneerimisega seotud otsustusprotsessides.

Ruumikasutuse ajalist dimensiooni võib vaadelda sageduse ja regulaarsusena, millega inimesed teatud tegevuskohti külastavad (Golledge & Stimson 1997). Schönfelder on Campbelli (1970) järgi jaganud reisi eesmärgid ehk tegevused nelja kategooriasse:

- tegevused, mis on ajas regulaarsed
- tegevused, mille sooritamise vajadus kasvab ajas
- hasarti tekitavad tegevused (*time-contagious*)

- juhuslikud tegevused

Joonisel 1 on kujutatud kohandatud versioon Golledge'i ja Stimsoni (1997) tööst selliste tegevuste jaoks tehtavate reisi sooritamise tõenäosus kohta ajas. Ajas regulaarsed tegevused on näiteks seotud regulaarse harrastusega või tööl käimisega. Sellisel juhul sooritatakse reis suure tõenäosusega väga kindlal ajal, mis muudab kõikidel muudel aegadel reisi sooritamise tõenäosuse omakorda madalaks. Poes käik on näiteks tegevusest, mille vajadus ajas kasvab ning kahaneb vahetult pärast tegevuse sooritamist. Selliseid tegevusi iseloomustab ajas võrdlemisi ühtlane jaotumine. Hasarti tekitavad tegevused on inimestes esile kutsunud vajaduse üsna pea antud tegevusega uuesti tegeleda. Need tegevused on enamasti ajas üsna tihedalt koos ning nende sooritamise tõenäosus kahaneb ajaga. Viimane grupp, kuhu kuuluvad juhuslikud tegevused nagu ootamatu arstilkäik, ei oma tegevuse sooritamise seotud tõenäosusfunktsiooni, mis võimaldaks ennustada, millal toimub uus sarnane tegevus.



Joonis 1. Erinevat tüüpi tegevuste toimumine ajas ja nende sooritamise tõenäosus (punane märgib tegevuse toimumist). Ülevalt vasakult: ajas regulaarsed tegevused; tegevused, mille sooritamise vajadus ajas kasvab; hasarti tekitavad tegevused; juhuslikud tegevused (Golledge & Stimson 1997).

Ajadimensiooni vaatlemist läbi erinevatele tegevustele kulutatud aja võib vaadelda ka utilitaristliku resursside paigutamise teooria raamistikus (Yamamoto & Kitamura 1999). See tähendab, et indiviid üritab investeerida aega tegevustesse enda jaoks võimalikult suure kasuga. Olemasolev transpordipoliitika ning erinevate võimaluste kättesaadavus võib inimeste ajakasutust sellises perspektiivis mõjutada. Vastavalt olemasolevate võimaluste sobivusele ja olemasolule tehakse valik tegevuses osalemise kasulikkuse kohta.

1.3 Variatsioonid ruumilises käitumises

Ruumilise käitumise uurimine on pikalt peatunud vaid lühikeste perioodide vaatlemisel (Spissu et al. 2009). Põhjused selleks tulenevad esiteks

transpordiplaneerimise fookusest, mis on seni keskendunud pigem üksikute päevade vaatlemisele. Teiseks on pikemate perioodide uurimisel traditsiooniliste meetoditega probleemiks respondentidele langev suur koormus, mille tulemusena vähenevad valimi mahud (Stopher & Greaves 2007) ning andmestiku täpsus (Pearson 2004, Wolf et al. 2003), samuti suureneb mittevastamiste osakaal (Wilson 2004). Pikemate perioodide vaatlemise olulisust eeskätt reisikäitumise kontekstis on aga rõhutatud juba mitu aastakümnet (Jones & Clarke 1988, Koppelman & Pas 1984, Schlich & Axhausen 2003, Tarigan et al. 2012), sealjuures saab ajalise variatiivsuse vaatlemine eeskätt oluliseks inimese enda ruumilise käitumise kontekstis. Pas (1987) leidis, et kuni 50% koguvariatiivsusest ruumilises käitumises on seletatav just muutustega inimese reisikäitumises erinevate kuude ja nädalapäevade lõikes.

Schlich ja Axhausen (2003) analüüsisid 6-nädala jooksul kogutud reisikäitumuslikku infot ning jõudsid järeldusele, et vaadeldud perioodi jooksul esines suurem variatiivsus nädalavahetustel ning väiksemad erinevused reisikäitumises kerkisid esile tööpäeviti. Sarnaste tulemusteni, küll erineva pikkusega ajaperioode käsitledes, jõudsid ka Roorda ja Ruiz (2008) ning Kamruzzaman ja Hine (2012). Buliung ja teised (2008) vaatlesid seitsme järjestikuse päeva jooksul muutusi inimeste ruumikasutuses, millest selgus, et suurem osa tegevusi leiab aset tegevuskohtades, mida kasutatakse korduvalt. Sarnased tulemused ilmnisid ka *MobiDrive* kuue nädalasest uuringust (Susilo & Axhausen 2007), mis keskendus ruumikasutuse stabiilsuse uurimisele. Töö, kodu ja kool on kohad, mida iseloomustab suurem ruumiline stabiilsus, võrreldes vaba aja ja isikliku asjaajamisega seotud tegevuskohtadega.

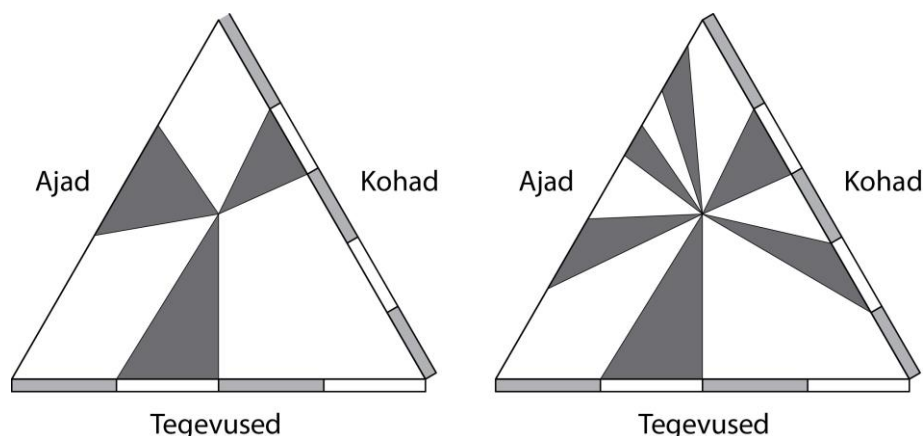
Lisaks ruumiliste aspektide vaatlemisele nagu reiside pikkus, kaugus kodust ning tegevuskohtade korduvkülastamine, on vaadeldud ka muutusi ajakasutuses erinevate nädalapäevade lõikes. Kang ja Scott (2010) kategoriseerisid väljaspool kodu toimuvad tegevused majapidamise hooldusega ning vaba aja tegevustega seonduvateks, millest eraldi eristati tegevused, mis viidi läbi üksi või kellegiga koos olles. Selgus, et nädalapäevade erinevad ajakasutuse mõttes nädalavahetustest märgatavalt. Samuti ei esine nädala sees "tüüpilist" päeva, mis kannaks endas iseloomulikke argipäeva ajakasutuslikke tunnuseid. Päevade vahelisi erinevusi ajakasutuses on uurinud ka Bhat ja Misra (1999) ning Yamamoto ja Kitamura (1999), kes käsitlesid vaba aja tegevustele kulutatava aja erinevusi argipäevade ja nädalavahetuse vahel. Argipäeviti veetsid uuritavad oluliselt rohkem aega väljaspool kodu, tegeledes vaba aja tegevustega. Nädalavahetusi seevastu iseloomustas koduga seotud vaba aja tegevuste rohkus.

GPS-tehnoloogiate kasutamine pikemate perioodide vältel inimese ruumilist käitumist puudutavates uuringutes on enam tähelepanu koguma hakanud seoses soodsate ning kättesaadavate vahendite laiaulatuslikuma levikuga. Üks esimesi selliseid longituuduuringuid viidi läbi Atlantas *Commute Atlanta* projekti raames, kasutades andmete kogumiseks isikliku sõiduautoga ühendatud GPS seadet (Schönfelder et al.

2006). Uuringu tulemused viitasid reisikauguse sesoonsele erinevusele, lähtudes kodust – võrreldes sügise ja talvega, on reisirid kevadel ja suvel enam ruumis hajutatud ning rohkem esineb reise, kus tegevuskoht asub kodust märgatavalt kaugemal (Järv et al. 2014 cit. Schönfelder & Axhausen 2010, p. 156). Stopher ja teised (2006) kinnitasid 28-päevase GPS uuringu põhjal inimkeskse variatiivsuse esinemist ning sellega arvestamise olulisust transpordikäitumise mudelite koostamisel. Samuti käsitlesid nad GPS uuringu sobiliku kestuse temaatikat ning leidsid, et eraldi kaasaskantavate GPS seadmete kasutamine uuringus võib pikema perioodi jooksul saada respondentidele koormavaks. Mobiiltelefonide kasutamine omab selles osas silmapaistvat perspektiivi, kuna inimesed kannavad telefone endaga kaasas pea kõikjal. Probleemne koht nutitefonide kasutamisel sellistes uuringutes peitub andmekogumise eest vastutavate rakenduste suures energiatarbes, mis vähendab aku kestvust märgatavalt.

1.4 IKT levik ja ruumilise käitumise uued suunad

Couclelis (2009) toob välja kaks olemuslikku aspekti, mida IKT laialdane levik ühiskonnas endaga kaasa toob: nõrgenenud assotsiatiivsed suhted tegevuste ja kohtade ning aja ja kohtade vahel. See tähendab sisuliselt suurenenud võimalusi sooritada tegevusi distantsilt – “*to act at a distance*” (Schwanen & Kwan 2008), samal ajal minimeerides tegevuste “vahetamiseks” kuluvat aega (Couclelis 2009). Eelpool nimetatut arvesse võttes pani Couclelis (2000) ette võtta kasutusele tegevuste fragmenteerituse mõiste (“*fragmentation of activity*”). “Fragmenteerumine on protsess, mille käigus kindel tegevus jagatakse mitmeks väiksemaks osaks, ning nende osade teostus toimub erineval ajal ja/või erinevas kohas,” (Couclelis 2003). Tegevuste fragmenteerumine lisab tegevuskohtade ja reisirikäitumise uurimisele uue dimensiooni, kus muutunud on erinevate tegevuskohtade vahelise kauguse osatähtsus, aga ka tegevuskohtade endi funktsioon (Kwan 2007). Joonisel 2 on esitatud tegevused enne (paremal) ja pärast (vasakul) IKT laialdasemat levikut – näiteks poed ning paljud ametiasutused võimaldavad kasutada teenuseid veebipõhiselt, mis võimaldab inimesel lisada oma füüsilise asukoha tavapärasele funktsioonile mitmeid lisafunktsioone, jagades sel moel tegevuste sooritamise mitmeteks väiksemateks osadeks. Sel moel on näiteks kodus võimalik tegeleda ostlemise ning asjaajamisega, liites seeläbi kodule mitmeid uusi tahke. Jooniselt ilmneb ka fragmenteerumise kaks peamist tüüpi: ruumiline ja ajaline (Alexander et al. 2011).



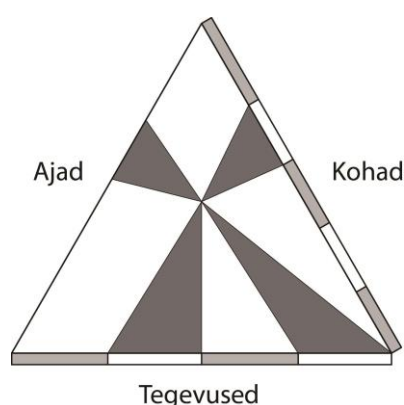
Joonis 2. Tegevused enne (vasakul) ja pärast (paremal) IKT laialdasemat levikut. Parempoolsel joonisel on näha ühe konkreetse tegevuse fragmenteerumine ajas ja ruumis (Couclelis 2009).

Kuivõrd Couclelise (2000) esialgne käsitus tegevuste fragmenteerumisest ei laiendanud teooriat empiirilisse sfääri, tekkis kiiresti selleks siiski vajadus – kohapõhised maakasutuse ja transpordiplaneerimise mudelid kaotasid valiidsust üha enam inimesekeskseks muutuvast ühiskonnas (Alexander et al. 2011). Alexander ja teised (2011) toovad välja fragmenteeritud tegevuste mõju avaldumise kohtade muutuvast funktsioonis, reisi aja kasuteguri tõusus ning mitmekesisistunud tegevustes. Ühtlasi toonitavad Alexander ja teised (2011), et selline fragmenteerumine võib viia sagenenud reisideni erinevate tegevuskohtade vahel, Schwanen ja Kwan (2008) püstitavad aga hüpoteesi, et tegevuste fragmenteeritus võib kokku langeda ka suurema tegevusruumiga. Kwan (2007) rõhutab omakorda IKT mõju olulisust pereelule ning sotsiaalsetele suhetele, mis on vihjeks muutuvast reisikäitumisest.

Lenz ja Nobis (2007) leiavad, et tegevuste fragmenteerumise uurimisel on oluline, kuidas muuta teoorias käsitletu praktikas arusaadavaks ja rakendatavaks. Nad toovad välja, et küsimused “*Millal jaguneb tegevus osadeks?*” ning “*Millal muutub see reisikäitumise kontekstis oluliseks?*” moodustavad nimetatud empiirika baasi. Eraldi on rõhutatud ka pikema ajaperioodi olulisust fragmenteerumise uurimisel. Samas ei ole seni veel pikemat perioodi tegevuste fragmenteerumisel käsitletud – enamasti on uurimisel kasutatud päeviku meetodit, kus uuringuperiood varieerub ühe-kahe päeva vahel (Lenz & Nobis 2007, Alexander et al. 2011, Ren & Kwan 2007) ning uuringute skoop on varieerunud vaid töökoha fragmenteerituse uurimisest (Alexander et al. 2011, Lenz & Nobis 2007) kuni Interneti kasutamiseni erineval kellaajal erinevates tegevuskohtades (Ren & Kwan 2007). IKT-d, mis kõnealustes uuringutes käsitluse all olid, sisaldasid enamasti lauaarvutit, laptopi või PDA-d (*personal digital assistant*). Vaid Lenz ja Nobis (2007) käsitlesid ka mobiiltelefoni kasutamist kõige muu hulgas. Seega on hetkel katmata pikemad ajaperioodid kui kaks päeva ning mobiiltelefoni kasutuse põhjalikum käsitus. Võttes arvesse, et mobiiltelefon on üks enim kasutatavaid IKT-sid (Kwan 2007), on see võrdlemisi oluline puudujääk.

Tegevuste fragmenteerumisega tihedalt seotud mõiste on *multitasking*. Kenyon ja Lyons (2007) defineerivad seda kui “kahe või enama tegevuse samaaegset läbiviimist

antud ajaperioodil”. Joonisel 3 on juba eelnevalt selgitatud Couclelis’ (2009) joonise (joonis 2) kohandatud versioon *multitasking* jaoks. Võib väita, et tegevuste fragmenteerumine ning *multitasking* on mõisted, mis sisalduvad teineteises või on ühe nähtuse vaatlemine erineval skaalal ja erineva rõhuasetusega. Teisisõnu võib öelda, et fragmenteerumine seab fookuse ühe tegevuse sooritamisele erinevates kohtades või erineval ajal, *multitasking* aga samal ajal ja samas kohas sooritatud tegevustele. Illustreerimaks eeltoodut: töötamine töökohas ja õhtuti kodus töökohustustega seotud e-kirjadele vastamine *versus* söögitegemise ajal tööalastele e-kirjadele vastamine. Seega võib *multitaskingut* näha kui võimalust efektiivsemalt ära kasutada olemasolevat 24 tundi.



Joonis 3. *Multitasking* kui kahe või enama tegevuse sooritamine samal ajal ja samas kohas (Couclelis 2009, autori kohandus).

IKT-de levikuga seoses on hakatud enam pöörama tähelepanu ka tegevuse ja koha vahelistele seostele. Kui klassikaline ajageograafia jagab tegevused fikseerituteks, kus oluliseks saab piirang tegevuse asukohale ja ajale, millal tegevus sooritatakse, ning paindlikemaks (Neutens et al. 2010), siis IKT levikuga on hakanud need piirid hägustuma – kodu võib seostada ostlemisega ning tööd koduse elu korraldamisega. Seetõttu on hakatud enam tähelepanu pöörama kohale ja sellega seotud tegevustele ka läbi IKT dimensiooni, kuivõrd *tegevuskoht=tegevus* seosed ei ole enam nii üheselt mõistetavad. Teisalt ei ole põhjust eeldada, et IKT mõju on igalpool ühesuunaline ja kõikehõlmav, vaid see mõju on sõltuv tegevuse tüübist ja tegevuskohast, inimesest endast, sotsiaalsest, kultuurilisest ning füüsilisest kontekstist (Schwanen & Kwan 2008, Valentine & Holloway 2002). Sellest tulenevalt võib eeldada, et tegevuskohtade lõikes esinevad erinevused IKT-de kasutamises, mis tulenevad nende kohtadega seotud tavadest, sotsiaalsetest normidest ning piirangutest. Millised on aga need kohad, kus rakenduvad kõnealused piirangud ja normid kõige esinduslikumalt, selle kohta puuduvad hetkel põhjalikumad uuringud. Järjest enam uuringuid viiakse läbi, pidades silmas tegevusi, kuid seosed koha ja IKT vahel on seni jäänud suuresti seletamata. Ren ja Kwan (2007) on küll visualisatsioonide näol laiendanud tegevuste-põhist lähenemist kohapõhisele ning kõrvutanud traditsioonilise

2D ruumi-aja teekonna “küberruumi” vastava teekonnaga, kuid selline näide ei ole prevaleeriv.

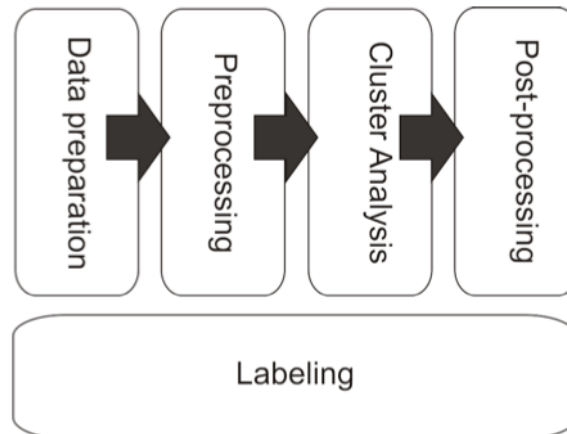
1.5 Tegevuskohad ja nende uurimine nutitelefonide andmestikuga

Reisikäitumise uuringud kujunesid ajalooliselt küsitlusuuringute-põhisel andmekogumise meetodikal (Golledge & Stimson 1997). Niisuguste andmete puhul on võimalik inimestelt saada vahetu informatsioon tegevuskohtade ja tegevuste iseloomu kohta. Uute, automaatsete andmekogumisvahendite (GPS, nutitelefonid ja erinevad sensorid) populaarsuse kasvades on aga andmed muutunud koheselt kättesaadava informatsiooni poolest vaesemaks (Wolf et al. 2001), tingides vajaduse meetodite järele, mis võimaldaksid vajalikku informatsiooni andmetest eraldada.

1.5.1 Tegevuskohtade tuvastamise algoritmid GPS andmetes

GPS-põhine asukohainfo kogumine geograafilistes uuringutes on seoses soodsate ja kättesaadavate vahendite esilekerkimisega saanud viimasel kümnendil märkimisväärset tähelepanu. Palju räägitakse GPS-põhiste andmete eelistest traditsiooniliste meetodite ees nagu reisipäevikud (Wolf et al. 2001, Chen et al. 2010) ning võimalustest, mida pakuvad asukohainfot koguda võimaldavad vahendid nagu nutitelefonid (Raento et al. 2009, Do & Gatica-Perez 2010, Nitche et al. 2013). Üks esimesi asukohainfo automaatse kogumise kasutusvõimalusi käsitlevatest uuringutest reisikäitumise valdkonnas viidi läbi Wolf ‘i ja teiste poolt (2001). Nende töö oli omal ajal unikaalne, kuivõrd võttis eesmärgiks GPS seadme poolt kogutud infost saada automatiseeritud andmetöötluse teel kätte võimalikult suur hulk kasulikke teavet inimese reisikäitumise kohta (Wolf et al. 2001).

Tegevuskohtade tuvastamist andmestikust saab vaadelda kui klasteranalüüsi teostamise ülesannet, kusjuures potentsiaalsed tegevuskohad eristuvad ühe või teise tunnuse poolest ümbritsevast foonist (Nurmi 2009, Thierry et al. 2013) ning omavad inimesele teatud tähendust. Protsess hõlmab endas mitut sammu, mis on kirjeldatud joonisel 4. Esimesed kaks sammu, andmete ettevalmistamine ja eeltöötlus, sisaldavad andmete viimist sobivale kujule ja võimalike vigade eemaldamist andmestikust. Klasteranalüüs on protsessi üks põhietappe, mille tulemusi hiljem viimistletakse järeltöötluse etapis, eesmärgiga eemaldada kõik sellised klastrid, millel puudub tähenduslik aspekt (Nurmi 2009).



