

# Päätöspuut ja markkinointisovellukset

Jyri Partanen  
28.3.2001

## Tiivistelmä

Tarkastelen päätöspuusovelluksia markkinoinnin suuntaamisen apuvälineenä. Lisäksi esitän menetelmien toteutuksia ja niiden yhteyden perinteisiin markkinointistrategioihin ja taloustieteellisiin teorioihin.

## 1 Johdanto

Yleistä. Mainontastrategian valinta ja toteutus vaikuttaa merkittävästi tuotteiden markkinaosuuden jakautumiseen. Tuottajan mielenkiinto kohdistuu kulujen minimointiin, minkä takia mainonnan osuus kuluissa on monesti tarkkailun alaisena.

Markkinointi. Kansantaloustieteen osana tutkittava markkinointi ja erityisesti mainonta ovat saaneet viime vuosina uusia tapoja parantaa tehokkuuttaan tutkimuksessa. Näitä tapoja ovat esimerkiksi älykkäiden järjestelmien käyttö talouden mallinnuksessa [Koh20].

Algoritmit. Päätöspuu esittää sellaiset muuttujat ja näiden muuttujien arvoalueet, joiden avulla luokiteltavan muuttujan arvot voidaan parhaiten erottaa toisistaan. Päätössolmut kuvaavat, miten aineisto jakautuu luokkiin soveltamalla sääntöjä juurisolmun ja jokaisen päätössolmun välillä. Solmuissa tehtävät aineiston lajittelut muodostavat sääntöjä.

Sovellukset. Käytännön sovellusesimerkkeinä esitellään kaksi päätöspuiden opetusalgoritmia. Ensimmäinen on perusalgoritmi, jolla mallinnetaan todennäköisyyksiä joilla tiettyyn ryhmään kuuluvat kuluttajat ostavat tuotteen. Toinen algoritmi eroaa ensimmäisestä siten että se mallintaa kuluttajaryhmien ostotodennäköisyyksiä laajentuen sisältämään tilanteet, joissa kuluttajaryhmälle joko lähetetään tai ei lähetetä mainosta tuotteesta. Algoritmien toimivuutta käytännössä verrataan toisiinsa ja yksinkertaiseen strategiaan, jossa mainos lähetetään kaikille.

## 2 Mainonta ja kansantaloustiede

Kansantaloustieteen lähtökohta markkinoinnissa on niukkuuden ongelma; tarpeemme ovat rajattomat, mutta resurssimme rajalliset. Niukan resurssin eri käyttötarkoitukset joutuvat kilpailemaan keskenään [Lah98]. Kansantaloustieteen kohteena on ihmisten tekemät valinnat niukkuuden vallitessa. Valinnan hinta ilmoitetaan vaihtoehtoiskustannuksena, joka kertoo päätökselle parhaan mahdollisen muun päätöksen.

Mikrotaloustieteessä tarkastellaan yksittäisten valintoja ja näiden seurauksia. Makrotaloustiede tutkii kansantalouksia kokonaisuuksina ja maailmantaloutta ja näiden kehittymistä. Kansantaloustieteessä ilmiöitä tarkastellaan mallien eli todellisuuden pelkistettyjen kuvausten avulla. Mallien takana on yleensä neljä perusoletusta, jotka ovat:

- yksilöillä on preferenssejä;
- yksilöillä on alkuvaranto resursseja
- yksilöt tekevät rationaalisia valintoja ja
- valinnat ovat koordinoituja. [Lah98]

Näiden yksinkertaisten mallien hyödyntämisessä on käytetty älykkäitä järjestelmiä selvittämään kansantalouden vaikuttimia mikro- ja makrotasolla [Koh20]. Mallit poikkeavat todellisista talouden tiloista karkeutensa vuoksi, koska kansantalouden todellisia vaikuttimia on yleisesti vaikea mallintaa tai edes havaita.

