

Time-series Analysis

Tausta

Aikasarja-aineistoa voidaan hyödyntää IT-alan tutkimuksessa esimerkiksi jonkin ohjelmiston tai palvelun käyttöä eri ajankohtina. Tämän tiedon perusteella voidaan tutkia mitä mahdollisia ongelmia käyttöpiikit ovat aiheuttaneet ja miten niistä voitaisiin selvittää. Normaalista poikkeava käyttö voidaan myös tunnistaa mahdolliseksi hyökkäykseksi. Toinen kohde voisi olla jatkuvasti pyöriväksi tarkoitettu ohjelmisto. Vaikuttaako kulunut aika jotenkin sen toimivuuteen.

Aikasarja-aineistoa voidaan käyttää sekä deduktiivisesti että induktiivisesti. Esimerkiksi aikuisviidesivuston käyttäjämäärässä nähdään toistuvasti suuri piikki iltaisin 23:00. Induktiivisesti voidaan olettaa tämän myös tulevaisuudessa. Pidemmällä aikavälillä saatettaisiin tiettyjen lomapyhien lisäävän piikkiä entisestään, jolloin näitä varten voitaisiin lisätä palvelinresursseja.

Toisaalta edellisessä esimerkissä on myös deduktiivinen oletus. Kaikkien verkkosivujen käyttäjämäärissä on piikki lomapyhien aikaan (premissi 1), aikuisviihdesivusto on verkkosivu (premissi 2), joten sivuston käyttäjäpiikki on "normaali" arkea korkeampi lomapyhien aikaan (johtopäätös).

Filosofisilta oletuksista aikasarja-aineisto sopii selviten postpositivismiin, sillä materiaali on hyvin tilastopohjaista ja neutraalisti kerättyä. Tilastot myös käyttäytyvät tietyllä tapaa. Metodologisella puolella on kuitenkin kaikuja sosiaaliseen konstruktivismiin ja pragmatismiin, sillä aineiston keräämisessä ja käsittelyssä voi olla hyvin induktiivinen lähestymistapa.

Ilman tarkempia kuvauksia Järvisen taksonomioista, aikasarja-aineiston voisi kuvitella kuuluvat sarjaan "conceptual-analytical approaches".

Esimerkkiartikkeli:

Xu F., Lin Y., Huang J., Wu D., Shi H., Song J., Li Y., (2016)

Big Data Driven Mobile Traffic Understanding and Forecasting: A Time Series Approach,

<http://dx.doi.org/10.1109/TSC.2016.2599878>

Datan kerääminen

Aineistotyyppiltään aikasarja-aineisto on IT-alalla hyvin numeerista.

Aineiston keräämiseen sopivimmat metodit ovat haastattelut, havainnointit ja kokeilut. Haastattelut toimivat parhaiten silloin, kun halutaan kerätä myös subjektiivisia kokemuksia tai automaattisten mittauksien kerääminen en jostain syystä onnistu. Esimerkiksi nettipalveluissa käyttöä on helpointa tutkia yksinkertaisella havainnoinnilla. Mitä, milloin ja missä tilanteissa käyttäjä tekee mitään. Ohjelmiston toiminnan tehokkuutta ajan kuluessa voi olla luontevinta tutkia kokeiluilla. Miten muutokset vaikuttavat toimintaan?

Haastatteluissa tarvitaan raakaa manuaalista voimaa. Muissa tapauksissa tärkeimmäksi tekijäksi nousee hyvin suunniteltu automaattinen aineiston kerääminen. Varmistaa, että kerätty aineisto saadaan pidettyä järkevästi tallessa ja luettavassa muodossa.

Soveltaminen

Jatketaan aiempaa aikuisviihdesivustoesimerkkiä. Sivustolla on videoita ja kuvia, jotka ovat maksumuurin takana. Rekisteröimätön käyttäjä voi katsoa lyhyitä teaser-pätkiä. Sivuston omistaja on kiinnostunut mihin aikaan sivuston käyttöpiikit ovat ja kuinka erityisen kiinnostavat uudet julkaisut vaikuttavat niihin. Käyttäjryhmää ei tässä tapauksessa voida erityisesti rajata tai valita, vaan tutkimus kohdistuu kaikkiin sivun käyttäjiin. Mikäli käyttäjä on rekisteröitynyt, voidaan hänen antamansa tiedot kerätä anonymisti aineiston tarkentamiseksi.

Tutkimuksesta havainnoimalla saatava tieto on hyvin käytännönläheistä. Saatua dataa voidaan suoraan hyödyntää oikean palvelinkapasiteetin mitoittamiseen kulloisenakin hetkenä. Toisaalta omistaja tietää minkälaista sisältöä kannattaa priorisoida.

Havainnointi on selvästi tehokkain keino aineiston keräämiseen, sillä hyvin suunnitellulla sivustolla jokaisesta klikkauksesta jää datajälki. Etenkin rekisteröityneiden käyttäjien toimintaa on helppo seurata. Heidän kohdallaan myös mahdollisesti muuttuva IP-osoite voidaan huomioida, jolloin tieto sivustolla käyvistä uniikeista käyttäjistä saadaan todenmukaisemmaksi.

Koska kaiken materiaallinen näkeminen vaatii kirjautumista, voidaan tutkimuksen validiteetin olevan korkea. Teaser-pätkissä validiteetti on vastaavasti hieman matalampi, sillä käyttäjien yksilöinnistä ei voida olla yhtä varmoja. Esimerkiksi botit voivat tuoda sivustolatauksia.

Sivuston "perusarjen" tutkimisessa reliabiliteetti on hyvä. Voidaan katsoa, että ihmisten tottumukset pysyvät varsin samanlaisina. Sen sijaan sivuston omistajan kiinnostus erityispiikkeihin on hieman haastavampi asia. Vaikka tietyt videot aiheuttaisivatkin erityisen suurta trafikkia hetkellisesti, ei ole varmaa, että tätä tapahtuisi tulevaisuudessakin. Jos sama tähti aiheuttaa kolme suurta piikkiä, ei neljättä välttämättä ole enää luvassa, sillä käyttäjät saattavat odottaa jo sitä seuraavaa uutta juttua.

Aikuisviihteen kulutus on monelle hyvin henkilökohtainen asia, joten aineistoa kerätessä on pidettävä huolta, ettei sitä voida yhdistää alkuperäiseen henkilöön. Toisaalta, jos käyttäjiä kategorisoidaan jollain tapaa, pitää myös kategorioista tehdä tarpeeksi laajoja, jottei käyttäjien voitaisi niiden kautta selkeästi olettaa kuuluvan tiettyyn ryhmään.