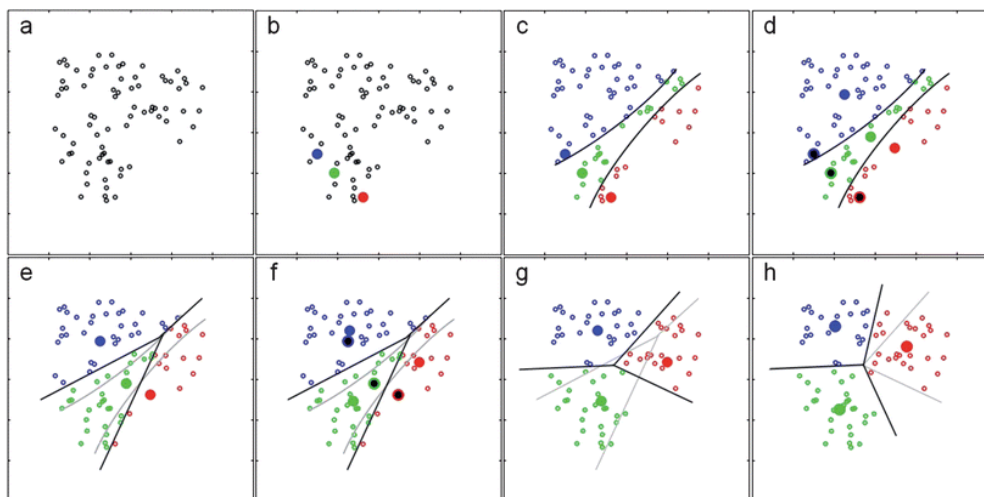
Joonis 4. Tegevuskohtade tuvastamise protsess GPS andmestikust (Nurmi 2009).

Erinevaid klasteranalüüsi variante on mitmeid ning nende klassifikatsioon suurematesse tüüpühmadesse varieerub kirjanduses mõningal määral. Zhou ja teised (2007) jagavad meetodid kolme suuremasse rühma, käsitledes eraldi ajaparaameetriga klasteranalüüsi (*time-based clustering*), tihedus-põhist klasteranalüüsi (*density-based*) ja etteantud klastrite arvuga klasterdamist (*partitioning clustering*).

Ajaparaameetril põhinevad meetodid vaatlevad kahe punkti vahelist aega ja vahemaad (määratletud raadius) ning määravad punktid klastrisse kuuluvaks, kui vahemaa on väiksem etteantud raadiusest ning kahe punkti vaheline aeg ei ületa teatud piirmäära (Marmasse & Schmandt 2002). Selliste meetodite puudus seisneb aga väheses üldistusvõimes, mis tähendab seda, et algoritm on enamasti andmestikuspetsiifiline (Nurmi 2009). Samas on tegemist väga lihtsate meetoditega, mille arvutuskeerukus ei saa takistuseks suuremate andmete puhul. Ajaparaameetriga klasterdamist võib kohata töödes, kus GPS andmete edastamise tekkinud pausi (signaali katkemist) kasutatakse selleks, et teha kindlaks inimese viibimine hoones (Ashbrook & Starner 2003, Marmasse & Schmandt 2002). Kang ja teised (2004) soovust töötasid välja ajal põhineva klasteranalüüsi meetodi, et eraldada tegevuskohad *Wi-Fi* võrgul põhinevatest positsioneerimisandmetest.

Etteantud klastrite arvuga meetodid on võrdlemisi robustsed tegevuskohtade tuvastamisel, kuivõrd eeldavad kasutajalt teadmist selle kohta, palju tegevuskohti andmestikus on. Meetod määrab esilagses andmestikus juhuslikult klastrite keskpunkti koordinaadid ning arvutab kõikide punktide jaoks kauguse määratud keskpunktidest. Punkt määratakse ühte klastrisse keskpunktiga, mis on talle kõige lähemal, seejärel arvutatakse uus keskpunkt kõikide klastrisse kuuluvate punktide põhjal. Protsessi korratakse seni, kuni keskmine enam oluliselt ei muutu (Ashbrook & Starner 2003, joonis 5). Meetodi suurimad puudused on seotud mürapunktide esinemisega andmestikus, mis võivad arvutatava keskmise väärtust oluliselt muuta, mõjutades hilisemat tegevuskoha asukoha täpsust (Zhou et al. 2007). Ashbrook ja Starner (2003) kasutasid k-keskmiste meetodit selleks, et ajaparaameetriga leitud

kohad üheks klastriks liita – antud juhul oli juba tagatud see, et andmestik, mida analüüsi, ei sisaldanud mürapunkte.

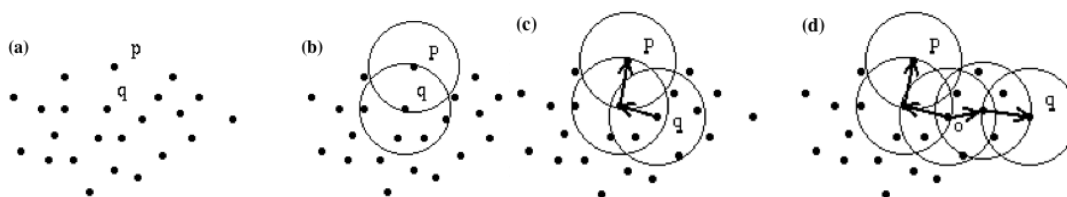


Joonis 5. K-keskmiste algoritmi tööprotsess: a – juhuslikult genereeritud punktid 2D tasandil; b – esialgsed klastrite keskpunkti koordinaadid valitakse juhuslikult; c – arvutatakse kõikide punktide kaugused tsentroidist ning määratletakse klastrisse vastavalt lähimale tsentroidile; d – tsentroidid liigutatakse klastrite keskpunkti (kõikide klastrisse kuuluvate punktide keskmine); e-g etapid kordavad eelnevat; h – klasteranalüüsi lõpptulemus (Konicek et al. 2012)

Tegevuskohtade tuvastamiseks, mida iseloomustab ümbritsevast alast tihedam koordinaadipunktide pilv, kasutatakse sageli erinevaid tihedusparameetritel põhinevaid klasteranalüüsi meetodeid, mis kasutavad parameetritena raadiust (*Eps*) ning minimaalset punktihulka (*MinPts*), mis antud raadiuse sisse peaks klastrite moodustumiseks jääma (Ester et al. 1996). Tihedusel põhineva klasteranalüüsi meetodi eelisena teiste meetodite ees nähakse sageli nende võimet eristada juhusliku kujuga klastreid, samuti ei kaasa sellised meetodid lõpptulemusse nn mürapunkte (*noise points*) (Zhou et al. 2007). Viimane on aga tegevuskohtade tuvastamisel oluline aspekt, kuivõrd müra lähevad arvesse ka kõikvõimalikud liikumised tegevuskohtade vahel. Samuti ei eelda sellised meetodid kasutajalt klastrinumbrite eelnevat määratlust nagu seda teeb etteantud klastrite arvuga k-keskmiste meetod. Suurimaks puuduseks selliste meetodite puhul on muutuva tihedusega klastritega mitte-arvetsamine (Ertöz et al. 2003). See saab GPS andmete puhul oluliseks pikemate perioodide uurimisel, kuna intuiivselt oletades on keeruline leida parameetrikomplekti, mis võimaldaks sellest perioodist optimaalselt eraldada tegevuskohad, mida külastatakse väga palju (kodu ja töö, toidupood) ning tegevuskohad, mille külastussagedus ei pruugi ületada ühte korda (turismiga seotud kohad).

Üheks sagedaseks kirjanduses ette tulevaks taoliseks meetodiks on DBSCAN (*Density-based spatial clustering of applications with noise*) (Ester et al. 1996, Nurmi

2009). DBSCANi tööpõhimõtte on kujutatud joonisel 6. Punktid klastris on jagatud tuumpunktideks (q) ja äärpunktideks (p) (joonis 6 (a)). Punkti q naabruskond (N), mis on ring raadiusega Eps , sisaldab alati minimaalselt kindla hulga punkte ($MinPts$), mis on kasutaja poolt defineeritud. Punkti p naabruskonnal sellist tingimust ei ole, küll aga peab punkt p ise kuuluma punkti q Eps naabruskonda. Seda võib vaadelda joonisel 6 (b) – kuigi punkti p naabruskonnas olev punktihulk ei ole suurem kui $MinPts=6$, kuulub see siiski klastrisse, jäädes punkti q naabruskonda. Joonised 6 (c) ja (d) kujutavad vastavalt punktide kättesaadavuse ja punktide ühenduvuse põhimõtet. Kõik punktid, mis nimetatud tingimustele ei vasta, määratletakse kui müra (*noise*).



Joonis 6. DBSCANi tööpõhimõtte ja klastri moodustumine vasakult paremale ($MinPts=6$): tuumpunkt (q) ja äärpunkt (p), äärpunkti kuulumine tuumpunkti naabruskonda, äärpunkti kättesaadavus ja punktide ühenduvus klastriks (Ester et al. 1996).

Ma et al. (2013) on kasutanud nimetatud meetodit, tuvastamaks veoautojuhtide liikumismustreid, Sun et al. (2013) uurisid Flickr'i geotätagiga fotosid, et tuvastada peamised turistide huvipunktid, Huang et al. (2013) selgitasid välja modifitseeritud DBSCANi abil sagedasti esinevad sarnaste tunnustega inimeste liikumistrajektorid ning Hwang et al. (2013) rakendasid DBSCANi, et uurida kogukonna-põhist mobiilsust raskemast vigastusest taastuvate patsientide puhul.

1.5.2 Tegevuskohtade semantiliste aspektide leidmine

Peatükkides 1.1 - 1.4 käsitletud inimese ruumilise käitumisega seotud erinevad aspektid nagu tegevuste planeerimine ajas ja ruumis ning nendega kaasnev suhestumine ümbritseva keskkonnaga (nt IKT kasutamine) on sageli allutatud teatavale rutiinile. Selline rutiin toodab inimkäitumises ruumilist stabiilsust ning ajalist perioodilisust nagu kirjeldatud peatükis 1.2 oleval joonisel 1 (Spissu et al. 2009, Hannes 2010). Mõned tegevused nagu töökäimine, ostlemine ning lastega seotud tegevuskohtade külastused (nt lasteaed) omavad nii ajas kui ruumis suurt stabiilsust ning on iseloomustatavad mitmete tunnustega nagu külastuse kellaeg, kestus ja nädalapäev (Frusti et al. 2002). Nutitelefoni lai levik ja suur kasutus inimeste seas on avardanud aga võimalusi telefonides paiknevate sensoritega koguda lisaks võrdlemisi intuiitivsetele kohakülastustunnustele veel lisaks teisi inimkäitumist ning ümbritsevat keskkonda puudutavaid andmeid (Do & Gatica-Perez 2013). Näited sellistest andmetest on ümbritseva keskkonna temperatuur, valgus- või müra ning ka mobiilikasutus ise. Selline mitmesugust informatsiooni sisaldav andmestik on

aluseks annoteerimismudelites, mis kasutavad kohtadele iseloomulikke või unikaalseid tunnuseid ja nende kombinatsioone selleks, et automaatselt lisada GPS andmetest leitud tegevuskohtadele semantiline aspekt.

Mudelid, mis automaatselt tuvastavad tegevuskoha semantilise tähenduse võib jagada kõige üldisemalt kaheks: protseduurilised ning masinõppelised (*machine learning*) (Bolbol et al. 2012). Protseduurilised mudelid hõlmavad endas teatud loogiliste eelduste rakendamist selle kohta, kuidas inimesed ühes või teises kohas käituvad, masinõppelised mudelid kasutavad kogutud andmestikku selleks, et seal olevad seosed töö käigus tuvastada ja “selgeks õppida”. Viimaste puhul tuleb aga teha optimaalne klassifikaatorite valik, see tähendab valida tunnused, mis kõige paremini kirjeldavad klassifitseeritavate elementide tüüpe (Bolbol et al. 2012). Bolbol ja teised (2012) rõhutavad oma töös statistiliste meetodite kasutamise vajadust klassifikaatorite valikuks, kuivõrd sel moel luuakse selgemad ja kindlamad alused tunnuste kaasamiseks mudelitesse.

Mitmed tööd koguvad aga lisaks automaatselt kogutavatele andmetele täiendavat informatsiooni respondentidelt, tagamaks võimalikult suure hulga kasuliku informatsiooni olemasolu. Näited kogutavast lisainformatsioonist hõlmavad sotsiaaldemograafilisi andmeid (Schönfelder et al. 2006), informatsioon reisi algus- ja lõppaegade, eesmärkide ja lähtekohtade kohta (Du & Aultman-Hall 2007) ning isegi võimalike GPS signaalide katkemiste kohta (Ogle et al. 2002). Üldine arusaam aga selle kohta puudub, kas automaatselt kogutav informatsioon peaks olema geograafiliste uuringute ainsaks andmesisendiks. Järjest pikemate perioodide uurimine soosib siiski respondentidele langeva koormuse vähendamist, mistõttu mudelid, mis tuvastavad automaatselt inimese tegevuskoha semantilise aspekti või reisi sooritamiseks kasutatava transpordivahendi tüübi, on muutumas järjest aktuaalsemaks.

2. Andmed

Käesolev töö põhineb Cambridge'i Ülikooli arvutiteaduste instituudis arendatud YouSense rakenduse (Linnap & Rice, ilmumas) poolt kogutavatel nutitelefoni sensorandmetel, mille kogumist alustati 2013. aasta esimeses pooles. Rakendus on välja töötatud eesmärgiga koguda energiasäästlikult võimalikult mitmekesiseid andmeid telefonikasutaja asukoha (GPS) ning telefonikasutuse kohta. YouSense'i põhises uuringus osaleb kokku üle 150 respondendi, kellest valiti käesoleva töö jaoks välja 18 inimest. Kõik välja valitud respondendid on Tartu Ülikooli geograafia osakonna töötajad või (endised) üliõpilased.

Uurimisperioodi jooksul on uuringus osalejatel kohustus kasutada oma peamise telefonina Androidi-põhist nutitelefoni, mis jagati kõikidele uuringus osalejatele, kes seda soovisid. Nendel, kes soovisid jätkata enda isikliku nutitelefoni kasutamist, seda võimaldati. Mõlemal juhul kehtib tingimus, et telefonivahetus uuringuperioodi jooksul tuleb kooskõlastada uuringu läbiviijaga, tagamaks andmekogumise pidevus ja kestvus, lisaks peab telefonis aktiivses olekus olema YouSense rakendus. Keskmiselt kogutakse iga respondendi kohta andmeid ligikaudu aasta.

Lisaks YouSense rakenduse andmetele koguti kõigilt 18-lt välja valitud respondendilt teavet nende tegevuskohtade semantika kohta. Selleks paluti uuringus osalejatel määratleda kõigi asukohaandmestikust tuvastatud tegevuskohtade kategooria, vastavalt etteantud kategooriate nimistule.

Uurimistöö hõlmab 6 kuu pikkust perioodi ajavahemikul 01.07.2013 – 31.12.2013. Poole aasta pikkune periood annab võimaluse töös käsitleda suuremal hulgal erinevaid tegevuskohti, võimaldades andmetest paremini tuvastada võimalikke tegevuskohaga seotud iseloomulikke tunnuseid. Töö uurimisel on kogu Eesti, uuringust on välja jäetud andmed, mis seostuvad välisreisidega, kuna välisreisidel võib telefonikasutus erineda tavapärasest (kallim andmeside, mitterutiinsem keskkond). Andmestik koosneb 10 519 803 geograafilisest koordinaadist ning 454 357 telefoni ekraaniga seotud sündmusest.

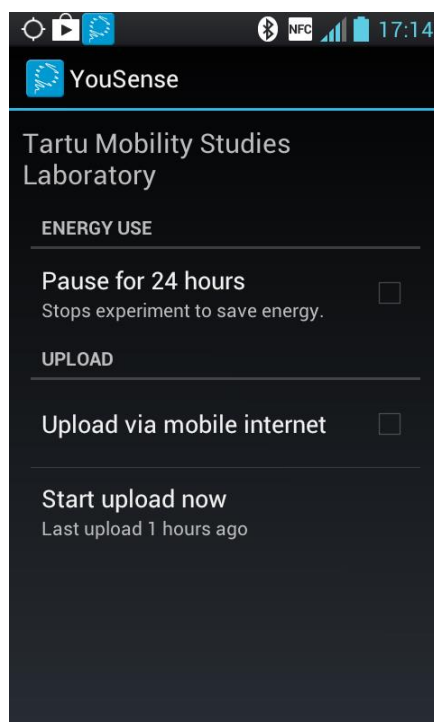
2.1. YouSense andmete kogumine ja salvestamine

YouSense rakendus salvestab andmeid andmebaasi sündmustena, mitte olekutena – see tähendab, et kirje andmebaasi tekib vaid oleku muutumisel. Kui muutusi ei esine, siis eelmise sündmusega seotud olek kehtib kuni järgmise sündmuseni (Linnap, isiklik kontakt, 2014). Tabelis 1 on välja toodud erinevat tüüpi andmed, mida YouSense rakendus mobiiltelefonidest kogub. Antud töös kasutatakse rakenduse poolt kogutud informatsioonist asukohainfot, telefoni ekraani sisse- ja väljalülitamist, klahviluku avamist ning süsteemi töösolekuga seotud teavet.

Tabel 1. YouSense rakenduse poolt kogutavad sündmused.

Kogutav sündmus
Asukoht GPS-ga
Telefoni ekraani sisse- ja väljalülitamine
Telefoni klahviluku avamine
Internetiühenduse olek: mobiiliinternet, Wi-Fi või mitte midagi
Laadija ühendamine ja lahtiühendamine
Aku täituvus 1% täpsusega
“Airplane mode” sisse- ja väljalülitamine
Äratuskella helisemine
Kella ja ajavööndi seadmine
Telefoni restart
Süsteemiinfo: rakenduse (taas)käivitamine, pausile panemine, “rakendus töötab” märged

Kuna YouSense kogub töösoleku ajal pidevalt informatsiooni inimese asukoha ning mobiilikasutuse kohta, on tegemist rakendusega, mis telefoniaku kestvust märgatavalt vähendab. Seetõttu on kasutajal soovi korral võimalus YouSense rakenduse töö peatada 24 tunniks (joonis 7), vältimaks telefoniaku liiga kiiret tühjenemist. Rakenduse võib manuaalselt ka taaskäivitada. Andmete edastamine serverisse toimub automaatselt, kas läbi Wi-Fi võrgu või mobiilse interneti, juhul kui kasutaja on seadetes nii määranud.



Joonis 7. Ekraanitõmmis YouSense rakendusest. Kasutaja muudetavad valikud on rakenduse töö peatamine 24 tunniks ning võimalus laadida andmed serverisse läbi mobiilse interneti.

2.1.1 Asukohainfo kogumine

Joonisel 8 on näide failist, kuhu on salvestatud erinevat tüüpi sündmused, mis on olulised antud töö kontekstis. Sündmus *sensor.gps* raames logitakse kellaaeg telefoni kella järgi, kellaaeg GPS moodulist, telefoni asukoha geograafiline laius ja pikkus, kõrgus, kiirus, täpsus ja suund. Kellaajad on UNIX *timestamp* formaadis ehk millisekundid 1970. aasta algusest UTC ajavööndis. Asukohainfo täpsus on lageda taeva all 3 meetrit, aga siseruumides või kõrgete majade vahel ka 50 või enam meetrit (Linnap 2014, isiklik kontakt).

```
1372657184309,app.heartbeat  
1372657210759,device.power.disconnected  
1372657210896,device.screen.on  
1372657217478,device.screen.unlocked  
1372657226184,device.screen.off  
1372657246322,user.prefs.pause,paused  
1372657256022,user.prefs.pause,unpaused  
1372657257331,sensor.gps,1372657257000,58.11188305,26.79800619,100.400,46.000,0.250,63.400  
1372657484318,app.heartbeat
```

Joonis 8. Erinevat tüüpi sündmused YouSense rakenduse andmefailist, mida töö raames kasutatakse

YouSense'i rakenduse eripära seisneb asukohainfo muutuva intervalliga kogumises - aeg, mis jääb kahe andmepunkti kogumise vahele sõltub liikumise kiirusest. Kiirustel alla 3 m/s kogutakse andmeid iga 16 sekundi järel, kiiremini liikudes iga 1 sekundi järel. Olekus, kus inimene telefoniga ei liigu, võib vahe aga kahe punkti kogumisel ulatuda mitme tunnini. - sel viisil on püütud andmekogumist muuta energiasäästlikumaks. Lisaks püüab rakendus energiat säästa sellega, et pärast telefoniga liikuma hakkamist ei hakata andmeid koguma koheselt, vaid teatud aja möödumisel. Hetkel on see aeg 30 – 60 sekundit.

GPS andmete kogumisse võib tekkida paus kolmel juhul: telefon lülitakse välja või aku saab tühjaks (*device.power.disconnected*), YouSense rakendus on käsitsi pandud pausile (*user.prefs.pause,paused*), kasutaja pani Androidi seadetest rakenduse kinni. Kahel esimesel juhul tekib sellest andmebaasi mäрге (joonis 8), viimasel juhul aga sellist märgi ei teki.

2.1.2 Telefonikasutusinfo kogumine

Sündmused *device.screen.on*, *device.screen.unlocked* ja *device.screen.off* (joonis 8) on seotud vastavalt mobiiltelefoni ekraani sisselülitamise (edaspidi sündmus ES), lahtilukustamise (edaspidi sündmus ELL) ja väljalülitamisega. Nende sündmuste toimumine salvestatakse iga kord, hoolimata sellest, kas inimene liigub telefoniga või mitte.

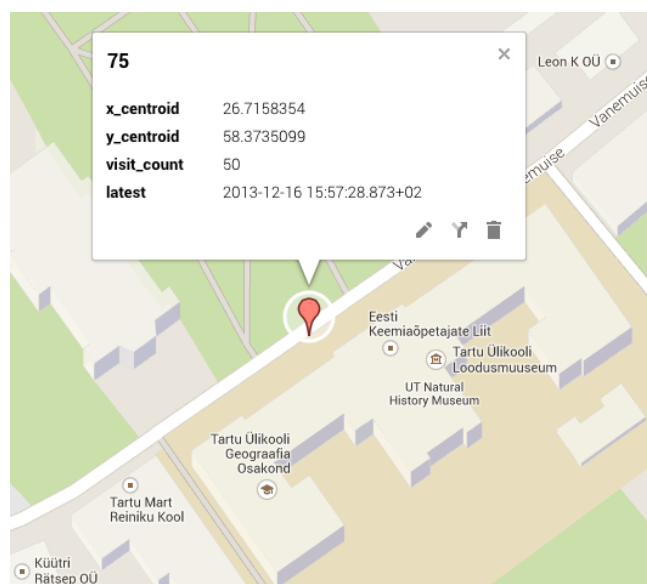
Sündmust *app.heartbeat* logitakse iga viie minuti järel eesmärgiga tuvastada andmekogumisse tekkinud pausi (telefonil on saanud aku tühjaks, telefon lülitati välja või mingil põhjusel ei ole YouSense rakendus sündmusi kirja pannud). Juhul kui viimasest *app.heartbeat* sündmusest on enam aega möödunud, võib eeldada, et vahepeal

logitud sündmuste vanad olekud enam ei kehti, aga uut kaa teada ei ole. Juhul kui kasutaja on YouSense rakenduse pannud käsitsi pausile, siis lakkab küll asukohainfo kogumine, kuid telefonikasutusega seotud informatsiooni kogutakse siiski edasi.