Kansantalous on mekanismi, joka jakaa niukat resurssit kilpailevien käyttötarkoitusten välillä. Kansantalouden päätöksentekijöitä ovat kotitaloudet, yritykset ja valtiovalta. Päätöksentekijät kohtaavat markkinoilla. Hyödykemarkkinoilla kotitaloudet ostavat tavaroita ja palveluita yrityksiltä ja tuotannontekijämarkkinoilla kotitaloudet myyvät tuotannontekijöitä yritysten käyttöön. Valtiovalta säätelee markkinoiden toimintaa, kerää kotitalouksilta ja yrityksiltä veroja ja huolehtii tulonsiirroista ja julkisista palveluista.

Mainonnan synty edellyttää toimintakentäkseen kapitalistiselle taloudelle rakentuvaa yhteiskuntaa. Kapitalistisen talouden yksi ominaispiirteistä on käyttöarvojen tuottaminen tavaroiden kautta. Tavaroiden mainonta ja massiivinen tuotekehittäminen ovat seurausta tavaroiden tuottajien hyödyn tavoitteluun perustuvasta intressistä kehittää uusia käyttöarvoja ja niiden levittämisestä mahdollisimman laajalle.

Jos kapitalistinen talous tuottaa tavaroita, niin mainonnan tehtävänä on puolestaan lisätä niiden kysyntää. Voidaan sanoa, että kysynnän tuotantona myös mainonta on tavaratuotantoa (kaksinkertainen tavarantuotanto). Eli mainostavaran erityinen rakenne viittaa tiettyyn tavarahan, mutta samalla se paljastaa tietyn mainostajan kyvyn lisätä tavarahan myyntiä; lisätä kysyntää.

Kapitalistisen talouden tunnusmerkkinä esiintyvä mainonta paljastaa sen tiiviin yhteyden vallitsevaan kulttuuriin ja yhteiskuntamuotoon. Kullakin kulttuurilla on omat merkitysulottuvuudet, joiden rajoissa mainonnalla on mahdollisuus toimia.

Ollakseen tehokas on mainostajan työstettävä juuri kulttuurin ytimessä olevia merkityksiä, sillä mainos rinnastetaan muihin kulttuurintuotteisiin. Mainoksella on kaksi ulottuvuutta. Se vaikuttaa vastaanottajaan yksilötasolla, mutta se on samanaikaisesti vuorovaikutuksessa myös ympäröivään kulttuuriin.

Vallalla olevat markkinointifilosofiat voidaan jakaa neljään ryhmään painopistealueensa mukaisesti: tuotantosuuntainen, tuotesuuntainen, myyntisuuntainen sekä markkinointisuuntainen. Tuotantosuuntainen edustaa vanhinta filosofiaa ja sen mukaan kuluttajat pitävät tuotteista, joita on tarjolla ja joihin heillä on varaa; keskittyttävä tuotantoon ja jakelun tehokkuuteen. Tuotesuuntaisessa ajattelussa kuluttajat suosivat tuotteita, jotka tarjoavat eniten laatua, suorituskykyä tai ominaisuuksia; organisaation panostettava jatkuvaan tuotekehittelyyn. Myyntisuuntainen filosofia uskoo "korkeapainemyyntiin" eli panostetaan tavaroiden myyntiin, joita kuluttajat eivät muuten ostaisi, esim. vakuutukset ja tietosanakirjat. Markkinasuuntaisen ajattelutavan mukaan tavoitteiden saavuttaminen riippuu kohdemarkkinoiden tarpeiden ja toiveiden määrittelystä ja siitä, miten paljon kilpailijoita paremmin asiakkaan tarpeet voidaan tyydyttää.

Markkinointi pyrkii uusimman näkemyksen mukaan luomaan, kehittämään ja ylläpitämään kestäviä ja molempia osapuolia tyydyttäviä yhteistyösuhteita hyödykkeiden tuottajien ja käyttäjien välillä erityisesti kaupallisilla markkinoilla. Mainonnan tavoitteet voidaan tiivistää kolmeen pääkohtaan: kuluttamiseen suostuttelu, tiedon jakaminen kuluttajalle sekä tuotteen selviytyminen kilpailusta toisten mainosten kanssa.