Kõigi sündmuste toimumise aeg salvestatakse telefoni enda kella järgi UNIX *timestamp* formaadis ehk millisekunditena 1970. aasta algusest UTC ajavööndis.

2.2 Tegevuskohtade annoteerimine

YouSense rakenduse poolt kogutav info võimaldab vaadelda seda, kus inimesed käivad, kuid nende tegevuskohtade semantika on teadmata. Kuivõrd töö üheks alameesmärgiks on välja selgitada erinevate tegevuskohtade külastustunnuste iseloomulikkus, siis on vajalik eelnevalt teada respondentide kuue kuu tegevuskohtade tähendusi. Selleks paluti väljavalitud 18 respondentil manuaalselt annoteerida kõik kuue kuu tuvastatud tegevuskohad, kus on vähemalt ühel korral viibitud enam kui 10 minutit ning mis ei asunud välismaal. Selleks koostati pärast tegevuskohtade tuvastamist ja külastustunnuste leidmist GPS andmestikust igale respondentile privaatne kaart Google Maps Engine veebirakenduses (joonis 9). Kaardile kanti kõik kuue kuu tegevuskohad ning igale tegevuskohale lisati respondentile abistavaks lisainformatsiooniks teave mitu korda antud tegevuskohta viimase kuue kuu jooksul külastati ning mis kuupäeval toimus viimane külastus. Kõik tegevuskohad olid markeeritud ID-ga, mis võimaldas hilisemat andmete sidumist andmebaasis olevate tegevuskohtadega. Tegevuskohtade annoteerimise etapp toimus 2014. aasta märtsis.



Joonis 9. Väljavõte kaardirakendusest, mille abil respondendid tegevuskohtadele kategooria määrasid.

Annoteerimisel paluti uuringus osalevatel inimestel valida tegevuskoha kategooria eelnevalt defineeritud 27 kategooria hulgast (tabel 2). Kategooriate koostamisel juhinduti Do ja Gatica-Perez (2013) tööst, mille eesmärgiks oli tegevuskohtade semantilise aspekti automaatne tuvastamine.

Kogu protsessi läbimiseks edastati kõikidele respondentidele juhend (lisa 1), kus oli etapi-kaupa kirjas, kuidas kategooriate lisamine toimub.

Tabel 2. Kategooriate nimistu, mille põhjal toimus tegevuskohtade annoteerimine.

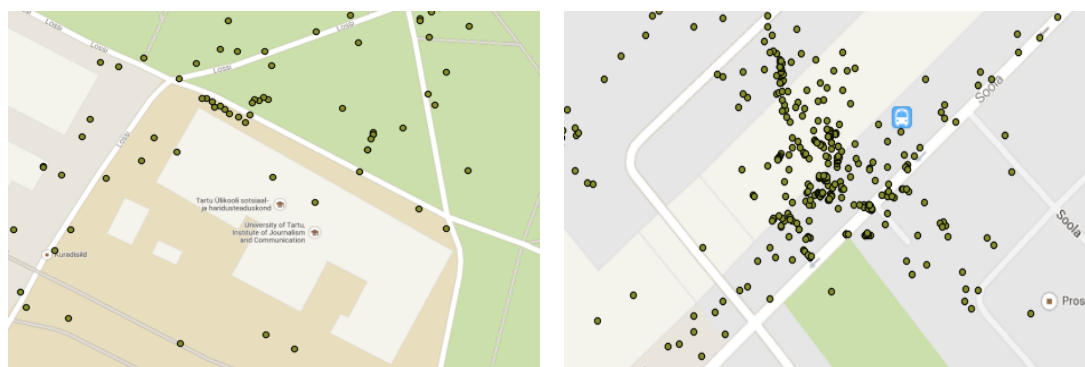
Kategooriad	Nr
Kodu	1
Töö	2
Töökohustustega seotud koht	3
Kooliga seotud kohad	4
Teine kodu	5
Sõbra/sugulase kodu	6
Sõbra/sugulase töö	7
Sõbra/sugulase kool	8
Lasteaed	9
Lapse kool	10
Lapse harrastus	11
Muu ametiasutus	12
Restoran/baar	13
Pood/ostukeskus	14
Meelelahutusasutus	15
Teater	16
Kino	17
Muuseum	18
Sport (sees)	19
Sport (väljas)	20
Spordiga mitte seotud hobi	21
Transpordiga seotud koht	22
Majutusasutus	23
Vaatamisväärsus	24
Puhkuse/turismiga seotud muu tegevuskoht	25
Muu	26
Ei tea	27

3. Tegevuskohtade tuvastamine

3.1 Tegevuskohad YouSense andmestikus

GPS andmete kogumine toimub sageli kindlaksseatud intervalliga (Herrera et al. 2010, Feng & Timmermans 2013, Wan & Lin 2013), mis on valitud, pidades silmas uuringu üldisi eesmärke ning neist tulenevaid tingimusi. YouSense'i rakendus kogub asukohaandmeid muutuva intervalliga, kusjuures aeg, mis jääb kahe andmepunkti kogumise vahele sõltub liikumise kiirusest. Asukohainfo kogumine lakkab, kui telefoniga ei liiguta.

Tulenevalt eelmainitust on ka andmetest äratuntavatel tegevuskohtadel duaalne iseloom, mis väljendub tegevuskohtade tuvastamisel kasutatava meetodikomplekti valikus.



Joonis 10. YouSense andmetes esinevad kahte tüüpi tegevuskohad. Vasakpoolsel joonisel on näha tegevuskoht, mis ei eristu ümbritsevast foonist tihedama punktivarvega, parempoolsel joonisel võib märgata aga tegevuskoha eristumist ümbritsevast keskkonnast.

Joonisel 10 on kujutatud andmetes esinevad kahte peamist tüüpi tegevuskohad. Parempoolsel joonisel asuv tegevuskoht on ümbritsevast foonist eristatav suhteliselt suurema punktutihedusega, vasakpoolsel joonisel asuv tegevuskoht aga ümbritsevast foonist suurema tihedusega esile ei kerki. Küll aga on vasakpoolsel joonisel asuvatele punktile iseloomulik oluliselt suurem eraldatus ajas, kui seda võiks eeldada kõrval oleval joonisel asuvatelt punktidelt. Olemuslikult tähendab see seda, et tegevuskohtade tuvastamiseks antud andmestikust ei piisa ühest meetodist või parameetrikomplektist, et kõik olulisemad kohad üles leida.

Töö autor otsustas erinevate klasteranalüüsi meetodite plussidele ja miinustele tuginedes (ptk 1.5) kasutada tegevuskohtade tuvastamiseks kahe meetodi kombinatsiooni – tihedusel põhinev klasteranalüüs ning ajaparaameetril põhinev meetod. Klasteranalüüsina kasutati DBSCANi kui laiemalt levinud meetodit, mis on kättesaadav erinevates vabavaralistes programmides ja pakettides. Ajaaukude tuvastamiseks kirjutas töö autor Pythoni skripti.

Kontrollimaks mõlema meetodi täpsust, paluti ühel respondendil üles märkida kõik ühe kuu tegevuskohad (tagasiulatuvalt). Sealjuures veenduti, et respondent oleks GPS punktide osas esinduslik – 1 kuu GPS punktide arv ei tohiks märkimisväärselt erineda 18 respondendi keskvärtusest (ca 100 000).

Leitud tegevuskohtadele külastustunnuste lisamiseks on töö autor kirjutanud Pythoni skriptid, mille tööpõhimõtte seletatakse lahti peatükis 3.3. Selguse mõttes on peatükis välja toodud vaid pseudokoodid, mis võimaldab lugejal paremini tehtud hoomata.

Otsitavateks külastustunnusteks on külastuse algusaeg, nädalapäev, kestus ja mobiilikasutuslikud tunnused (tabel 3). Mobiilikasutustunnustena käsitletakse ekraani sisselülitamise ja lahtilukustamise sagedust külastuse ajal, mobiili aktiivse kasutamise kestuse mediaani (aega, mil ekraan oli lukustamata) iga külastuse jaoks ning aktiivse mobiilikasutuse osakaalu külastusest.

Tabel 3. Otsitavad ja hilisemas analüüsis kasutatavad külastustunnused.

Külastustunnus	Mõõde
Külastuse kestus	min
Külastuse algusaeg	hh:mm:ss
Nädalapäev	E,T,K,N,R,L,P
Mobiilikasutus	
...ekraani sisselülitamise sagedus	korda tunnis
...ekraani lahtilukustamise sagedus	korda tunnis
...aktiivse telefonikasutuse osakaal külastusest	%
...aktiivse telefonikasutuse kestuse mediaan külastuse jaoks	min

Kõik ruumilised päringud ning muud andmebaasi toimingud on teostatud PostgreSQL andmebaasis.

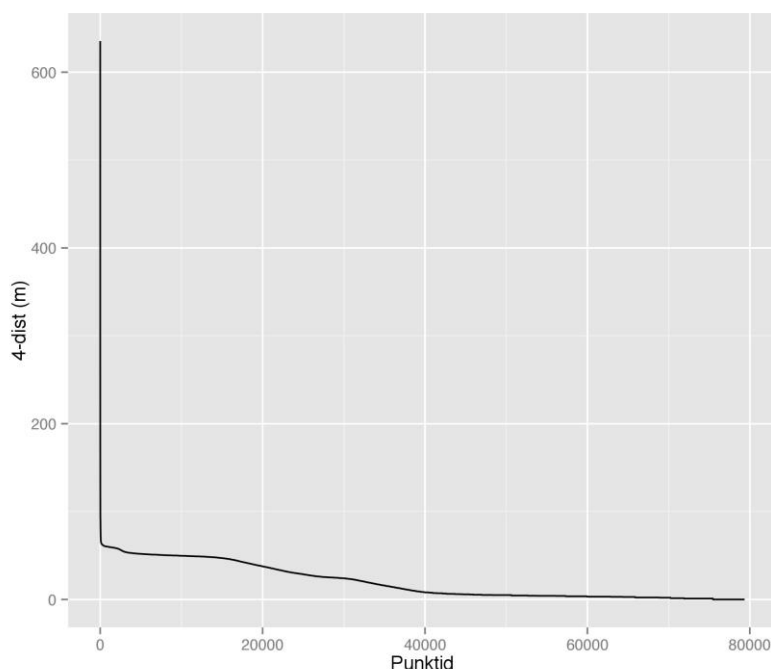
3.2 Tegevuskohtade tuvastamine naabruskonna tiheduse alusel

Nagu selgus peatükist 1.5, siis DBSCAN vajab kahte parameetrit klastrite moodustamiseks (*MinPts* ja *Eps*) ning nende parameetrite valik on olulise tähtsusega optimaalse tulemuse saavutamiseks. Optimaalseks tulemuseks peab autor reaalsusega enam-vähem kooskõlas olevat klastrite (tegevuskohtade) arvu ning klastrite suurust, mis peaks iseloomustama ka tegevuskoha enda suurust, samuti ei tohiks ühes klastris asuda mitu tegevuskohta.

3.2.1 Parameetrite *Eps* ja *MinPts* valik

DBSCANi kahe parameetri (*Eps* ja *MinPts*) väärtuse valimiseks soovib Ester et al. (1996) kasutada sorteeritud *k*-distanti graafikut, mis saadakse, arvutades andmebaasis oleva iga punkti jaoks tema *k* lähima naabri (*k*-th nearest neighbour) kaugus ning sorteerides saadud tulemused kahanevas järjekorras. Sel viisil on

teoreetiliselt võimalik eristada esimene murdepunkt graafikul ning antud k väärtus saab parameetri $MinPts$ väärtuseks ning murdepunktiks oleva punkti kaugus k -ndast naabrist saab parameetri Eps väärtuseks (Ester et al. 1996). See, kui suur k analüüsi tarbeks valida, jääb Ester et al. (1996) artiklis otseselt määratlemata – viidatakse vaid autorite enda kogemusel ja andmestikul põhinevale k väärtusele, mis on 4. Joonisel 11 on näidatud selline graafik YouSense andmebaasi juhuslikult valitud respondendi ühe kuu andmete kohta. On näha, et esimene murdepunkt ilmneb ligikaudu väärtuse 70 meetrit juures. Parameetri väärtuste $Eps=70$ ja $MinPts=4$ juures aga selgub klasteranalüüsi tulemuste visuaalsel vaatlusel, et väike $MinPts$ väärtus ja võrdlemisi suur Eps väärtus on tekitanud olukorra, kus klastrisse kuuluvad ka punktid, mis otseselt ei seostu tegevuskohaga (tegevuskohtade vahel liikumiseks kasutatud teed). Sellest annab aimu ka graafiku kuju – valdav osa punkte jääb valitud murdepunktist paremale, mis tähendab, et kõik need punktid kaasatakse ühte või teise klastrisse. See ei ole tegevuskohtade leidmisel kõige optimaalsem tulemus.



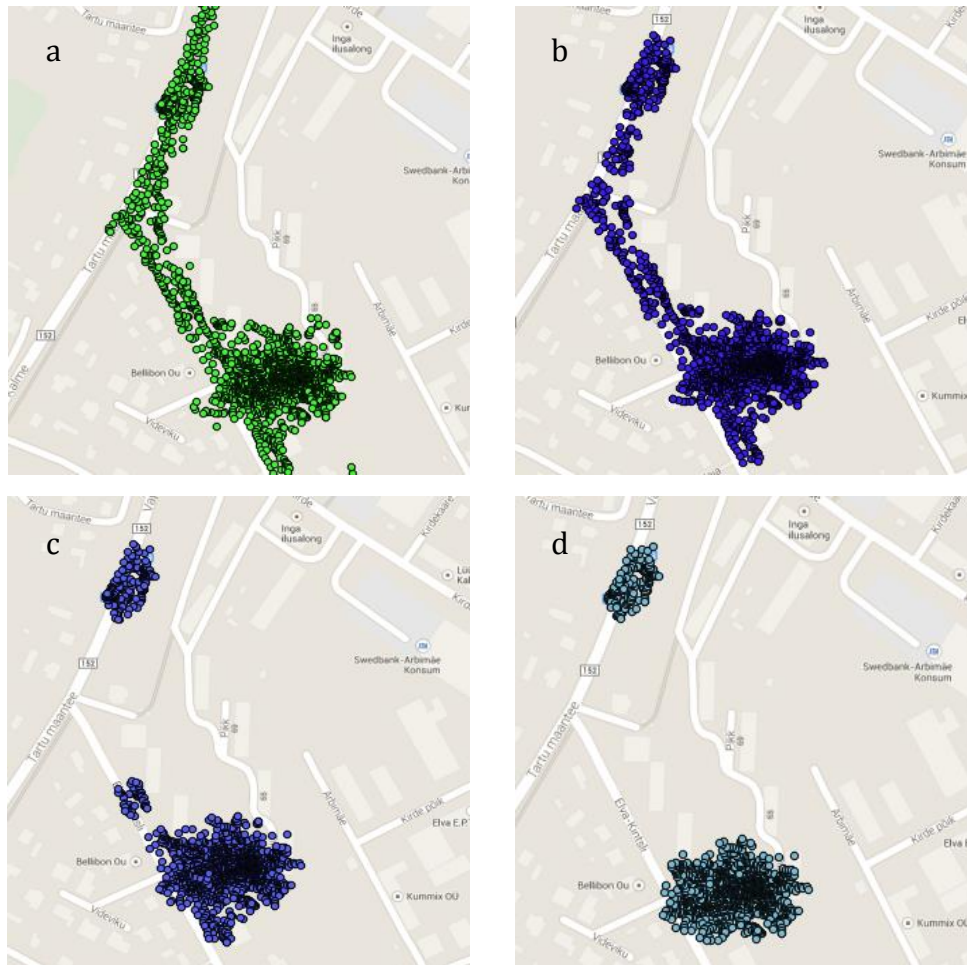
Joonis 11. k -nn analüüsi tulemused respondendi ühe kuu andmete kohta k väärtuse 4 juures.

Teiseks võimaluseks on tihedusel põhineval klasteranalüüsi parameetrite valikul võtta orientiiriks GPS andmete täpsus (Zhou et al. 2004) ning kasutada seda Eps parameetri väärtusena, mida antud töös ka tehti. Testides erinevaid $MinPts$ väärtusi, võib vaadelda, millised muutused ilmnevad (tabel 4). Parameetrikomplekt Eps 10 – $MinPts$ 10 puhul leitakse üles küll kõige suurem hulk tegevuskohti, kuid sageli on tegevuskohale liidetud ka lähedalasuvad teed ja teised tegevuskohad (joonis 12). Komplekti Eps 10 – $MinPts$ 40 puhul iseloomustab leitud klastreid kooskõla tegevuskoha suurusega, samuti ei ole klastritele liidetud lähedalasuvat teedevõrku (joonis 12).

Tabel 4. DBSCANi tulemused erinevate parameetri väärtuste juures.

Eps	MinPts	Klastrite arv	Tuvastatud	Olukord
10	10	562	78% (21)	Probleemid tekivad tegevuskohtadega, mida sagedasti kasutatakse ja mis asuvad ruumis lähedal – ühendatakse üheks klastriks. Teed on ühendatud tegevuskohtadega klastrisse. Sagedasti kasutatavad teed joonistuvad klastrina välja. Suur tähenduseta klastrite osakaal.
10	20	139	67% (18)	Suured klastrid sagedasti kasutatavatel teedel. Sagedasti kasutatava tegevuskoha lähedal asuvad kohad liidetakse klastriga.
10	30	67	60% (16)	Sagedasti kasutatavad kohad omavahel ühenduses. Teedel lõiguti klastrid. Tähenduseta klastrite osakaal hakkab vähenema.
10	40	40	52% (14)	Klastrid optimaalse suurusega. Sagedasti kasutatavatel teedel väiksemad klastrid fooride ja ristmike juures, samuti maanteed teatud lõikudel.
10	50	27	41% (11)	Klastrid teedel sagedasti kasutatavatel ristmikel ja fooride juures. Lähedal asuvad vähem külastatud kohad satuvad ühte klastrisse.

Parameetrikomplekti valik on siinkohal suuresti sõltuv töö eesmärkidest. Ühelgi juhul ei kerki esile väga head parameetrite paari, mis ei nõuaks kompromisside tegemist. Kui töö seab eesmärgiks tuvastada vaid DBSCANi abil YouSense'ga sarnasest hõredast andmestikust võimalikult palju tegevuskohti, siis väike *Eps* ja *MinPts* väärtus leiab üles suurema hulga tegevuskohti. Miinusena kerkib esile vajadus eristada leitud paljude tegevuskohtade hulgast vaid need kohad, millel võiks respondendile olla reaalne tähendus. Juhul kui töö üheks eesmärgiks on vaadelda ka tegevuskohtadega seonduvaid külastustunnuseid, siis eelmainitud võimalus ei ole kõige sobivam – suured teedega ühenduses olevad tegevuskohad ei võimalda hinnata külastustunnuseid piisava täpsusega.



Joonis 12. Parameetri *MinPts* muutumine ja sellega kaasnevad muutused klastri struktuuris: a) *Eps* 10 – *MinPts* 10; b) *Eps* 10 – *MinPts* 20; c) *Eps* 10 – *MinPts* 30; 4) *Eps* 10 – *MinPts* 40. Lõpptulemusena on sagedasti külastatavast tegevuskohast eraldunud lähedalasuv bussijaam.

Kuivõrd käesoleva töö puhul on sõnastatud eesmärk vaadelda tegevuskohtadega seonduvaid külastustunnuseid, soovitakse valida sellest aspektist võimalikult sobivad parameetrid. Antud juhul selgub (tabel 4), et väärtused *Eps* 10 – *MinPts* 40 on külastustunnuste seisukohalt kõige sobivam valik. Puudu olevate tegevuskohtade arvu on võimalik vähendada aga andmeauke tuvastava skripti abil.

3.2.2 DBSCANi rakendamine

DBSCANi algoritmi rakendati vabavaralise tarkvaraga ELKI (joonis 13), mis tagab olemasolevatest DBSCANi rakendada võimaldavatest programmidest (testiti veel Wekat ja R-i paketti “*fpc*”) kõige kiirema ning paindlikuma andmetöötluse, võimaldades valida erinevate indekseerimismeetodite ning distantsifunktsioonide vahel.

Parameter	Value
rtree.reinsertion-...	Default: CloseReinsert
rtree.reinsertion...	Default: 0.3
rtree.reinsertion...	geo.LatLngDistanceFunction
geo.model	SphericalHaversineEarthModel
spatial.bulkstrategy	SortTileRecursiveBulkSplit
time	Default: false
algorithm	clustering.DBSCAN
algorithm.distancefu...	geo.LatLngDistanceFunction
geo.model	Default: SphericalVincentyEarthModel
dbscan.epsilon	
dbscan.minpts	
evaluator	Default: [class de.lmu.ifi.dbs.elki.evaluation.AutomaticEvaluation]
resulthandler	Default: [class de.lmu.ifi.dbs.elki.visualization.gui.ResultVisualizer]
vis.window.title	
vis.window.single	Default: false
vis.sampling	Default: 10000
visualizer.stylesheet	Default: default
vis.enable	Default: ^\Qde.lmu.ifi.dbs.elki.visualization\E\.*

[Saved Settings] Load Save Remove Run Task

Joonis 13. ELKI graafiline kasutajaliides ja osa selle poolt pakutavatest võimalustest.

Andmetöötluse kiirendamiseks kasutati R-puu (*R-tree*) indekseerimist ning distantsifunktsioonina nii indekseerimise kui ka DBSCANi puhul pikkus- ja laiuskraadide vormis 2D vektoritele mõeldud funktsiooni koos Haversine'i Maa mudeliga. DBSCANi parameetrid olid $Eps=10$ meetrit ja $MinPts=40$.

Klasteranalüüs toimus respondentide kaupa, kusjuures iga respondendi 6 kuu andmestik jagati 1 kuu pikkusteks alamhulkadeks. Alamhulkadeks jagamine tähendab ühe sammu lisandumist hilisemasse andmetöötlusesse – saadud tulemused peab edasise töö tarbeks taas kokku liitma. Seda on võimalik saavutada, konstrueerides iga kuu kohta samasse klastrisse kuuluvatele punktidele kumerad katted (*convex hullid*) ning seejärel liites omavahel lõikuvad polügonid kokku (nt QGIS *Dissolve* tööriist).

3.3. Tegevuskohtade tuvastamine “aukude” meetodil

YouSense rakendus on loodud olema võimalikult energiasäästlik, mistõttu andmeid ei koguta kindla intervalliga, vaid see on otseses sõltuvuses telefoni kasutamisest ja sellega liikumisest, millest tulenevalt võib andmetesse tekkida nõ “auke”. Kuivõrd GPS sensori poolt andmebaasi edastatud koordinaatide hulk ei pruugi sellistes tegevuskohtades olla piisavalt suure tihedusega, eristamaks tegevuskohti ümbritsevast foonist tihedusel põhineva klasterdamise meetodiga, on tarvis kasutada meetodit, mis võimaldab leida tegevuskohad üles, lähtudes tekkinud “aukudest” ehk ajast kahe punkti logimise vahel. Idee seisneb selles, et uus tegevuskoht leitakse siis, kui vahemaa uue ja vana koha vahel on suurem kui seatud piir ning kui uues kohas on viibitud teatud aja vältel. Käesolevas töös on kasutatud sarnast lähenemist, kus tehakse kindlaks need koordinaadipaarid andmestikus, mille vahemaa ajaskaalal on enam kui seatud piir t , kuid ruumis asuvad need punktid teineteisele lähemal kui maksimumiks seatud piirmäär s .

Tegevuskohtade tuvastamiseks ajaparaameetriga kirjutati Pythoni skript (joonis 14), mis kontrollib andmestiku iga GPS punkti kaugust ajas ja ruumis sellele punktile

järgneva punkti suhtes ning väljastab seejärel järjendi punktide koordinaatidega, mis rahuldavad seatud tingimust:

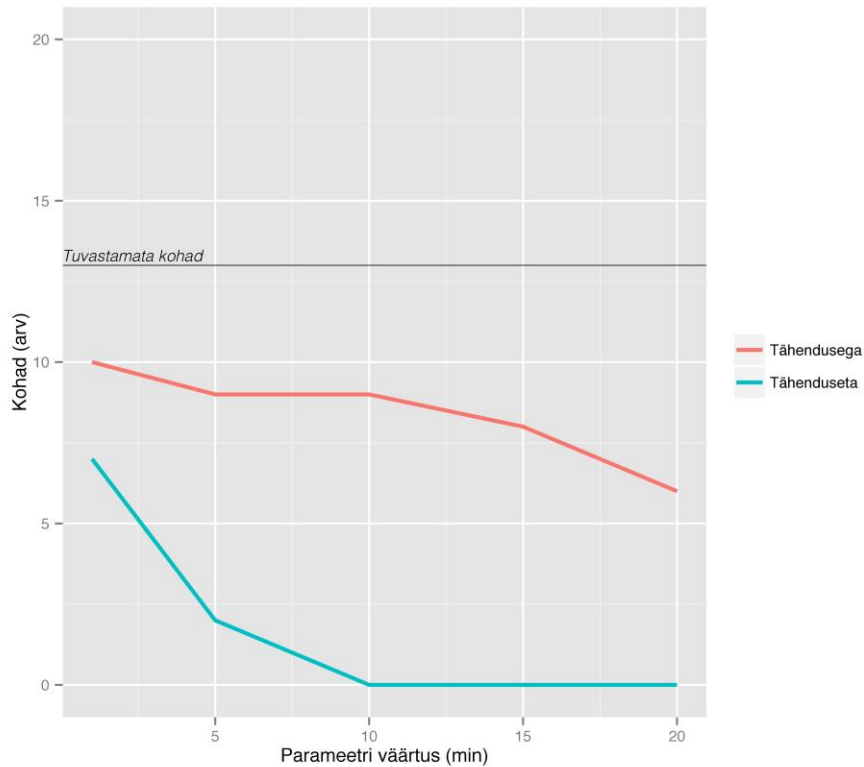
$$ajavahe_{mik} > t \text{ ja Haversine'i kaugus} < s$$

```
1. for every pair of geographic coordinates in a list
2.   compute time difference with the next pair of geographic coordinates
3.   compute Haversine distance between current pair and the next pair
4.   if time difference >= threshold and distance <= threshold then
5.     append current pair to the set of timecaps points
6.   end if
7. endfor
```

Joonis 14. Pythoni skripti pseudokood, millega tuvastati töö raames andmetes esinevad “augud”.

Kaugusparameetri s väärtus on valitud, lähtudes teadmised, et pärast uuesti liikumahakkamist kulub YouSense rakendusel 30-60 sekundit, et uuesti asukohainfot koguma hakata. See tähendab, et vastavalt liikumisviisile (jalgsi, transpordivahendiga) võib järgmine kogutav punkt olla eelmisest üsna kaugel. Proovides erinevaid kaugusparameetrite väärtusi selgus, et 250 meetri juures leitavate tegevuskohtade hulk stabiliseerus.

Ajaparametri valimiseks kasutati taas ühe respondendi teadaolevaid tegevuskohti ning seejärel testiti kaugusparameetriga (50 m) võimalikke ajaparametri väärtusi, tuvastamaks, milline väärtus võimaldab koos DBSCANi tulemustega tuvastada tegevuskohti kõige täpsemini. Sobivaks peeti parameetrit, mis ei dubleeriks DBSCANi tulemusi ning samas ei tekitaks juurde tegevuskohti, millel respondendile sisuline tähendus puudub. Tulemuste hindamiseks loeti kokku tegevuskohad, mis jäid DBSCANi poolt tuvastamata ning vaadeldi, mitu kohta tuvastamata jäänud tegevuskohtadest ajaparameter väärtusvahemikus 1 kuni 20 minutit üles suudab leida. Lisaks vaadeldi iga parameetri väärtuse juures, kui palju sisulise tähenduseta tegevuskohti juurde tekib. Selleks paluti respondendil uued tekkinud kohad üle vaadata ning kinnitada või ümber lükata tegevuskoha tähenduslikkus. Tulemused on esitatud joonisel 15. Selgub, et parameetri väärtus 1 minut tuvastab küll enim puuduolevaid tegevuskohti, kuid sealjuures loob juurde kõige rohkem neid tegevuskohti, mis jäid tähenduseta. Optimaalseim valik antud juhul on väärtus 10 minutit - tähenduseta tegevuskohti juurde ei loodud ning tuvastatud kohtade arv on lähedane maksimaalsele võimalikule.



Joonis 15. “Aukude” meetodi efektiivsus erinevate ajaparametri t väärtuste juures. Parim on selline väärtus, mis tuvastab võimalikult suure hulga tähendusega kohti, sealjuures loomata juurde uusi tähenduseta kohti.

Kuivõrd antud meetod tagastab tegevuskohta markeeriva ühe punkti, mis on olnud kooskõlas seatud tingimusega, on edasise analüüsi jaoks tarvis kaasata ka leitud punkti lähiümbruses asuvad teised punktid, ilma milleta ei ole võimalik antud tegevuskohas külastustunnusi vaadelda. Nende punktide hulgas võib esineda veel teisi ajaparametriga tuvastatud punkte, kuid see ei ole reegel. Selleks, et sellised punktid analüüsi samuti kaasata, loodi igale ajaauguga seotud punktile ümber 50 meetrine puhver, mille sisse jäävad punktid arvestati kõik samasse klastrisse. Selline raadius valiti välja, arvestades GPS andmete täpsust. Omavahel lõikuvad puhvid ühendati üheks klastriks ning juhul kui “aukude” meetodil tuvastatud klastrid lõikusid tihedusel põhinevate klastritega, ühendati need samuti omavahel. Sel moel ühendati kahe meetodi (DBSCAN ja “augud”) tulemused ühtseks tegevuskohtade kogumiks. Tekkinud polügonide abil on 6 kuu GPS andmestikust võimalik andmebaasis eraldada klastritesse kuuluvad punktid.

Selleks, et kasutatud meetodika tulemusi valideerida, vaadeldi ortofoto taustal kõigi 18 uuringus osaleja tekkinud klastreid. Juhul kui mõni tegevuskoht oli nähtavalt killustunud (suuremad kaubanduskeskused), liideti klastrid kokku manuaalselt üheks klastriks. Liiga suured, potentsiaalselt mitut tegevuskohta sisaldavad klastrid (töö+kodu ning Tartu Raekojaplatsi mõnel juhul tekkinud üks suur klaster) eemaldati edasisest analüüsist

4. Külastustunnuste lisamine tegevuskohtadele ja statistilise analüüsi meetodika

4.1 Külastuse alguse ja kestuse arvutamine

Sarnaselt Do ja Gatica-Perez (2013) tööle, on antud töös on tähenduslik tegevuskoht defineeritud kui ala, kus on vähemalt ühe külastuse vältel viibitud kauem kui 10 minutit. “Aukude” meetod kätkeb endas sisulist tegevuskoha definitsiooni, mistõttu võib eeldada, et kõik sel moel leitud klastrid on juba tegevuskohad *per se*. Sama ei saa kinnitada tihedusklasterdamise meetodi põhjal leitud klastrite kohta – klaster võib olla tekkinud igapäevaselt kasutatava sõidutee mõne foori juurde, kuid kunagi ei ole seal viibitud kauem kui 10 minutit. Seetõttu on vajalik selgitada välja kõikide potentsiaalsete tegevuskohtade iga külastuse kestus.

See on saavutatud, järjestades andmebaasis vastava respondendi kõik punktid, mis kuuluvad ühte või teise klastrisse, aja alusel ning otsides üles kohad andmebaasis, kus klastrinumbrid vahetuvad (joonis 16).

```
59.416290979999999;24.709285019999999;353089057256113.00000000;1372662575139.0000000000;64
59.097936429999999;25.376396929999999;353089057256113.00000000;1372666236652.0000000000;109
59.097826540000000;25.376735790000001;353089057256113.00000000;1372666237653.0000000000;109
59.097732020000002;25.377082020000000;353089057256113.00000000;1372666238656.0000000000;109
58.888436200000001;25.574669390000000;353089057256113.00000000;1372667930065.0000000000;2
58.888310470000000;25.574770269999999;353089057256113.00000000;1372667946103.0000000000;2
58.888309040000003;25.574769289999999;353089057256113.00000000;1372667961461.0000000000;2
58.888245140000002;25.574765010000000;353089057256113.00000000;1372667981115.0000000000;2
58.887813520000002;25.575557849999999;353089057256113.00000000;1372672480437.0000000000;2
58.371691939999999;26.716792219999999;353089057256113.00000000;1372678156622.0000000000;15
```

Joonis 16. Külastused andmebaasis. Klastritesse kuuluvad GPS punktid on järjestatud aja alusel – klastrinumbrid vahetuvad sellises järjestuses märgivad ühe külastuse lõppu ja teise algust.

Vastava toiminguteostamiseks on koostatud Pythoni skript (joonis 17), mis nimetatud kohad üles leiab. Skript arvutab sellises sarnaste numbrite järjestuses esimese ja viimase numbriga seotud ajahetkede vahe (igal GPS punktiga on seotud kellaeg). See vahe ongi otsitav külastuse pikkus ning esimene ja viimane element sellises järjestuses märgib vastavalt algus- ja lõpukuupäeva. Kõik klastrid, kus ükski külastus ei ületanud 10 minuti alampiiri eemaldati edasisest analüüsist.

```
1.for every cluster number
2. find sequences from time-ordered GPS points
3. if cluster number is equal to GPS point cluster number
4. append GPS point to chunk
5. else
6. if length of chunk is not null
7. calculate time difference between first and last element in a chunk
8. append cluster number, time difference, first and last element to a 2D list
9. end if
10. end if
11.endfor
```

Joonis 17. Pseudokood külastuse pikkuse, algus- ja lõpukuupäeva ning kellaaja leidmiseks aja alusel järjestatud, klastrinumbriga varustatud GPS andmestikust.

4.2 Mobiilikasutustunnuste sidumine külastustega

Mobiilikasutustunnuste sidumiseks külastustega on antud andmestiku puhul võimalik kasutada mobiilisündmuse toimumise aega. Iga telefonikasutust puudutava sündmuse salvestamisega andmebaasi salvestatakse ka selle sündmuse toimumise aeg. Kui sündmuse toimumise aeg jääb mõne külastuse ajalistesse piiridesse, võib selle sündmuse siduda külastusega. Siinkohal tuleb märkida, et külastustunnuste täpsus on otseses sõltuvuses klasteranalüüsi tulemuste täpsusest, mis tähendab, et ebasobiva parameetrite komplekti valikul võib juhtuda, et külastused on tükeldatud väiksemateks osadeks või on need liiga pikad. Tulemuste täpsust on võimalik hinnata hilisema andmete ülevaatamisega, kontrollimaks, kas mõni tegevuskoht sisaldab ebanormaalselt pikki külastusi või liiga sagedasi väikseid külastusi (nt kodu külastused on sagedased ja väga lühikesed).

Käesoleva töö raames tehti järgmised loogilised eeldused, tuginedes andmete vaatlusel selgunud seaduspäradele:

- ekraani lahtilukustamise sündmusele eelneb alati ekraani sisselülitamise sündmus;
- ekraani lahtilukustamise sündmusele järgneb alati ekraani väljalülitamise sündmus (mitte uuesti lahtilukustamine või sisselülitamine).
- ekraani lahtilukustamise ja väljalülitamise vahel esinevad sündmused nagu telefoni sisse- või väljalülitamine ning rakenduse käivitamine võivad andmetesse tekitada erindeid või eelnimetatud loogikale vasturääkivaid tulemusi.

Joonisel 18 on loodud Pythoni skripti põhimõtteline ülesehitus, mis leiab ekraani sisselülitamiste ja lahtilukustamiste arvu iga külastuse jaoks. Skript loeb kokku kõik ekraani sisselülitamise ja lahtilukustamisega seotud sündmused, mis jäävad külastuse algus- ja lõppkuupäeva ning kellaaja piiridesse.

```
1.for every visit
2. for every mobile event check if time of 'device.screen.on' or 'device.screen.unlocked' is between check in and check out
3.   if true
4.     count 'device.screen.on'
5.     count 'device.screen.unlocked'
6.   else
7.     continue
8.   end if
9. append counted events to visit
10. endfor
11.endfor
```

Joonis 18. Pseudokood ekraani sisselülitamise ja lahtilukustamise sageduse arvutamiseks iga külastuse jaoks.

Mobiili aktiivse kasutuse kestuse arvutamiseks tuleb lisaks sündmuse toimumise ajalitele piiridele kontrollida ka seda, et andmetes ei esineks loogilisi vasturääkivusi, mis võivad tuleneda vigadest andmekogumisel, andmeedastusse tekkinud pausidest (telefoni aku sai tühjaks, telefon lülitati välja), mõjutades aktiivse mobiilikasutuse kestuse täpsust. Selleks kontrollitakse ajaliselt järjestatud telefonisündmuste järjendis kõigepealt ekraani lahtilukustamise sündmuse (*device.screen.unlocked*) ja sellele

vahetult järgnenud ekraani väljalülitamise (*device.screen.off*) sündmuse sattumist külastuse ajalistesse piiridesse. Seejärel kontrollitakse, et kahe ekraani sisselülitamise ja väljalülitamise vahele ei jääks sündmusi, mis võivad mõjutada tulemuse täpsust (joonis 19).

```
1.for every visit
2. for every mobile event
3. if event=='device.screen.unlocked' and time parameter is between check_in and check_out
4. usage start time=time parameter of 'device.screen.unlocked'
5. else if event=='device.screen.off' and time parameter between check_in and check_out and usage start time > 0
6. check if outlier events occur between usage start time and device.screen.off time parameter
7. if false
8. usage end time=time parameter of 'device.screen.off'
9. usage time=usage end time-usage start time
10. append usage time to visit
11. end if
12. endfor
13.calculate sum of all usage times
14.calculate median
15.endfor
```

Joonis 19. Pseudokood aktiivse telefonikasutuse kestuse arvutamiseks iga külastuse jaoks.

Eespool kirjeldatud meetodil on võimalik külastusi iseloomustada läbi nende mobiilikasutuslike tunnuste. Teatud olukorras võib aga tekkida vajadus mobiilikasutustunnuseid vaadelda eraldi lähtuvalt nende toimumise täpsest ruumilisest asukohast (näiteks mobiilikasutuse vaatlemine reisi ajal). Sellisel juhul ei ole kirjeldatud meetodid enam adekvaatsed. Probleemi lahendamiseks tuleb iga sündmusega siduda sellele sündmusele vahetult eelnenud GPS sensori poolt salvestatud asukoha geograafilised koordinaadid ning kontrollida seda, et kasutaja ei oleks vahepeal YouSense rakendusest GPS andmete kogumist peatanud. Autor on antud meetodi skripti näol realiseerinud, kuid käesoleva töö ulatust arvestades seda eraldi ei kajastata.

4.3 Külastustunnuste analüüs

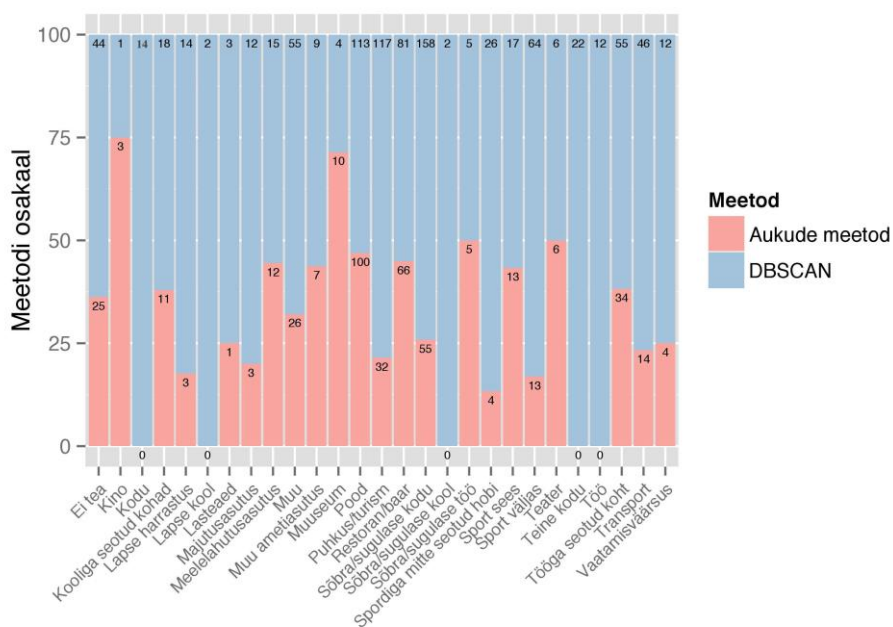
Külastustunnuste analüüsi etapis selgitatakse välja, millised varieeruvused esinevad külastustunnustes erinevate tegevuskohakategooriate lõikes. Külastustunnustena käsitletakse külastuse kestust, külastuse algusaega (kellaaeg), nädalapäeva, ekraani sisselülitamise ja lahtilukustamise sagedust, aktiivse mobiilikasutuse kestust ning osakaalu külastusest. Arvesse võttes uuringu väikest valimit, kasutan eesmärgi täitmiseks mitteparameetrilisi meetodeid, kuivõrd normaaljaotuse kontroll ei pruugi anda usaldusväärseid tulemusi.

Tegevuskohtade vaheliste erinevuste väljaselgitamiseks peab autor sobivaks Friedman'i testi, mis on sõltuvatele üldkogumitele mõeldud mitteparameetiline alternatiiv ühefaktorilisele dispersioonanalüüsile. Kuna Friedman'i test võimaldab öelda vaid seda, kas andmetes esinev teadaolev varieeruvus on suurem kui teadmata varieeruvus, tuleb rühmade vaheliste erinevuste väljaselgitamiseks kasutada *post hoc* testi. Käesolevas töös kasutatakse paarikaupa võrdlemiseks Wilcoxon'i astakmargitesti, mis testib nullhüpoteesi, et kahe sõltuva valimi keskmine astak on null.

5. Tulemused

5.1 Tegevuskohtade tuvastamine

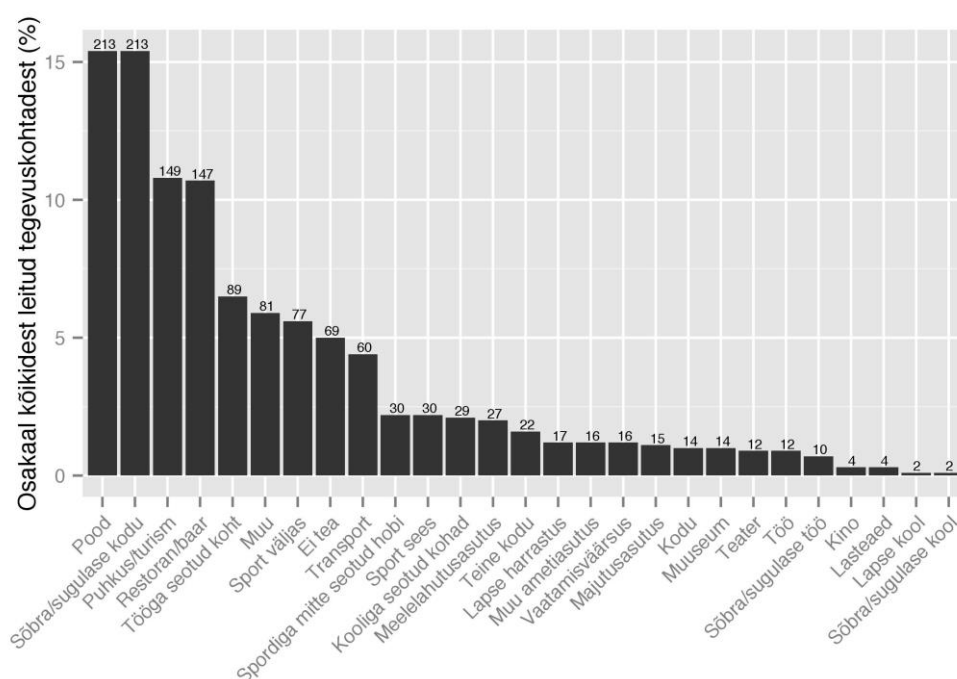
Kahe algoritmi tulemusena leiti ühtekokku 1379 tegevuskohta ning tuvastati 8186 külastust. Ligikaudu kolmandik külastustest (447) olid leitavad *vaid* “aukude” meetodil ning kaks kolmandikku (932) olid leitavad DBSCANiga. Joonisel 20 on kujutatud tulemused ka tegevuskoha kategooriate lõikes. Selgub, et sagedasti külastatavad kohad nagu kodu, töö ja teine kodu on kõikidel juhtudel leitavad tihedusel põhineva meetodiga nagu DBSCAN. Lisaks neile kohtadele tuvastas meetod kõik leitud tegevuskohad ka kategooriates lapse ja sõbra või sugulase kool. Kõige vähem erinevaid tegevuskohti tuvastas DBSCAN kino ja muuseumi kategoorias (vastavalt 25% ja 29%), mis tähendab, et nendes kategooriates oli enamik tegevuskohti leitavad *vaid* “aukude” meetodil. Mõlemad meetodid tuvastasid ligikaudu võrdse hulga tegevuskohti kategooriates sõbra/sugulase töö, teater, pood, restoran – baar, meelelahutusasutus, ametiasutus ning sport sees.



Joonis 20. DBSCANi efektiivsus erinevates tegevuskohtades. Vertikaalteljel on kujutatud meetodi poolt leitud kohtade osakaal (%) ning tulpadel on märgitud tuvastatud tegevuskohtade absoluutväärtused.

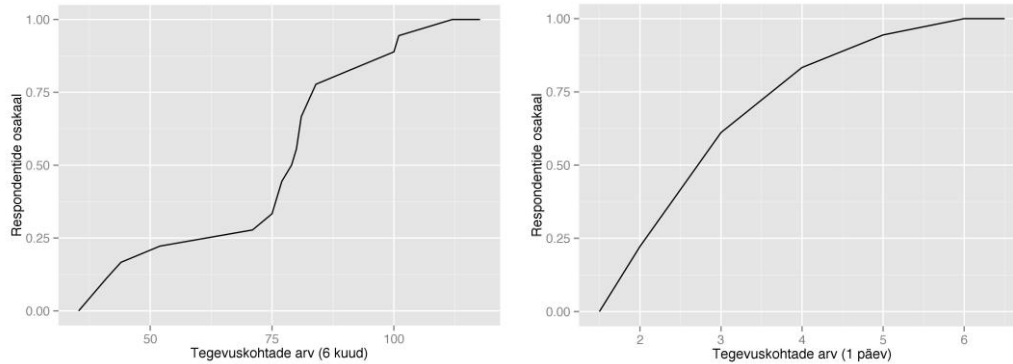
Kokku kõige enam tegevuskohti tuvastati poodide ning sõbra-sugulase kodu kategoorias, moodustades kõikidest vaadeldud tegevuskohtadest ligikaudu 30% (joonis 21). Neile järgnevad puhkuse ja turismiga seotud tegevuskohad (11%) ning toitlustusega seotud kohad nagu restoranid ja baarid (11%). Kõige vähem tegevuskohti asub aga kategooriates sõbra – sugulase kool (0.1%), lapse kool (0.1%), lasteaed (0.3%) ja kino (0.3%). Kategoorias “ei tea” oli kokku 69 tegevuskohta, moodustades kõikidest tuvastatud tegevuskohtadest 5%.

Eraldi tasub tähelepanu pöörata kategooriale kodu – 18 respondendi kohta on kodusid vaid 14. Selline tulemus on tingitud asjaolust, et neljal respondendil asusid kodu ja töö ruumis teineteisele väga lähedal, mis tingis nende tegevuskohtade sattumise ühte klastrisse.



Joonis 21. Tuvastatud tegevuskohtade jaotumine kategooriate lõikes. Vertikaalteljel on kujutatud tegevuskoha kategoorias asuvate kohtade osakaal kõikidest leitud tegevuskohtadest ning igal tegevuskohal on sildina kujutatud ka leitud kohtade arvuline väärtus.

Respondentide lõikes tuvastatud tegevuskohti vaadeldes selgub, et enamik inimesi külastas 6 kuu jooksul 50 – 90 erinevat tegevuskohta (joonis 22). Inimesi, kellel oli enam kui 90 tegevuskohta, oli kokku 4, moodustades uuritavatest 20%. Väga väikese tegevuskohtade arvuga (<50) respondente oli kokku 3 (17%).



Joonis 22. Respondentide tegevuskohtade arvu jaotusfunktsioon kuue kuu (vasakul) ning ühe päeva kohta (paremal).

5.2 Tegevuskohtade külastustunnused

Käesoleva peatüki eesmärk on kasutada leitud külastustunnuseid selleks, et statistiliste vahenditega hinnata nende iseloomulikkust erinevate tegevuskohtade jaoks. Kuivõrd eelnevast peatükist selgus, et nii mõneski kohakategoorias on tegevuskohti väga vähe, otsustati edasise analüüsi tarbeks liita kokku järgmised sarnase külastusiseloomuga kategooriad:

- tööga liideti tööga ja kooliga seotud kohad (töö/kool)
- puhkuse ja turismiga seotud tegevuskohale liideti majutusasutus ning vaatamisväärsus (puhkus)
- sport sees ja sport väljas liideti üheks kategooriaks (sport)
- ühendati muu meelelahutusasutus, teater, kino, muuseum, spordiga mitte seotud hobi (vaba aeg).

Statistilisse analüüsis on väikese tegevuskohtade ja respondentide arvu tõttu välja arvatud kategooriad sõbra/sugulase töö/kool, lasteaed, lapse kool, lapse harrastus ning ametiasutused. Samuti ei käsitleta eraldi kategooriaid “muu” ning “ei tea”.

5.2.1 Külastuspäeva tüüp kui iseloomulik tunnus

Päevatüüpina käsitletakse käesolevas töös argipäevi (esmaspäev-reede) ja nädalavahetusi (laupäev-pühapäev) ning mõõdetavaks tunnuseks on külastuste osakaal argipäevadel ja nädalavahetustel. Kuna tegevuskoha kategooriate-vahelised erinevused osakaaludes jäävad samaks nädalavahetustel ning argipäevadel, viiakse statistiline analüüs läbi vaid ühe päevatüübiga. Analüüsi tulemused on seejärel laiendatavad ka teisele päevatüübile.

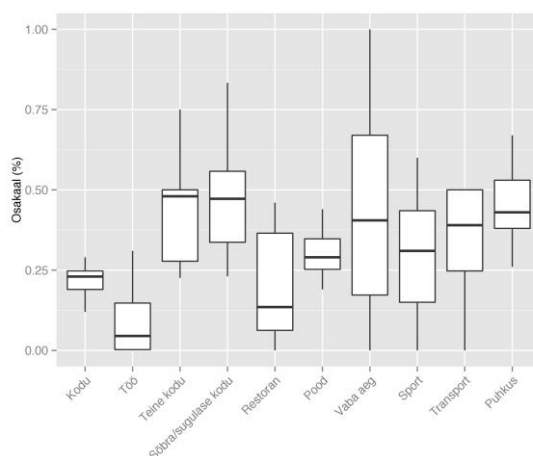
Muutused külastuste arvus ja osakaaludes külastuspäeva tüüpide lõikes on kajastatud tabelis 5. Kõikide kategooriate puhul on märgatav külastuste osakaalu langus nädalavahetustel, kuid suuremad erinevused tegevuskoha kategooriate vahel on märgatavad vaid kodu (23%), töö-kool (6%), restoranide (22%) ja spordiga seotud

tegevuskohtade (26%) puhul, mis erinevad väiksema osakaalu poolest teistest kategooriatest.

Tabel 5. Ülevaade külastuste arvust ja osakaalust tegevuskoha kategooriate lõikes argipäevadel ja nädalavahetustel.

Kategooria	Argipäev		Nädalavahetus		Kokku
	Külastuste arv	Osakaal %	Külastuste arv	Osakaal %	
Kodu	1737	77	519	23	2256
Töö/kool	1367	94	84	6	1451
Teine kodu	380	61	246	39	626
Sõbra/sugulase kodu	263	52	238	48	501
Restoran/baar	601	78	172	22	773
Pood	744	69	340	31	1084
Vaba aeg	88	60	58	40	146
Sport	271	74	93	26	364
Transport	118	63	68	37	186
Puhkus	161	55	133	45	294

Kategooriate võrdlemisel joonisel 23 selgub, et teistest kategooriatest madalama nädalavahetuste külastuste osakaaluga on kodu ja töö kategooriad. Kodu puhul on madal külastuste osakaal nädalavahetustel selgitatav kahe asjaoluga – 1) inimesed veedavad enam aega kodus, mistõttu külastuse toimumist registreeritakse vähem; 2) inimesed külastavad teist kodu, mistõttu külastuse toimumist koju ei registreerita. Teistest kategooriatest madalama külastussagedusega jäävad veel silma restoranide-baaride kategooria ning poed ja ostukeskused, kus pooltel juhtudel jääb külastuste osakaal vahemikku 6%–36% ja 25%–35%.



Joonis 23. Külastuste osakaalu varieerumine tegevuskohtade lõikes nädalavahetusel.

Tegevuskohtade omavahelise võrdlemise tulemused Friedman'i testiga osutusid statistiliselt oluliseks ($p < 0.05$, hiir-ruut=41.12), mistõttu viidi läbi kategooriate vahel paarikaupa võrdlemine Wilcoxon'i testiga. Statistiliselt olulised tulemused ($p < 0.05$) on esitatud tabelis 6. Nagu tabeli 6 ja joonise 23 vaatlemisest selgub, on päevatüübist

põhjastatud külastuste osakaal kodu, töö ja restoranide-baaride puhul kõige enam erinevusi põhjustav. Nimetatud kategooriates on nädalavahetustel statistiliselt olulisel määral väiksem külastuste osakaal, võrreldes teiste kategooriatega. Töö-kooli kategooria eristus kõikidest teistest kategooriatest, kodu puhul ei suudetud erinevusi tõestada vaid restorani ja spordi kategooriaga ning restoranide-baaride puhul kodu ja spordiga seotud tegevuskohtade puhul. Samuti võib märgata poekategooria suuremat statistiliselt oluliste erinevuste hulka teiste kategooriatega – erinevusi ei suudetud tõestada vaid vaba aja, spordi ja transpordi kategooriatega, kõikide teiste kategooriate puhul ilmnisid statistiliselt olulised erinevused.

Tabel 6. Tunnuse “külastuspäeva tüüp” koondtabel statistiliselt olulistest ($p < 0.05$) paarikaupa võrdlemise tulemustest nädalavahetusel.

Kategooria	Erinevused	Erinevusi ei leitud
Kodu	pood, puhkus, sõbra/sugulase kodu, teine kodu, töö, transport, vaba aeg	restoran, sport
Töö	pood, puhkus, restoran, sõbra/sugulase kodu, sport, teine kodu, kodu, transport, vaba aeg	
Teine kodu	pood, restoran, kodu, töö	sõbra/sugulase kodu, vaba aeg, sport, transport, puhkus
Sõbra/sugulase kodu	pood, restoran, kodu, töö	teine kodu, vaba aeg, sport, transport, puhkus
Restoran	pood, puhkus, sõbra/sugulase kodu, teine kodu, töö, transport, vaba aeg	kodu, sport
Pood	kodu, restoran, sõbra/sugulase kodu, teine kodu, töö, puhkus	vaba aeg, sport, transport
Vaba aeg	restoran, töö	kodu, teine kodu, sõbra/sugulase kodu, pood, sport, transport, puhkus
Sport	töö	kodu, teine kodu, sõbra/sugulase kodu, restoran, pood, vaba aeg, transport, puhkus
Transport	kodu, restoran, töö	teine kodu, sõbra/sugulase kodu, pood, vaba aeg, sport, puhkus
Puhkus	kodu, pood, restoran, töö	teine kodu, sõbra/sugulase kodu, vaba aeg, sport, transport

5.2.2 Külastuse algusaeg kui iseloomulik tunnus

Külastuse algusaega vaadeldakse käesolevas töös kui pidevat ajalist tunnust, mis on indiviidide lõikes iga tegevuskoha kohta keskmistatud. Teooriale tuginedes võib eeldada erinevusi külastuse algusaegades argipäeviti ning nädalavahetuseti, mistõttu statistiline analüüs teostati nende perioodide kohta eraldi.

Tabelis 7 on summaarne ülevaade iga tegevuskoha külastuse algusaja keskmise ning mediaani kohta nädalavahetustel ning argipäeviti. Kategooriate lõikes võib märgata, et argipäeviti külastatakse enamikke tegevuskohti pärast tööpäeva lõppu. Erandiks on siinkohal toitlustusasutused nagu restoranid, mille varasem külastusaeg on tõenäoliselt seotud lõunapausidega. Vastupidiselt argipäevadele, mil inimeste päevane aeg on seotud töö- ja koolikohustustega, on nädalavahetustel märgata enamike

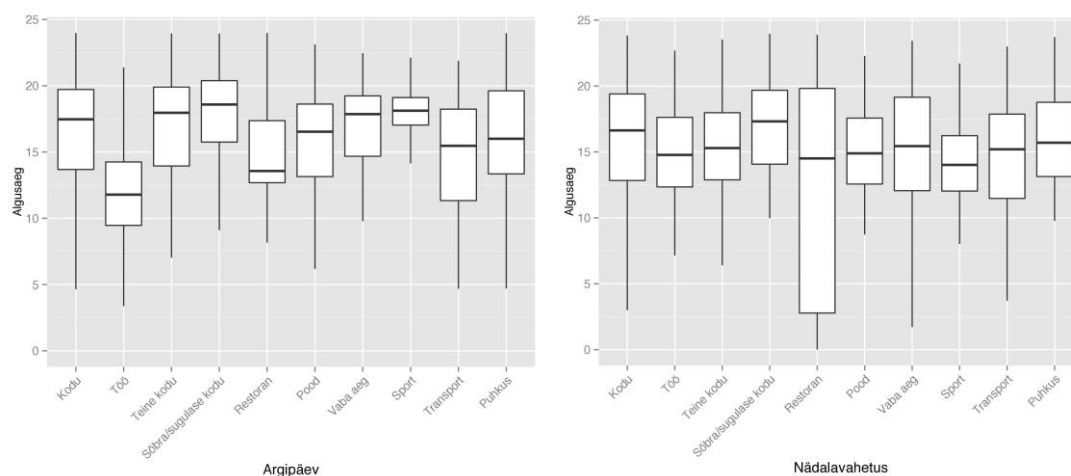
kategooriate puhul varasemat külastuse algusaega, mis viitab suuremale vabadusele tegevusi planeerida kogu päeva lõikes.

Tabel 7. Tegevuskohtadega seotud külastuste algusaja keskmine ning mediaan argipäeviti ning nädalavahetustel.

Kategooria	Argipäev		Nädalavahetus	
	Keskmine	Mediaan	Keskmine	Mediaan
Kodu	16.26	17.47	15.28	16.63
Töö	12.22	11.78	15.14	14.78
Teine kodu	16.61	17.96	15.24	15.29
Sõbra/sugulase kodu	17.47	18.58	16.27	17.32
Restoran/baar	14.68	13.57	12.55	14.51
Pood	16.00	16.53	15.03	14.89
Vaba aeg	16.82	17.86	15.17	15.44
Sport	17.42	18.12	14.09	14.02
Transport	14.59	15.47	14.67	15.20
Puhkus	16.02	16.00	15.62	15.70

Kategooriate lähemal vaatlusel (joonis 24) selgub, et argipäeviti omavad võrdlemisi sarnast külastuse algusaega kodu, teise kodu, sõbra/sugulase kodu ning puhkusega seonduvad kohad – 50% kõikide külastuste algusaegadest jääb nende kategooriate puhul kella 13 ja 20 vahele. Toitlustusasutused, poed, vaba aja, transpordi ning spordiga seotud kohad moodustavad teise grupi, mille külastuste algusajad jäävad pooltel juhtudel vahemikku 11 – 19. Mainitud tegevuskohtadest varasema algusajaga eristuvad selgelt töö ja kooliga seonduvad kohad, mis on argipäeviti ootuspärane tulemus.

Nädalavahetustel on külastuste algusajad kategooriate lõikes sarnasemad, kui argipäeviti (joonis 24). Teistest eristub restoranide – baaride kategooria, mille puhul võib märgata suuri variatsioone algusaegades.



Joonis 24. Külastuste algusaegade varieerumine argipäeviti ning nädalavahetustel.

Gruppide omavahelisel võrdlemisel Friedman'i testiga selgub, et statistiliselt olulised ($p < 0.05$) erinevused tegevuskoha kategooriate vahel ilmnevad vaid argipäeviti ($p < 0.05$, hii-ruut=36.91). Tulemust kinnitavad ka juba eelnevalt vaadeldud karpdiagrammid. Wilcoxon'i paarikaupa võrdlemise test viidi seetõttu läbi vaid argipäeva külastustega ning statistiliselt olulised tulemused ($p < 0.05$) on esitatud tabelis 8.

Tabel 8. Tunnuse "külastuse algusaeg" koondtabel statistiliselt olulistest ($p < 0.05$) paarikaupa võrdlemise tulemustest argipäevadel.

Kategooria	Erinevused	Erinevusi ei leitud
Kodu	restoran, töö	teine kodu, puhkus, sõbra/sugulase kodu, pood, sport, transport, vaba aeg
Töö	pood, puhkus, restoran, sõbra/sugulase kodu, sport, teine kodu, kodu, vaba aeg	
Teine kodu	restoran, töö, transport	kodu, sõbra/sugulase kodu, sport, pood, vaba aeg, puhkus
Sõbra/sugulase kodu	pood, puhkus, restoran, töö, transport	vaba aeg, kodu, sport, teine kodu
Restoran/baar	pood, kodu, sõbra/sugulase kodu, teine kodu, töö, sport, vaba aeg	puhkus
Pood	restoran, sõbra/sugulase kodu, töö	teine kodu, kodu, vaba aeg, sport, transport, puhkus
Vaba aeg	restoran, töö	sport, kodu, teine kodu, sõbra/sugulase kodu, puhkus, pood, transport
Sport	restoran, töö	vaba aeg, kodu, teine kodu, sõbra/sugulase kodu, puhkus, pood, transport
Puhkus	sõbra/sugulase kodu, töö	kodu, teine kodu, restoran, pood, vaba aeg, sport, transport
Transport	sõbra/sugulase kodu, teine kodu	kodu, töö, restoran, pood, vaba aeg, sport, puhkus

5.2.3 Külastuse kestus kui iseloomulik tunnus

Sarnaselt külastuse algusajale on külastuse kestus pidev tunnus, mida antud analüüsis on mõõdetud minutites. Gruppide omavahelise võrdlemise jaoks on iga uuringus osaleja vastava kategooria külastuse kestus keskmistatud. Külastuse kestuseid vaadeldakse eraldi argipäeviti ning nädalavahetustel.

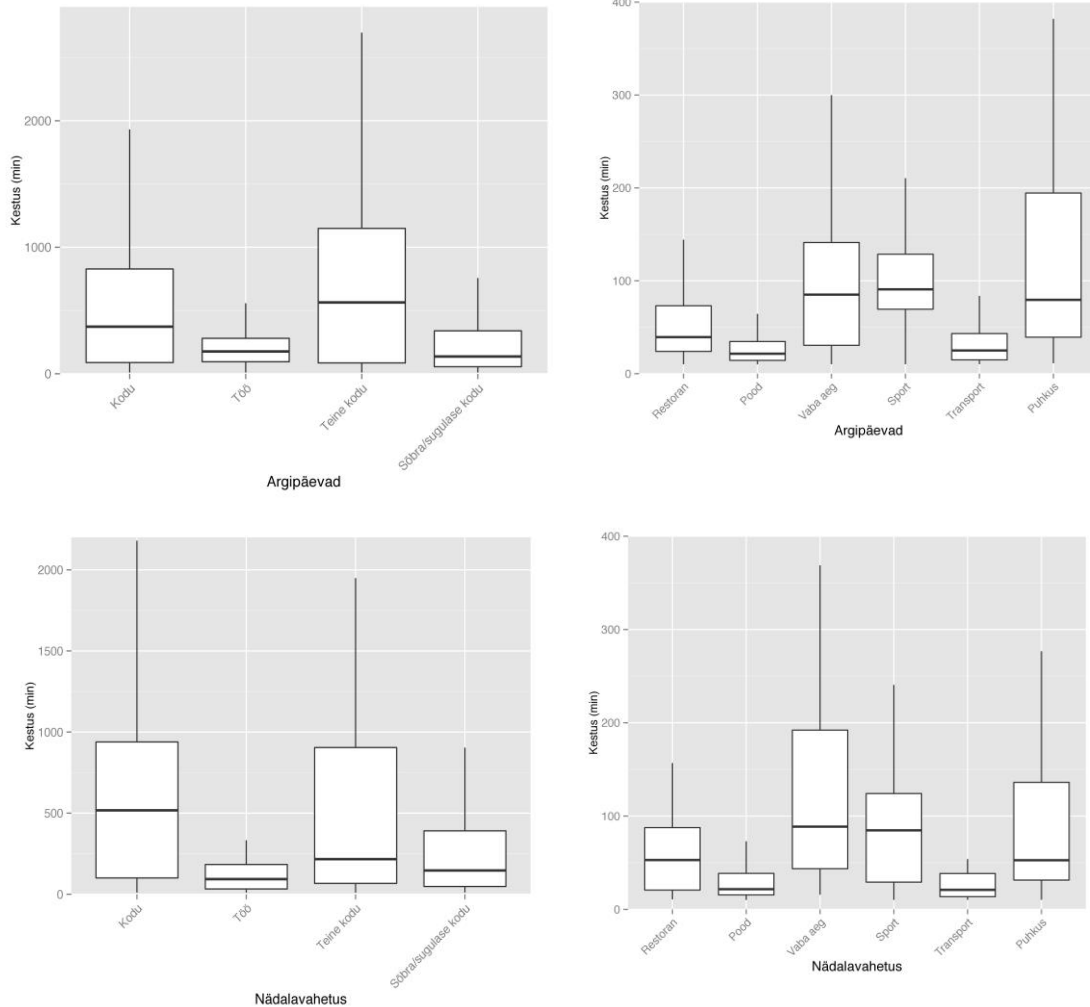
Tabel 9 toob välja esmase ülevaate analüüsis kasutatava tunnuse väärtuste varieerumisest kategooriate lõikes nädalavahetustel ning argipäevadel. Tegevuskoha külastuse keskmised annavad võimaluse ka hinnata tegevuskohtade tuvastamiseks kasutatud meetodite ja nende parameetrite valikutäpsust. Antud juhul võib öelda, et tegevuskohtade külastuste pikkused võiksid olla kooskõlas reaalsusega. Töö/kooli kategooria puhul on lühike külastus tõenäoliselt tingitud tööaja sisse tekkivatest pausidest nagu lõuna-aeg, samuti osalevad uuringus üliõpilased, kelle kooliga seotud tegevuskohad asuvad erinevates asukohtades, mis külastuse pikkuse lühemaks muudab.

Tabel 9. Külastuste keskmine ja mediaan tegevuskoha kategooriate lõikes argipäeviti ja nädalavahetustel.

Kategooria	Argipäevad		Nädalavahetus	
	Keskmine	Mediaan	Keskmine	Mediaan
Kodu	535.74	372.28	638.37	517.29
Töö	235.65	176.54	185.56	93.33
Teine kodu	896.83	563.87	639.00	216.37
Sõbra/sugulase kodu	299.76	136.50	291.76	146.60
Restoran/baar	63.55	39.35	73.18	52.90
Pood	32.45	21.48	33.61	21.63
Vaba aeg	106.07	85.10	188.41	88.68
Sport	93.08	90.79	103.01	84.71
Transport	46.79	24.97	83.98	20.92
Puhkus	197.44	79.49	158.66	52.67

Jooniselt 25 võib märgata, et külastuste kestustes tekivad argipäeviti kategooriate vahel mõnel juhul erinevused. Teistest kategooriates eristuvad poed ning restoranid, kus pooled külastused jäävad vahemikku 14-28 minutit ning 17-51 minutit (vastavalt), samuti transpordiga seotud kohad, kus pooled külastused on 14-44 minutit pikad. Pikema keskmise külastusega tegevuskohad nagu kodu ning teine kodu erinevad teistest tegevuskoha kategooriatest märgatavalt – kodu puhul on pooled külastused pikkusega 89-829 minutit ning teise kodu puhul 85-1149 minutit. Sõbra või sugulase kodu ja respondendi enda töökoha külastuste pikkused on argipäeviti võrdlemisi sarnase väärtusvahemikuga – 56-339 minutit ja 95-280 minutit (vastavalt).

Nädalavahetustel ilmnevad argipäevadega võrreldes sarnased tendentsid – teistest kategooriatest eristuvad poed ning restoranid, kus külastuse pikkused (50%) jäävad 15-39 ja 21-88 minuti piiridesse, transpordiga seotud tegevuskohtade puhul on see vahemik 14-40 minutit. Tegevuskohad nagu kodu, töö, teine kodu ja sõbra-sugulase kodu omavad nädalavahetustel samuti märgatavalt pikemaid külastusi – 100-940 minutit, 31-186 minutit, 65-905 minutit ja 48-393 minutit (vastavalt). Võib märgata, et külastusaegade pikkustes on kattuvusi, mis võivad tunnuse iseloomulikkust mõjutada.



Joonis 25. Külastuse pikkuse variatsioonid erinevates tegevuskoha kategooriates nädalavahetustel ja argipäevadel. Selguse mõttes on pikema keskmise külastusega kategooriad tõstetud eraldi graafikule. Autor juhib tähelepanu sellele, et vasak- ja parempoolsete graafikute puhul on y-teljed erineva ulatusega.

Friedman'i test tõestas kategooriate vahelisi statistiliselt olulisi erinevusi nädalavahetadel ($p=0.00001$, hii-ruut=59.81). Tabelis 10 on esitatud Wilcoxon'i paarikaupa võrdlemise tulemused. Selgub, et külastuse pikkus on parem tunnus kategooriate eristamiseks kui seda oli külastuse algusaeg – 7 kategooria puhul oli see tunnus võrdlemisi iseloomulik, erinevusi ei suudetud tõestada vaid ühe-kahe teise kategooriaga. Poe ning restoranide-baaride puhul võib külastuse pikkus kattuda transpordiga seotud kohtadega, kodu puhul teise koduga, töö puhul sõbra/sugulase kodu ja puhkusega seotud kohtadega, teise kodu ja sõbra/sugulase kodu puhul kodu ning puhkuse ja tööga (vastavalt). Kõige enam sarnasusi teiste kategooriatega ilmnis puhkuse, transpordi ning spordiga seotud kategooriates, kuid sarnaste kategooriate hulk oli siiski väiksem kui külastuse algusaja tunnuse puhul.

Tabel 10. Tunnuse “külastuse kestus” koondtabel statistiliselt olulistest ($p < 0.05$) paarikaupa võrdlemise tulemustest argipäevadel.

Kategooria	Erinevused	Erinevusi ei leitud
Kodu	pood, puhkus, restoran, sõbra/sugulase kodu, sport, töö, vaba aeg, transport	teine kodu
Töö	pood, restoran/baar, sport, teine kodu, kodu, transport, vaba aeg	sõbra/sugulase kodu, puhkus
Teine kodu	vaba aeg, transport, sport, restoran, pood, puhkus, sõbra/sugulase kodu, töö	kodu
Sõbra/sugulase kodu	pood, restoran, kodu, teine kodu, sport, transport, vaba aeg	puhkus, töö
Restoran/baar	pood, puhkus, kodu, sõbra/sugulase kodu, teine kodu, töö, vaba aeg	transport
Pood	kodu, restoran, sõbra/sugulase kodu, teine kodu, töö, puhkus, sport, vaba aeg	transport
Vaba aeg	transport, pood, restoran/baar, kodu, sõbra/sugulase kodu, teine kodu, töö	sport, puhkus
Sport	kodu, pood, sõbra/sugulase kodu, teine kodu, töö, transport	puhkus, vaba aeg, restoran
Transport	puhkus, kodu, sõbra/sugulase kodu, teine kodu, töö, vaba aeg	pood, restoran, sport
Puhkus	pood, transport, kodu, restoran/baar, teine kodu	töö, sõbra/sugulase kodu, vaba aeg, sport

Nädalavahetustel ilmsid Friedman'i testis sarnaselt argipäevadega statistiliselt olulised erinevused tegevuskoha kategooriate vahel ($p < 0.05$, hii-ruut=43.9). Wilcoxon'i testi statistiliselt olulised tulemused ($p < 0.05$) on esitatud tabelis 11. Võrreldes argipäevadega, ilmsid nädalavahetusel statistiliselt olulisi erinevusi paaride vahel mõnevõrra vähem – vastavalt 35 ja 31 erinevuste paari. Töö kategoorias jäi tegevuskohtade vahelisi erinevusi vähemaks, mis oli ka ootuspärane tulemus, arvestades nädalavahetustel töö ja kooliga seonduvate kohtade külastuspikkuse vähenemist. Külastuspikkuse vähenemine muutis tööga seonduvad külastused ajaliselt sarnaseks teiste pikemate külastustega nagu sõbra-sugulase kodu ja puhkus. Tunnuse suuremat eristusvõimet võib vaadelda poe kategoorias ning samuti kodu, teise kodu, sõbra/sugulase kodu ning restoranide-baaride kategoorias, kus erinevusi ei suudetud tõestada vaid ühe tegevuskoha tüübiga.

Tabel 11. Tunnuse “külastuse kestus” koondtabel statistiliselt olulistest ($p < 0.05$) paarikaupa võrdlemise tulemustest nädalavahetusel.

Kategooria	Erinevused	Erinevusi ei leitud
Kodu	pood, puhkus, restoran, sõbra/sugulase kodu, sport, töö, vaba aeg, transport	teine kodu
Töö	pood, restoran/baar, teine kodu, kodu	sõbra/sugulase kodu, vaba aeg, sport, transport, puhkus
Teine kodu	pood, puhkus, restoran/baar, sõbra/sugulase kodu, sport, töö, transport, vaba aeg	kodu
Sõbra/sugulase kodu	pood, restoran, kodu, teine kodu, sport, transport, vaba aeg, puhkus	töö
Restoran/baar	pood, puhkus, kodu, sõbra/sugulase kodu, teine kodu, töö, vaba aeg	transport
Pood	kodu, restoran, sõbra/sugulase kodu, teine kodu, töö, puhkus, sport, vaba aeg, transport	
Vaba aeg	pood, restoran/baar, kodu, sõbra/sugulase kodu, teine kodu	sport, puhkus, töö, transport
Sport	kodu, pood, sõbra/sugulase kodu, teine kodu	puhkus, vaba aeg, restoran, töö, transport
Transport	pood, kodu, sõbra/sugulase kodu, teine kodu	töö, vaba aeg, restoran, sport, puhkus
Puhkus	pood, kodu, restoran/baar, teine kodu, sõbra/sugulase kodu, sport	töö, transport, vaba aeg

5.2.4 Mobiilikasutus kui iseloomulik tunnus

Käesoleva peatüki eesmärk on välja selgitada ja kirjeldada mobiilikasutuse kui külastustunnuse iseloomulikkus uuringus osalevate inimeste näitel, võttes arvesse tähtsamad kohakategooriad, mis inimese igapäevaste tegevuskohtadena arvesse tulevad. Mobiilikasutuse tunnustena on arvatud igale külastusele ekraani sisselülitamise sagedus tunnis, ekraani lahtilukustamise sagedus tunnis, telefoni aktiivse kasutamise kestus (mediaan iga külastuse jaoks) ning aktiivse mobiilikasutuse summaarse kestuse osakaal külastuse kestusest.

Tabelis 12 esitatud üldine kirjeldav statistika mobiilisündmuste kohta. Kõikide mobiilisündmuste ja tegevuskoha kategooriate puhul võib märgata mediaani võrdlemisi suurt erinevust keskmisest, mis viitab kõrvalekalletest normaaljaotusest. Eelnevast tulenevalt kasutatakse statistilises analüüsis respondentide lõikes mediaane, mitte keskmisi.

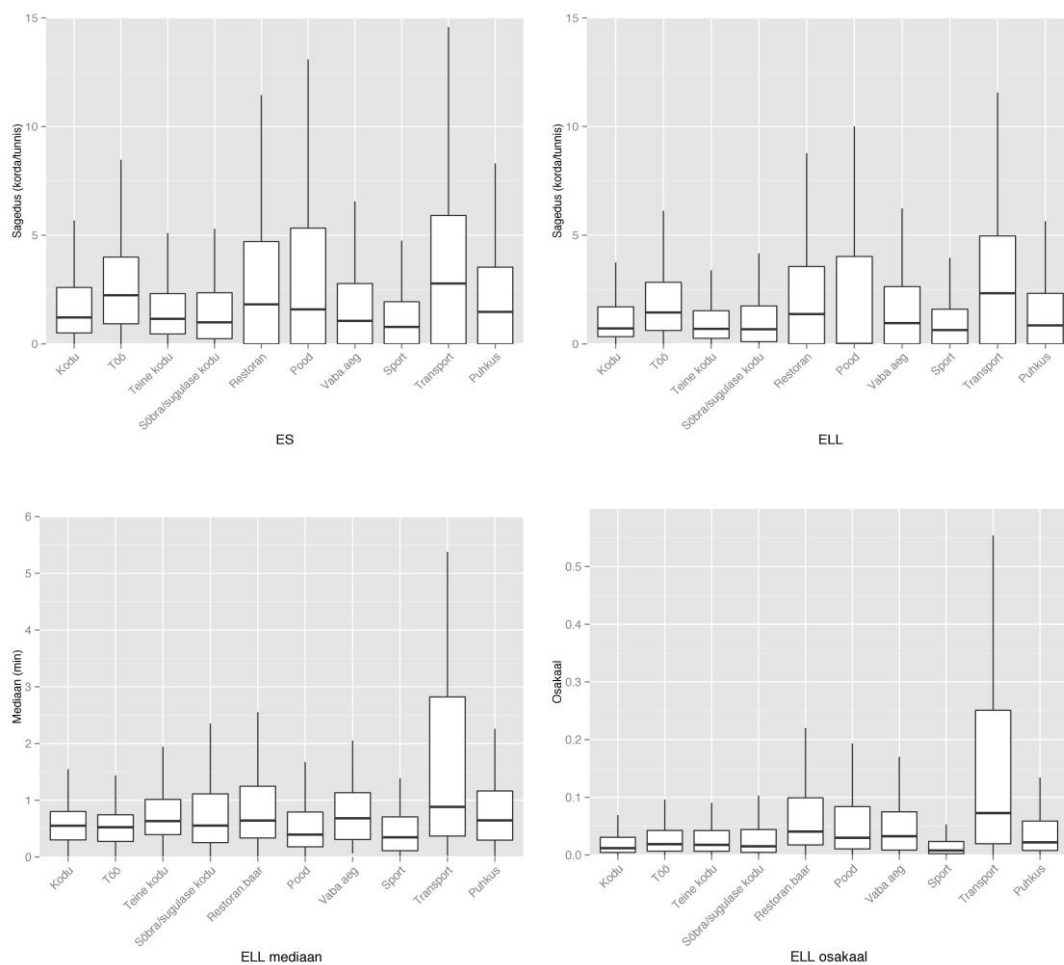
Ekraani sisselülitamise sagedus on suurim tööl (2.24 korda tunnis), transpordiga seotud kohtades (2.78) ning restoranides ja baarides (1.82). Kõige väiksem on see aga spordiga seotud kohtades (1.77) ning sõbra-sugulase kodus (0.99). Ekraani lahtilukustamise sagedus järgib sisselülitamisega sarnaseid tendentse, olles suurim transpordiga seotud kohtades (2.33 korda tunnis), tööl (1.44) ja restoranides-baarides (1.37) ning väiksem poes (0.00) ning spordiga seotud kohtades (0.64). Aktiivse telefonikasutuse kestus on suurim transpordiga seotud kohtades (0.88 min), vaba ajaga seotud kohtades (0.68) ning restoranides-baarides ja puhkusega seotud kohtades (0.45) ning väiksem poes (0.39) ja spordiga seotud kohtades (0.35). Telefoni aktiivse

kasutamise osakaal on suurim transpordiga seotud kohtades (0.07) ja restoranides-baarides (0.04) ning väikseim kodus (0.01), sõbral või sugulasel külas olles (0.01) ja spordiga seotud kohtades (0.01).

Tabel 12. Üldine kirjeldav statistika mobiilisündmuste kohta. ES – ekraani sisselülitamise sagedus tunnis, ELL – ekraani lahtilukustamise sagedus tunnis, ELL_k – aktiivse telefonikasutuse kestuse mediaan tegevuskohas, ELL_ok – ELL_k sündmuse kestuse osakaal külastuse ajast.

Kategooria	ES		ELL		ELL_k (min)		ELL_ok	
	Keskm.	Med.	Keskm.	Med.	Keskm.	Med.	Keskm.	Med.
Kodu	2.31	1.22	1.43	0.71	0.77	0.55	0.03	0.01
Töö	3.05	2.24	2.08	1.44	0.75	0.52	0.07	0.02
Teine kodu	2.01	1.15	1.29	0.69	1.09	0.63	0.05	0.02
Sõbra/sugulase kodu	2.06	0.99	1.33	0.67	1.25	0.55	0.05	0.01
Restoran/baar	3.14	1.82	2.46	1.37	1.23	0.64	0.09	0.04
Pood	3.48	1.59	2.53	0	0.96	0.39	0.12	0.03
Vaba aeg	2.2	1.06	1.75	0.96	0.98	0.68	0.07	0.03
Sport	1.77	0.78	1.27	0.64	0.66	0.35	0.03	0.01
Transport	4.24	2.78	3.28	2.33	3.38	0.88	0.38	0.07
Puhkus	2.73	1.47	1.77	0.85	1.05	0.64	0.06	0.02

Ekraani sisselülitamise ja lahtilukustamise puhul võib märgata restoranide-baaride, poodide ning transpordiga seotud kohtade eristumist suuremate väärtusvahemike poolest (joonis 26). Teisalt võib märgata, et ekraani sisselülitamise puhul on poe kategoorias 50% külastuste ajal sagedus olnud null. Transpordiga seotud kohad eristuvad teistest kategooriatest lisaks eelnevale ka suurema aktiivse telefonikasutuse osakaalu väärtusvahemiku poolest ning telefoni aktiivse kasutamise mediaani poolest. Spordiga seotud kohtades on telefoni aktiivse kasutamise osakaal, võrreldes teiste tegevuskohtadega, madalam. Ükski kategooria aga ei oma mobiilikasutustunnuste lõikes väärtusvahemikku, mis ei kattuks mõne teise tegevuskoha väärtusvahemikuga.



Joonis 26. Telefonikasutuse varieerumine vaadeldavates kategooriates.

Tegevuskoha kategooriate omavaheliste mobiilikasutustunnuste statistiliste erinevuste väljaselgitamiseks kasutati Friedman'i (tabel 13) ning Wilcoxon'i testi (tabel 14). Kõikide käsitletud mobiilikasutustunnuste puhul esinesid tegevuskoha kategooriate vahel statistiliselt olulised erinevused ($p < 0.05$).

Tabel 13. Friedman'i testi tulemused tegevuskoha kategooriate vaheliste erinevuste leidmiseks mobiilikasutustunnuste lõikes. ES – ekraani sisselülitamise sagedus tunnis, ELL – ekraani lahtilukustamise sagedus tunnis, ELL_k – aktiivse telefonikasutamise mediaan, ELL_ok – ELL_k sündmuse kestuse osakaal külastuse ajast.

Tunnus	Respondentide arv	Hii-ruut	Olulisustõenäosus
ES	14	29.617	0.001
ELL	14	25.380	0.003
ELL_k	14	26.159	0.001
ELL_ok	14	23.687	0.01

Ekraani sisselülitamise sagedus on mobiilikasutustunnus, mis tõi välja gruppide vahel kõige enam erinevusi (tabel 14). Uuritud kümnest kategooriast oli transpordi kategoorial teiste gruppidega kõige enam statistiliselt olulisi erinevusi. Selgus, et

transpordiga seotud kohtades lülitatakse ekraani sisse enam kui poes, teises kodus, spordiga seotud kohtades, kodus ning sõbra või sugulase kodus. See võib olla seotud vajadusega kella kontrollida, mis on transpordiga seotud kohtades võrdlemisi arusaadav tegevus.

Tööl olles on ekraani sisselülitamise sagedus suurem kui sõbra või sugulase kodus, spordiga seotud kohtades, kodus ning teises kodus. Restoranis viibides seevastu lülitatakse ekraani sisse sagedamini kui puhkusega seotud tegevuskohtades, sõbra või sugulase kodus ning kodus.

Ekraani lahtilukustamine on statistiliselt olulisel määral sagedam transpordiga seotud kohtades kui puhkusega seotud kohtades, sõbra või sugulase kodus, spordiga seotud kohtades, teises kodus ja kodus. Restoranis lukustatakse ekraani lahti enam kui puhkusega seotud kohtades, kodus ning sõbra või sugulase kodus. Tööl, sarnaselt restoraniga, enam kui puhkusega seotud kohtades ning sõbral või sugulasel külas olles.

Aktiivse telefonikasutuse kestus ei osutunud tunnuseks, mis kategooriate vahel palju erinevusi oleks välja toonud. Erinevus ilmnis vaid poe kategoorias, kus sündmuse kestus oli väiksem kui transpordiga seotud kohtades, teises kodus ja sõbra või sugulase kodus. Kui arvestada ühe külastuse ajal aktiivse telefonikasutuse osakaalu, siis selgus, et statistiliselt olulised erinevused ilmnevad poe ning teise kodu vahel – poes on ekraani lahtioleku summaarse kestuse osakaal väiksem kui teises kodus. Samuti on statistiliselt oluliselt väiksem vaadeldava tunnuse osakaal spordi kategoorias, võrreldes töö ning teise koduga.

Tabel 14. Mobiilisündmuste koondtabel statistiliselt olulistest ($p < 0.05$) paarikaupa võrdlemise tulemustest.

Kategooria	Erinevused	Erinevusi ei leitud
Ekraani sisselülitamise sagedus		
Transport	Pood, teine kodu, sport, kodu, sõbra/sugulase kodu	töö, restoran/baar, vaba aeg, puhkus
Sõbra/sugulase kodu	Töö, restoran, transport	Kodu, teine kodu, pood, vaba aeg, sport, puhkus
Pood	Transport	Kodu, töö, teine kodu, sõbra/sugulase kodu, restoran, vaba aeg, sport, puhkus
Puhkus	Restoran	Kodu, töö, teine kodu, sõbra/sugulase kodu, pood, vaba aeg, sport, transport
Restoran	Puhkus, sõbra/sugulase kodu, kodu	Töö, teine kodu, pood, vaba aeg, sport, transport
Teine kodu	Transport, töö	Kodu, sõbra/sugulase kodu, restoran, pood, vaba aeg, sport, puhkus
Sport	Töö, transport	Kodu, teine kodu, sõbra/sugulase kodu, restoran, pood, vaba aeg, puhkus
Kodu	Transport, töö, restoran	Teine kodu, sõbra/sugulase kodu, pood, vaba aeg, sport, puhkus
Töö	Sõbra/sugulase kodu, sport, kodu, teine kodu	Restoran, pood, vaba aeg, transport, puhkus

Ekraani lahtilukustamise sagedus		
Puhkus	Töö, restoran/baar, transport	Kodu, teine kodu, sõbra/sugulase kodu, pood, vaba aeg, sport
Töö	Puhkus, sõbra/sugulase kodu	Kodu, teine kodu, restoran/baar, vaba aeg, sport, transport.
Restoran/baar	Puhkus, kodu, sõbra/sugulase kodu	Kodu, töö, teine kodu, pood, vaba aeg, sport, transport
Transport	Puhkus, sõbra/sugulase kodu, sport, teine kodu, kodu	Töö, restoran/baar, vaba aeg, pood
Sõbra/sugulase kodu	Transport, töö, restoran	Kodu, töö, teine kodu, restoran/baar, vaba aeg, sport, puhkus
Sport	Transport	Kodu, töö, teine kodu, sõbra/sugulase kodu, restoran/baar, pood, puhkus
Kodu	Restoran/baar, transport	Töö, teine kodu, sõbra/sugulase kodu, vaba aeg, sport, puhkus
Teine kodu	Transport	Kodu, töö, teine kodu, sõbra/sugulase kodu, restoran/baar, pood, vaba aeg, sport, puhkus
Telefoni aktiivse kasutuse kestus (mediaan)		
Pood	Transport, teine kodu, sõbra/sugulase kodu	Kodu, töö, restoran/baar, vaba aeg, sport, puhkus
Transport	Pood	Kodu, töö, teine kodu, sõbra/sugulase kodu, restoran, vaba aeg, sport, puhkus
Teine kodu	Pood	Kodu, töö, transport, sõbra/sugulase kodu, restoran, vaba aeg, sport, puhkus
Sõbra/sugulase kodu	Pood	Kodu, töö, teine kodu, restoran, vaba aeg, sport, puhkus, transport
Telefoni aktiivse kasutamise osakaal külastusest		
Pood	Teine kodu	Kodu, töö, transport, sõbra/sugulase kodu, restoran, vaba aeg, sport, puhkus
Sport	Töö, teine kodu	Kodu, sõbra/sugulase kodu, restoran/baar, pood, vaba aeg, transport, puhkus
Töö	Sport	Kodu, teine kodu, sõbra/sugulase kodu, restoran/baar, pood, vaba aeg, transport, puhkus
Teine kodu	Sport	Kodu, töö, sõbra/sugulase kodu, restoran/baar, pood, vaba aeg, transport, puhkus

6. Arutelu ja järeldused

Mobiiltelefonide, eeskätt nutitefonide kasutamine geograafilistes uuringutes omab mitmeid eeliseid traditsiooniliste andmekogumisvahendite ees (Raento 2009). Mitmekülgsete käitumis- ning asukohaandmete automatiseeritud kogumisvõimalus ja seadme võrdlemisi odav hind muudavad nutitefonide kasutamise inimekäitumise uurimisel perspektiivseks vahendiks, mis üha enam populaarsust kogub (Nitsche et al. 2013, Higuera de Furtos 2014). Üheks olulisemaks omaduseks, mida mobiiltelefonide puhul tasub tähele panna, on nende seotus inimese igapäevaeluga. See teeb võimalikuks pikemate ajaperioodide täpse uurimise, samal ajal avaldades vähem koormust respondentile. Teisalt seab respondentide passiivsem osalus uuringutes uurijale endale suurema koormuse eeskätt just kasuliku informatsiooni kättesaamiseks kogutud andmetest. Informatsioon nagu reisimiseks kasutatav transpordiliik või tegevuste toimumiskohad ja nende kohtade semantika on oluline ruumilise käitumise uurimise seisukohalt, kuid pikema perioodi vaatlemisel ei pruugi respondentilt saadav lisainformatsioon olla piisav ning vajaliku täpsusega. Eelnevast tulenevalt kasvab uuringute arv, mis püüavad kasulikku informatsiooni kogutud andmetest tuvastada automaatselt, ilma respondendi olulise panuseta (Bolbol et al. 2012), püüdes minimaliseerida respondentile langevat koormust ning võimalikke vigu respondendi poolt esitatud andmetes (Wolf et al. 2001).

Käesolevas töös võeti eesmärgiks tuvastada poole aasta jooksul kogutud YouSense andmestikust inimesele tähenduslikud kohad ning siduda nende kohtadega peamised külastustunnused nagu külastuse kestus, külastuse algusaeg ja nädalapäev ning mobiilkasutustunnused. Statistilise analüüsi etapis hinnati, millised vaadeldud tunnustest omavad iseloomulikkust erinevate kohakategooriate lõikes ning kui perspektiivikas on nende tunnuste kasutamine klassifikatsioonimudelite muutujatena.

6.1 Tegevuskohtade tuvastamine

Tegevuskohtade tuvastamiseks YouSense rakenduse poolt kogutud asukohaandmestikust kasutati antud töös kahte meetodit: tihedusel põhinevat klasteranalüüsi DBSCAN ja ajaparameetriga klasterdamist. Viimase jaoks kirjutati autori poolt Pythoni skript. Tulemuste analüüs näitas, et tihedusel põhinev klasteranalüüsi meetod ei võimalda YouSense andmetest tuvastada suurt hulka respondentile tähenduslikke tegevuskohti, mistõttu täiendava meetodi kasutamine tegevuskoha tuvastusprotsessis on YouSense andmete puhul vajalik. Ajaparameetrit sisaldava meetodi kasutamine osutus antud juhul võrdlemisi tõhusaks vahendiks, suutes kategooriates nagu pood ja restoran/baar tuvastada ligikaudu pooled DBSCANi poolt tuvastamata jäänud tegevuskohad. Ajaparameetriga meetodi tõhusus võib YouSense andmestiku puhul tuleneda mitmest põhjusest. Esiteks on tegu GPS andmestikuga, kus saab oluliseks GPS satelliidi signaali kättesaadavus nutitefonis

paiknevale sensorile. Signaali kättesaamist võivad mõjutada mitmed asjaolud nagu hoonetes viibimine (Nurmi 2009) või telefoni kotis kandmine (Nitsche et al. 2013). Teiseks lakkab YouSense rakenduse asukohainfo kogumine juhul kui inimene ei liigu, mis muudab harvem külastatavate ning vähest liikumist tingivate kohtade tuvastamise vaid tihedusparameetritega keeruliseks ülesandeks. Samas tuleb tähele panna, et harvem külastatavaid tegevuskohti nagu teater ja kino on ka arvuliselt oluliselt vähem, mistõttu suur "aukude" meetodi osakaal sellistes kohtades võib olla tingitud ka juhusest.

Tegevuskohtade tuvastamisel nimetatud meetodite kombinatsiooniga leiti iga vaadeldud respondendi 6 kuu asukohaandmestikust kohad, mida uuringus osalejad on selle aja jooksul vähemalt korra külastanud. Kuivõrd puudub informatsioon selle kohta, kui paljud tegevuskohtadest on jäänud tuvastamata, siis tugineb töö autor kirjanduses olevale informatsioonile, hindamaks algoritmide efektiivsust. Leitud kohtade hulk sarnaneb Do ja Gatica-Perez (2013) GPS andmete-põhises töös leitud päevasele ning 14 kuu tegevuskohtade hulga, teisalt aga on mõnevõrra väiksem Järv ja teiste (2014) passiivse mobiilpositsioneerimise teel hinnatud tegevuskohtade hulgast. Kuna GPS andmed on oluliselt täpsemaks vahendiks inimeste tegevuskohtade vaatlemisel kui passiivse mobiilpositsioneerimise andmed (Do & Gatica-Perez 2013), võib tegevuskohtade tuvastusprotsessi edukust antud töös kinnitada.

Kuivõrd nelja respondendi puhul ei suutnud kasutatud meetodid eristada lähedalasuvat tööd ja kodu kaheks erinevaks klastriks võib eeldada, et sarnaseid probleeme võib esineda ka teiste tegevuskohtadega, mis asusid ruumis lähedal ning mida sageli külastatakse. Üheks selliseks näiteks võib olla Tartu Raekojaplats ja selle lähiümbrus, kus ruumiliselt koos asuvad kontorid, kohvikud ning ülikooli õppehooned. DBSCANi probleemiks on võrdlemisi suur tundlikkus kahe sisendparameetri suhtes (Milenova & Campos 2002) ning lokaalsete tihedusmuutustega mitteametamine (Ankerst et al. 1999), mistõttu võib alternatiivina kaaluda mõne hierarhilise klasteranalüüsi meetodi kasutamist, mis loob andmebaasi esmalt klastrite tihedusstruktuuri (Ankerst et al. 1999). Kaaluda võib ka respondentide jaotamist gruppidesse GPS punktide arvu alusel ning igale grupile seejärel määrata sobiv tihedusparameetrite komplekt.

Töö raames teostatud ruumiandmete töötlus klasteranalüüsi ja andmeaukude meetodil loob juurde mitmeid lisaetappe, mis muudavad andmetöötlusprotsessi võrdlemisi mahukaks ning aeganõudvaks. Neid kahte sammu on reaalsuse võimalus ühendada, defineerides distantsifunktsioon, mis sisaldab aega kaaluna. Töö autor on antud võimalust rakendanud Pythoni skripti näol väiksemat ajaperioodi hõlmavate andmete puhul. Pikema perioodi puhul osutub kood liiga aeglaseks, et seda ilma andmeid indekseerimata edukalt rakendada. Kuivõrd ELKI *open source* projekt võimaldab kasutada R-puu indekseerimist, on edaspidi võimalik ja soovitatav defineerida kõnealune distantsifunktsioon Java keskkonnas sobiva klassina, avades võimaluse

loobuda töö mitmesse etappi jagamisest, kasutades sealjuures vaid ühte algoritmi. See saab oluliseks eeskätt suurema hulga respondentide andmete töötlemisel ja analüüsimisel.

6.2 Külastustunnused ja nende iseloomulikkus tegevuskoha kirjeldamisel

Korruga erinevat tüüpi andmete kogumine ja kasutamine on kasvav trend, mille kasutamise vajadust nähakse eeskätt inimese asukohaandmete turvalisuse tagamise kontekstis (Liu et al. 2013). Järjest vähem püütakse inimese tegevuskohtade semantika tuvastusprotsessi siduda konkreetse geograafilise asukohaga ning kaardisobitamise (*map matching*) protseduuriga (Do & Gatica-Perez 2013), mis võib lisaks kahanevale asukohaturvalisusele anda tulemuseks ka kohmakaid ja ebatäpseid tõlgendusi (Nurmi 2009). Kuivõrd nutitelefonide sensorid võimaldavad koguda erinevat informatsiooni inimese käitumise kohta, on rõhuasetus nihkunud geograafiliselt asukohalt inimkäitumise mustrite uurimisele (Liu et al. 2013). Käesoleva töö raames vaadeldi peamiste kohakülastustunnuste nagu külastuse kestus, külastuse algusaeg ja nädalapäev variatsioone tegevuskohtade lõikes. Eraldi pöörati tähelepanu inimese mobiilikasutustunnuste variatsioonidele kui ühele võimalikule nutitelefonidega kogutavale ja tegevuskohti iseloomustavale tunnuskomplektile sellises andmestikus.

Leitud külastustunnuste väärtused ning nende tulemuste kooskõla reaalsusega on sõltuv mitmest asjaolust. Esiteks on külastustunnuste täpsus ja võrreldavus reaalse toimunud tegevusega suuresti tulenev klasteranalüüsi parameetrite valikust ja analüüsi lõpptulemusest. Liiga suured klastrid, kus sees on mitu tegevuskohta või tegevuskohad, mis on klasteranalüüsi käigus killustunud, ei võimalda adekvaatselt hinnata nendega seotud külastustunnuseid. Eeskätt mõjutavad mainitud tulemused just tegevuskohas veedetud aja hinnangut, mis võib vastavalt osutada ülehinnanguks või alahinnanguks. Teiseks omavad mitmed tegevuskohad mitut erinevat funktsiooni (suured keskused sisaldavad poode, aga ka mitmeid vaba aja ning sportimisega seotud võimalusi), mistõttu leitud külastustunnuseid ei pruugi kajastada tegevuskoha iseloomu kõige paremini. Antud töös leitud külastuste pikkuste kohta koostatud esmane statistiline ülevaade võimaldab öelda, et leitud külastuspikkused kajastavad võrdlemisi hästi tegelikkust. Viimasele annab kinnitust ka manuaalselt kõikide respondentide tegevuskohtade inspekteerimine. Samas ilmnes läbiviidud annoteerimisprotsessist vajadus edaspidi võimaldada tegevuskoha määratlemist mitmesse kategooriasse. Sel moel on paremini võimalik kindlaks teha tegevuskoha funktsionaalsus.

Külastuse nädalapäevaline kuuluvus esindab valikuid inimese ajakasutuses ning võib seetõttu tegevuskohtade lõikes varieeruda – võrdlemisi piiratud vaba ajaga argipäevadel võib tegevustes osalemine olla vähemtõenäoline kui nädalavahetustel, mistõttu võib tegevuskohtade külastamises esineda teatud rütmilisus või regulaarsus (Buliung et al. 2008). Võrdlemaks erinevaid tegevuskohti omavahel, vaadeldi antud

töös kategooriate lõikes tegevuste osakaalu nädalavahetusel. Tunnuse iseloomulikkus kerkis esile vaid töö kategoorias, mis on väga kindla nädalase rütmiga tegevuskoht. Kategooriate puhul nagu kodu ja restoranid – baarid, kerkis esile erinevus enamikest teistest kategooriatest – vaid 1 kuni 2 tegevuskoha kategooriaga võrdlemisel erinevusi ei tuvastatud.

Rütmilisus võib tuleneda ka inimese igapäevastest valikutest, mis tingib erinevate tegevuste jaotumise päevas võrdlemisi kindlatele aegadele (Eagle & Pentland 2009). Selliste rutiinide sarnasus erinevate inimeste igapäevases elus kajastus ka antud uurimuse tulemustes argipäevadel töö ning restoranide-baaride kategoorias. Kuivõrd vaba aja, sportimise ning sotsiaalsete kohustustega seotud tegevused on ajas kõige paindlikuma graafikuga (Arentze & Timmermans 2004) ei tuvastanud kategooriatevaheline võrdlemine külastuse kellaaja kui tunnuse iseloomulikkust sellistes tegevuskohtades. Restoranide – baaride külastusi on samuti vaadeldud kui vabama graafikuga tegevusi (Arentze & Timmermans 2004), kuid antud töös selliste tegevuskohtade erinemine teistest tegevuskohtadest on tõenäoliselt seotud väga sarnase valimi sarnaste igapäevaste valikutega. Nädalavahetus kui suurema ajaplaneerimisvabadusega aeg nädalas tegevuskoha kategooriate-vahelisi erinevusi külastuse algusaegades vaadelda ei võimaldanud.

Külastuse kestus on, võrreldes külastuse algusaja ning nädalapäeva tüübiga, antud uuringus osalejate puhul paremini tegevuskohti omavahel eristav tunnus, võimaldades omavahel võrdlemisi täpselt eristada 70% vaadeldud kategooriatest. Teisalt on igal kategoorial vähemalt üks kategooria, millega külastuse pikkused on sarnase väärtusvahemikuga. Nädalavahetuste puhul oli tunnuse statistiliselt oluline eristusvõime mõnevõrra väiksem. Antud tunnuse puhul võib taas märgata seda, et suur hulk erinevusi kategooriate vahel eksisteerib, kuid iseloomulikkust, mis eristaks üht kategooriat kõikidest teistest ei esine.

Mobiilikasutust iseloomustavate näitajate kasutamine tegevuskohale iseloomulike külastustunnustena omab antud respondentide puhul mõningast potentsiaali tegevuskohtade puhul nagu transpordiga seotud kohad ning restoranid, kus mobiiltelefonikasutus analüüsitud respondentide puhul oli mõnevõrra suurem kui mujal. Suurima erinevuste arvuga kategooriana kerkib esile transpordikategooria, mida on põhjalikumalt IKT kasutamise kontekstis vaadeldud ka teistes uuringutes (Tillema et al. 2009). Transpordi ja restorani kategooria esilekerkimine on kooskõlas ka Google poolt läbi viidud mahuka uuringuga nutitelefonide kasutajate seas (The Mobile Movement 2011), kus selgus, et 87% uuritavatest on kasutanud transpordiga seonduvates kohtades nutitelefoni. Restoranide puhul on sama näitaja 73%. Samas selgus uuringust, et kõige sagedamini leiab nutitelefoni kasutust kodus (93%) ning kolmandal kohal on poodide kategooria (77%). Need on aga kaks tegevuskohta, mis kerkisid YouSense andmetest esile pigem väiksemate kasutusnäitajatega. Kuivõrd poodide puhul oli märgata ka suurt hulka tegevuskohti, mis olid tuvastatavad vaid “aukude” meetodil, võib eeldada, et paljudel juhtudel oli poekülastuse ajaks telefon

jäänud näiteks autosse. Seda kinnitab ka tegevuskohtade vaatlemine ortofotodel – paljudel juhtudel oli poodi tähistav klaster moodustunud poe kõrval asuvasse parklasse.

Teisalt ei kerkinud mobiilikasutustunnuste puhul siiski esile vaadeldavat iseloomulikkust, kus mõni kategooria eristus enamikust teistest kategooriatest ja oleks seetõttu tuvastatav pelgalt mobiilikasutustunnuste alusel. Põhjuseid märkimisväärsete erinevuste puudumisele võib leida mitmeid. Esiteks võivad käsitletud mobiilikasutustunnused olla liiga üldised, võimaldamaks käsitleda IKT kasutamisega seotud eripärasid kõikide tegevuskohtade lõikes. Teiseks võib mobiilikasutus ise olla inimese igapäevaelu loomulik osa, mis ei allu erinevatele sotsiaalsetele ja kontekstipõhiste piirangutele, vaid kujuneb läbi jooksvate, tegevuskohast sõltumatute vajaduste. Siinkohal võib oletada, et mobiilikasutustunnuste eristusvõimet aitaks parandada nende suurem spetsiifilisus – kogudes informatsiooni ka väljuvate kõnede ja lühisõnumite kohta või vaadeldes internetikasutust. Sellise teabe lisamine aitaks paremini mõista ka seda, mida inimesed teevad siis kui “ekraan on lukust lahti”.

Nutitelefoni andmestikus peituvate külastustunnuste potentsiaal ei seisne nende tunnuste eraldi vaatlemises, vaid koos käsitlemises. See tähendab seda, et tegevuskoha kategooria unikaalsus või iseloomulikkus peitub temale ainulaadses külastustunnuste komplektis, mis sisaldab endas nii käitumulikke, ruumikasutuslikke kui ajakasutuslikke aspekte. Käesolevas töös vaadeldi lisaks mobiilikasutusele peamisi külastustunnuseid, kuid võimalikke vaadeldavaid tunnuseid on reaalsuses oluliselt rohkem. Uuringu raames suudeti küll tuvastada erinevuste esinemine teatud kategooriate lõikes, kuid tunnuste puhul ei esinenud enamasti iseloomulikkust, mis tegevuskohti kõikidest teistest oleks eristanud.

Antud tulemustest selgub, et nii mõnegi kategooria puhul võib tunnuste omavaheline kombineerimine tuua välja tegevuskoha iseloomulikkuse (tabel 15). Näitena võib vaadelda poe kategooriat, kus kõige enam erinevusi välja toonud tunnus (külastuse kestus) ei suutnud eristada kategooriat transpordikategooriast. See tähendab seda, et klassifikatsioonimudel, mis põhineb külastuse kestuse tunnusel, võib klassifitseerida poed transpordiga seotud kohtadeks ja vastupidi. Kui aga sellisele mudelile lisada tunnuseks ekraani sisselülitamise sagedus, teeb mudel selliseid vigu tõenäoliselt vähem, kuna neis kahes kohas on tunnuse väärtuste erinevus statistiliselt oluline. Nutitelefoni kasutamist puudutavate andmete puhul võivad telefonikasutuslikud tunnused panustada mudeli täpsusastme paranemisse kuni mõne protsendipunkti võtta (Do & Gatica-Perez 2013). Seda, kui palju tunnused reaalselt mudeli “headusesse” panustavad, saab hinnata vaid läbi reaalselt koostatud mudelite.

Tabel 15. Külastustunnuste kombinatsioonid ja nende käigus välja kujunev tegevuskoha iseloom. Tunnus 1 tähistab külastustunnust, mis toob vaadeldava kategooria puhul välja kõige enam erinevusi teiste kategooriatega; valik 1 tähistab kategooriaid, mis tunnus 1 põhjal ei eristunud; tunnus 2 on lõplikuks eristamiseks vajaminev külastustunnus.

Kategooria	Tunnus 1	Valik 1	Tunnus 2
Kodu	külastuse kestus	kodu/teine kodu	nädalapäev
Töö	nädalapäev/algusaeg		
Teine kodu	külastuse kestus	kodu/teine kodu	algusaeg
Sõbra/sugulase kodu	külastuse kestus	sõbra-sugulase kodu/puhkus/töö	algusaeg
Restoran	külastuse kestus	restoran/transport	algusaeg
Pood	külastuse kestus	pood/transport	mobiil
Vaba aeg	-		
Sport	-		
Transport	külastuse kestus	pood/restoran/sport/transport	mobiil
Puhkus	-		

Käesoleva töö olulised esmased järeldused on väljatoodavad kolme peamise punktina:

- 1) Külastuse kestus on antud pilootuuringu andmetele tuginedes suurima potentsiaaliga tegevuskohti omavahel eristav tunnus, mis võiks teiste tunnustega kombinatsioonis eristada suurema osa tegevuskohti.
- 2) Mobiilikasutustunnused omavad potentsiaali vaid teatud tegevuskohti eristavate tunnustena. Nende tunnuste eristusvõimet tegevuskoha kategooriate lõikes oleks võimalik parandada läbi suurema spetsiifilisuse.
- 3) Tegevuskohtade semantika tuvastamise edu peitub võimalikult mitmekesiste käitumisandmete kaasamises klassifikatsioonimudelisse – ühte head tunnust, mis suudaks tegevuskohta eristada teistest, ei ole.

Kokkuvõte

Info- ja kommunikatsioonitehnoloogia laiaulatuslik levik ühiskonnas on avardanud võimalusi inimese ruumilise käitumise uurimiseks. Üha enam pöörduakse automatiseeritud andmekogumissüsteemide poole, mis võimaldavad uurida pikemaajalisi perioode võimalikult väikese koormusega respondendile. Nutitelefonid on näide tehnoloogiast, mida inimesed igapäevaselt kasutavad ja endaga kaasas kannavad, võimaldades ühtlasi automaatselt koguda mitmekesist informatsiooni inimese liikumise ja käitumise kohta. Samas vähendavad sellised süsteemid uurijale koheselt kättesaadava kasuliku informatsiooni hulka, mistõttu on järjest kasvav vajadus meetodika järele, mis võimaldaks efektiivselt kasulikku informatsiooni kogutud andmetest “kaevandada”.

Käesolevas töös arendati välja meetodilised võtted YouSense rakenduse poolt kogutavate inimese asukoha- ja käitumisandmete töötlemiseks, eesmärgiga tuvastada andmetest tegevuskohad ja neile vastavad külastustunnused. Väljatöötatud meetodikat rakendati pilootuuringu raames 18 inimese 6 kuu tegevuskohtade ja külastustunnuste leidmiseks, võimaldamaks hinnata külastustunnuste iseloomulikkust tegevuskohtade lõikes. Külastustunnustena käsitleti nädalapäeva tüüpi, külastuse algusaega ja kestust ning mobiilikasutuslikke tunnuseid.

Tegevuskohtade tuvastamiseks rakendati YouSense asukohaandmete puhul tiheduspõhist klasteranalüüsi DBSCAN kombinatsioonis ajaparameetrilise meetodiga. Rakendatud meetodite efektiivsus oli võrreldav olemasolevate GPS-põhiste uuringutega. Töö käigus tuvastati ka kasutatud meetodite peamised kitsaskohad ning tehti ettepanekud nende parandamiseks.

Tegevuskohtade võrdlemisel külastustunnuste lõikes selgus, et ajakasutuslikest tunnustest on külastuse kestus antud 18 inimese andmete põhjal kõige paremini tegevuskohti eristav tunnus, kuid pelgalt nimetatud tunnusest ei piisa, et tegevuskohti iseloomustada. Mobiilikasutuslikud tunnused omavad eristava tunnusena teatud potentsiaali vaid mõnes üksikus tegevuskohas, mistõttu on neid võimalik käsitleda kui teisejärgulisi iseloomustajaid. Käsitletud külastustunnuste efektiivsus peitub nende kooskäsitlemises, võimaldades sel moel iseloomustada kahte kolmandikku vaadeldud kategooriatest.

Summary

Master's thesis "Finding human activity places from smartphone gathered behavioral data"

The spread of information and communication technologies in society has given new opportunities for studying human spatial behaviour. Automated systems are gaining more popularity due to the ability to gather longitudinal data without extra respondent burden. Smartphones are example of such technologies that are being used in everyday lives, allowing to gather accurate movement and behavioral data. On the other hand, automately gathered data contains less readily available information for the researcher to use, thus there is increasing need for the methods that make the contained information usable.

The aim of this master thesis was to develop methodological steps for extracting significant places and relevant visitation variables from Android-based YouSense smartphone data. The methods were then used in a pilot study to extract places and visitation variables from 6 months data of 18 respondents in order to evaluate variation of the selected variables among main activity place categories. Selected variables were the type of day, start time and duration of visit and four different variables for smartphone usage.

Density-based clustering DBSCAN was used in combination with method involving time parameter for extracting significant places. The effectiveness of two combined methods were comparable to the ones with previous studies. The work also gives further instructions how to improve the efficiency of the methods for YouSense data.

Multiple comparisons between selected activity locations revealed that visit duration has the best differentiation power among selected timeuse variables. However visit duration itself is not enough to reveal the nature of activity locations. Variables for the smartphone usage gave relatively poor results, allowing to differentiate only some of the activity places, giving them potential as secondary variables. The most effective way of describing activity places in observed data would be using variables in combination.

Tänuavaldused

Täna oma juhendajat professor Rein Ahast, kelle juhiste ja näpunäidete abil antud töö valmis. Eraldi tahaksin tänada Erki Saluveeri Positiumist, YouSense rakenduse loojat Mattias Linnapit, Artjom Lindu ja dr. Amnir Hadachit STACC-st, kes olid nõu ja kommentaaridega abiks ajal kui seda väga vaja oli. Suurt tänu tahaksin avaldada ka oma sõpradele ja perekonnale kannatlikkuse, toe ja vajalike näpunäidete eest.

Kasutatud kirjandus

Alexander, B., Ettema, D., Dijst, M., 2011. Fragmentation of work activity as a multi-dimensional construct and its association with ICT, employment and sociodemographic characteristics. *Journal of Transport Geography* 18, 55–64.

Ankerst, M., Breuning, M. M., Kriegel, H.-P., Sander, J., 1999. OPTICS: Ordering points to identify the clustering structure. In *Proc. SIGMOD '99*, 49-60.

Arentze, T. A., Timmermans, H. J. P., 2004. A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological* 38(7), 613-633.

Bhat, C. R., Koppelman, F. S., 1993. A conceptual framework of individual activity program generation. *Transportation Research A* 27(6), 433–446.

Bhat, C. R., Koppelman, F. S., 1999. A retrospective and prospective survey of time-use research. *Transportation* 26, 119–139.

Bhat, C. R., Misra, R., 1999. Discretionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends. *Transportation* 26, 193–209.

Birkin, M., Malleson, N., 2012. Investigating the behaviour of Twitter users to construct an individual-level model of metropolitan dynamics. *NCRM Working Paper*.

Bolbol, A., Cheng, T., Tsapakis, I., Haworth, J., 2012. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Computers, Environment and Urban Systems* 36, 526-537.

Buliung, R. N., Roorda, M. J., Rimmel, T. K., 2008. Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto Travel-Activity Panel Survey (TTAPS). *Transportation* 35, 697 – 722.

Campbell, J., 1970. A stochastic model of human movement. *Discussion Paper*, 13, Department of Geography, The University of Iowa, Iowa City.

Couclelis H., 2000. From sustainable transportation to sustainable accessibility: can we avoid a new tragedy of the commons? In *Information, Place, and Cyberspace: Issues in Accessibility*, 342 – 356. Springer: Berlin.

Couclelis, H., 2003. Housing and the new geography of accessibility in the information age. *ICT, Housing & the Built Environment* 28(4), 7-13.

Couclelis, H., 2009. Rethinking time geography in the information age. *Environment and Planning A* 51, 1556 – 1575.

- Dijst, M., 1999. Two-earner families and their action spaces: a case study of two Dutch communities. *GeoJournal*, 48(3), 195–206.
- Do, T. M. T., Gatica-Perez, D., 2013. The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Transactions on Mobile Computing* 13 (3), 638 – 648.
- Du, J., Aultman-Hall, L., 2007. Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: automatic trip end identification issues. *Transportation Research Part A: Policy and Practice* 41 (3) 220–232.
- Eagle, N., Pentland, A.S., 2009. Eigenbehaviors: identifying structure in routine. *Behavioural Ecology Sociobiology* 63, 1057–1066.
- Ertöz, L., Steinbach, M., Kumar, V., 2003. Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data. In *Proc. Second SIAM International Conference on Data Mining*.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. KDD-96*.
- Feng, T., Timmermans, H. J. P., 2013. Transportation mode recognition using GPS and accelerometer data. *Transportation Research Part C* 37, 118 – 130.
- Frusti, T., Bhat, C., Axhausen, K., 2002. Exploratory analysis of fixed commitments in individual activity – travel patterns. *Transportation Research Record: Journal of the Transportation Research Board* 1807, 101-108.
- Golledge, R.G., Stimson, R.J., 1997. *Spatial Behavior: A Geographic Perspective*. Guilford Press, New York.
- Golob, T.F., Meurs, H., 1986. Biases on response over time in a seven-day travel diary. *Transportation* 13, 163-181.
- Hägerstrand, T., 1952. The propagation of innovation waves. *Lund studies in geography, Series B*, 4.
- Hannes, E., Liu, F., Vanhulsel, M., Janssens, D., Bellemans, T., Vanhoof, K., Wets, G., 2012. Tracking household routines using scheduling hypothesis embedded in skeletons. *Transportmetrica* 8(3), 225-241.
- Herrera, J. C., Work, D. B., Herring, R., Ban, X., Jacobson, Q., Bayen, A. M., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment. *Transportation Research Part C* 18, 568–583.
- Higuera de Frutos, Castro, S. M., 2014. Using smartphones as a very low-cost tool for road inventories. *Transportation Research Part C* 38, 136–145.
- Horton, F.E., Reynolds, D.R., 1971. Effects of urban spatial structure on individual behavior. *Econ. Geogr.* 47, 36–48.

- Huang, X., Luo, J., Wag, X., 2013. Finding frequent sub-trajectories with time constraints. In *Proc. UrbComp'13*.
- Hwang, S., Hanke, T., Evans, C., 2013. Automated extraction of community mobility measures from GPS stream data using temporal DBSCAN. *Lecture Notes in Computer Science* 7972, 86 – 98.
- Järv, O., Ahas, R., Witlox, F., 2014. Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C* 38, 122–135.
- Jones, P., Clarke, M., 1988. The significance and measurement of variability in travel behaviour. *Transportation* 15, 65–87.
- Kamruzzaman, M., Hine, J., 2012. Analysis of rural activity spaces and transport disadvantage using a multi-method approach. *Transport Policy* 19, 105–120.
- Kang, H., Scott, D. M., 2010. Exploring day-to-day variability in time use for household members. *Transportation Research Part A* 44, 609–619.
- Kenyon, S., Lyons, G., 2007. Introducing multitasking to the study of travel and ICT: Examining its extent and assessing its potential importance. *Transportation Research Part A* 41, 161–175.
- Konicek, A. R., Lefman, J., Szakal, C., 2012. Automated correlation and classification of secondary ion mass spectrometry images using a k-means cluster method. *Analyst* 137, 3479-3487.
- Koppelman, F. S., Pas, E. I., 1984. Estimation of disaggregate regression models of person trip generation with multiday data. In Volmuller J., Hamerslag R. (eds.) *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*, pp. 513–529. VNU Science Press, Utrecht, The Netherlands.
- Kurani, K.S., Lee-Gosselin, M. E. H., 1996. Synthesis of past activity analysis applications. Presented at the Travel Model Improvement Program (TMIP) Conference on activity-based travel forecasting, New Orleans, June 2-5.
- Kwan, M. P., 2007. Mobile communications, social networks, and urban travel: Hypertext as a new metaphor for conceptualizing spatial interaction. *The Professional Geographer* 59, 434–446.
- Lenz, B., Nobis, C., 2007. The changing allocation of activities in space and time by the use of ICT – “Fragmentation” as a new concept and empirical results. *Transportation Research Part A* 41, 190–204.
- Line, T., Jain, J., Lyons, G., 2011. The role of ICTs in everyday mobile lives. *Journal of Transport Geography* 19(6), 1490-1499.
- Linnap, M., Rice, A., forthcoming. Managed Participatory Sensing with YouSense. *Journal of Urban Technology* xx, xxx-xxx.

- Liu, F., Janssens, D., Wets, G., Cools, M., 2013. Annotating mobile phone location data with activity purposes using machine learning algorithms. *Expert Systems with Applications* 40, 3299-3311.
- Loseto, G., Ruta, M., Scioscia, F., Di Sciascio, E., Mongiello, M., 2012. Mining the user profile from a smartphone: a multimodal agent framework. In *Proc. CEUR Workshop*, 47-53.
- Ma, X., Wu, Y.-J., Wang, Y., Chen, F., Liu, J., 2013. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C* 36, 1–12.
- Marmasse, N., Schmandt, C., 2002. A user-centered location model. *Personal and Ubiquitous Computing*, 6(5-6), 318–321.
- Meloni, I., Guala, L., Loddo, A., 2004. Time allocation to discretionary in-home, out-of-home activities and to trips. *Transportation* 31, 69–96.
- Milenova, B. L., Campos, M. M., 2002. O-cluster: scalable clustering of large high dimensional data sets. In *Proc. ICDM '02*, 290 – 298.
- Neutens, T., Schwanen, T., Witlox, F., 2010. The prism of everyday life: Towards new research agenda for time geography. *Transport Reviews: A Transnational Transdisciplinary Journal* 31(1), 25-47.
- Nitsche, P., Widham, P., Breuss, S., Brändle, N., Maurer, P., 2013. Supporting large-scale travel surveys with smartphones – A practical approach. *Transportation Research Part C* xxx, xxx–xxx.
- Noulas, A., Scellato, S., Mascolo, C., Pontil, M., 2011. An empirical study of geographic user activity patterns in foursquare. *ICWSM'11*.
- Nurmi, P., 2009. Identifying meaningful places: PhD thesis. University of Helsinki, Helsinki.
- Ogle, J., Guensler, R., Bachman, W., Koutsak, M., Wolf, J., 2002. Accuracy of Global Positioning System for determining driver performance parameters. *Transportation Research Record* 1818, 12–24.
- Pas, E. I., 1987. Intrapersonal variability and model goodness-of-fit. *Transportation Research Part A* 21 (6), 431–438.
- Pearson, D., 2004. A comparison of trip determination methods in GPS-enhanced household travel surveys. In *84th Annual Meeting of the Transportation Research Board*, Washington, DC.
- Petrenko, A., Bell, S., Stanley, K., Qian, W., Sizo, A., Knowles, D., 2013. Human spatial behaviour, sensor informatics, and disaggregate data. *Lecture Notes in Computer Science* 8116, 224–242.
- Pinjari, A.R., Bhat, C.R., 2010. Activity-based travel demand analysis. *Handbook of*

Transport Economics, Chapter 10, pp. 213-248, eds. A. de Palma, R. Lindsey, E. Quinet, and R. Vickerman, Edward Elgar Publishing.

Raento, M., Oulasvirta, A., Eagle, N., 2009. Smartphone: an emerging tool for social scientists. *Sociological Methods & Research* 37(3), 426-454.

Raux, C., Ma, T.-Y., Joly, I., Kaufmann, V., Cornelis, E., Ovtracht, N., 2011. Travel and activity time allocation: an empirical comparison between eight cities in Europe. *Transport Policy* 18, 401–412.

Ren, F., Kwan, M.-P., 2007. Geovisualization of human hybrid activity-travel patterns, *Transactions in GIS* 11, 721–744.

Roorda, M.J., Ruiz, T., 2008. Long- and short-term dynamics in activity scheduling: a structural equations approach. *Transportation Research Part A: Policy and Practice* 42, 545–562.

Schlich, R., Axhausen, K.W., 2003. Habitual travel behaviour: evidence from a six-week travel diary. *Transportation* 30, 13–36.

Schönfelder, S., 2006. Urban rhythms – modelling the rhythms of individual travel behaviour: PhD thesis. ETH Zürich, Zürich.

Schönfelder, S., Axhausen, K.W., 2010. Urban rhythms and travel behaviour: spatial and temporal phenomena of daily travel. Ashgate Publication Company, England.

Schönfelder, S., Li, H., Guensler, R., Ogle, J., Axhausen, K. W., 2006. Analysis of commute atlanta instrumented vehicle GPS data: destination choice behavior and activity spaces. *Arbeitsberichte Verkehrs- und Raumplanung* 350, ETH Zürich.

Schönfelder, S., Li, H., Guensler, R., Ogle, J., Axhausen, K. W., 2006. Analysis of commute Atlanta instrumented vehicle GPS data: destination choice behavior and activity spaces. Paper presented at *the 85th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2006.

Schwanen, T., Kwan, M.P., 2008. The internet, mobile phone, and space-time constraints. *Geoforum* 39, 1362–1377.

Spissu, E., Pinjari, A. R., Bhat, C. R., Pendyala, R. M., Axhausen, K. W., 2009. An analysis of weekly out-of-home discretionary activity participation and time-use behavior. *Transportation* 36, 483 – 510.

Stopher, P., Fitzgerald, C., Britain, P., Zhang, J., 2006. Variability in day-to-day travel: analysis of a 28-day GPS survey. Australasian Transport Research Forum (ATRF) 29, Queensland.

Stopher, P.R., Greaves, S.P., 2007. Household travel surveys: Where are we going? *Transportation Research Part A* 41, 367–381.

Sun, Y., Fan, H., Bakillah, M., Zipf, A., 2013. Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems* xxx, xxx–xxx.

- Susilo, Y. O., Axhausen, K. W., 2007. Stability in individual daily activity-travel-location patterns: a study using the Herfindahl-Hirschman Index. *Arbeitsberichte Verkehrs- und Raumplanung* 435, ETH Zürich.
- Tarigan, A. K. M., Fujii, S., Kitamura, R., 2012. Intrapersonal variability in leisure activity- travel patterns: the case of one-worker and two-worker households. *The International Journal of Transportation Research* 4, 1-13.
- The Mobile Movement, 2011. Understanding smartphone users. Google/Ipsos OTX MediaCT.
- Thierry, B., Chaix, B., Kestens, Y., 2013. Detecting activity locations from raw GPS data – a novel kernel based algorithm. *International Journal of Health Geographics* 2013, 12 – 14.
- Tillema, T., Schwanen, T., Dijst, M., 2009. Communicating something confidential while travelling by train: The use of a telephone conversation versus silent modes. *Transportation* 36, 541-564.
- Valentine, G., Holloway, S.L., 2002. Cyberkids? Exploring children's identities and social networks in on-line and off-line worlds. *Annals of the Association of American Geographers* 92 (2), 302–319.
- Wan, N., Lin, G., 2013. Life-space characterization from cellular telephone collected GPS data. *Computers, Environment and Urban Systems* 39, 63 – 70.
- Wilson, J., 2004. Measuring personal travel and goods movement (Transportation Research Board Special Report 277). Transportation Research Board, Washington, DC.
- Wolf, J., Guensler, R., Bachman, W., 2001. Elimination of the travel diary: An experiment to derive trip purpose from GPS travel data. In *80th Annual Meeting of the Transportation Research Board*, Washington, DC.
- Wolf, J., Oliveira, M., Thompson, M., 2003. Impact of underreporting on mileage and travel time estimates – results from Global Positioning System-enhanced household travel survey. *Transportation Research Record* 1854, 189–198.
- Yamamoto, T., Kitamura, R., 1999. An analysis of time allocation to in-home and out-of-home discretionary activities across working days and no-working days. *Transportation* 26, 211–230.
- Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L., 2004. Discovering personal gazetteers: an interactive clustering approach. In *Proc. ACMGIS*, 266 – 273.
- Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L., 2007. Discovering personally meaningful places: an interactive clustering approach. *ACM Transactions on Information Systems* 25 (3).

Lisad

Lisa 1. Respondentidele edastatud juhend annoteerimiseks

JUHEND

Kirjaga on kaasas **tegevuskohad.xls** fail ning **link** veebirakendusele, kus asuvad Teile tegevuskohad ajavahemikul 01.07.2013–31.12.2013.

1. Avage kaart ning kõrvale tegevuskohad.xls fail.
2. Kaardil on näha tegevuskohtade markerid. Neil klikkides avaneb aken, kus on järgnev info kohatuvastuseks:
 - a. *visit_count* – märgib, mitu korda antud kohta külastatud on;
 - b. *latest* – viimase külastuse ligikaudne aeg formaadis "aasta-kuu-päev kellaaeg + ajavöönd";
 - c. tegevuskoha ID ülemises vasakus servas;
3. Valige failis asuvast kategooriate loetelust tegevuskoha jaoks sobiv kategooria ja märkige see tegevuskoht_ID veerus asuva sobiva ID taha, arvestades järgnevat:
 - a. Kui marker asub mõne Teile teadaoleva tegevuskoha naabruses ja selles naabruskonnas pole tegevuskohale paremini sobivat markerit, siis märkige .xls faili antud tegevuskoha kategooria.
 - b. Kui markeri naabruses asub mitu Teile teadaolevat tegevuskohta, kuid markereid on vaid üks, siis nimetage markeri kategooriaks tegevuskohta kõige enam iseloomustav kategooria. Näiteks Tasku keskuses olete külastanud paar korda kino ja üsna sagedasti toidupoodi, siis valige kategooriaks "Pood/ostukeskus".
 - c. Kui marker ei asu mitte ühegi Teile teadaoleva tegevuskoha naabruses, siis nimetage kategooriaks "Ei tea".
4. Kustutage marker kaardilt, kasutades nuppu "Kustuta funktsioon".
5. Korrake punkte 2-4 seni kuni kõik lahtrid .xls failis on täidetud ning kaardil pole enam markereid.
6. Salvestage tegevuskohad.xls fail ning saatke aadressil vkaisa@ut.ee.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kaisa Vent

(sünnikuupäev: 01.02.1987),

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Inimese tegevuskohtade leidmine nutitelefonipõhiste käitumisandmestike alusel,

mille juhendaja on Rein Ahas,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace'i lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.3. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile;

1.4. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartu, 19.05.2014