Markkinointitutkimusten tavoitteena on tiedon kerääminen markkinoijalle. Tietoa kuluttajista, kulutustottumuksista sekä tuotteista markkinoija tarvitsee ollakseen kilpailukykyinen ja vastataksaan entistä paremmin kuluttajien tarpeisiin. Tavallisimpia markkinointitutkimuksia ovat esim. markkinapotentiaalin mittaaminen, myynnin analysointi, trendien tutkimus, kilpailijoiden tuotteiden tutkiminen sekä lyhyen- ja pitkäkätähtäimen ennustukset. Markkinointitutkimusten lisäksi markkinoijat ovat turvanneet selustansa tiedon hankkimiseksi oman informaatiojärjestelmän kautta. Kyseinen järjestelmä koostuu ihmisistä, laitteista ja menettelytavoista, joiden avulla kerätään, järjestetään, analysoidaan, arvioidaan ja jaetaan tarvittavaa ajankohtaista ja täsmällistä informaatiota markkinoinnin päätöksentekijöille. Usein markkinointitutkimukset eivät yksin pysty antamaan markkinoijalle riittävästi tietoa tai tarvittava tieto saadaan liian myöhään.

### 3 Mainoskampanjan suunnittelu

Mainoskampanjan ensimmäinen ja samalla hyvin ratkaiseva vaihe on kampanjan tavoitteen täsmentäminen. Yleisimmät tavoitteet ovat myynnin lisääminen, tuotteen tai yrityksen tunnetuksi tekeminen ja yrityskuvan parantaminen. On selvítettävä tarkasti, mitä tarvetta tuote tyydyttää, ennen kuin sitä voidaan tehokkaasti mainostaa. Mainonnan suunnittelussa on keskeistä tietää, mitä hyviä ja huonoja ominaisuuksia tuotteella on. Mainoksissa ei kuitenkaan tuoda esille kuin keskeisimmät hyödyt.

Kilpailutilanne on yleensä melko pysyvä, mutta siinä tapahtuu kuitenkin joskus muutoksia. Tämän vuoksi koko kilpailijajoukkoa on tarkasteltava toisinaan tavanomaista syvällisemmin ja pantava kilpailijat aika ajoin paremmuusjärjestykseen. On analysoitava erityisesti kilpailijoiden toimintoja sekä heidän vahvuuksiaan ja heikkouksiaan. Vaikka kilpailutilanteen tarkkailu on selvästi yksi tärkeimmistä tekijöistä uutuuksien omaavien tuotteiden markkinoille tuonnissa, eivät kirjoittajat Chickering ja Heckerman kuitenkaan ole sisällyttäneet ominaisuutta päätöspuidensa kasvatusalgoritmeihin.

Kohderyhmän valinta on perinteisten taloustieteidenkin näkökannalta katsottuna ratkaiseva tekijä myynnin onnistumisessa. Tämän kirjoittajat ovat sisäistäneet, ja heidän kirjoituksensa sisältääkin pääasiassa juuri tämän markkinoinnin vaiheen analysointia.

Kampanjaan käytetyn median ja sanoman valinta on perinteisesti ollut myös tärkeässä asemassa. Ohjelmistojen ja niiden sisällön myynissä näiden eroavuus vanhoista tekniikoista on kuitenkin muuttunut selvästi. Uusi sukupolvi on tottunut täysin uudenlaiseen mainontaan, eikä enää nykyään tulisi mainontaa suunnitellessaan väheksyä mainoksien arvoa viihdykkeenä kuten kirjoittajat tekevät [Chi20][Lah98].

Muita mainonnan ominaisuuksia joihin tämän kirjoitelman esittämää päätöspuurakenteita kannattaisi laajentaa, ovat pitkäaikaisessa mainonnassa tulosten tutkiminen mainonnan kuluessa ja mainosbudjetin maksimikoon sisällyttäminen saatavan hyödyn ja mainonnan laajuuden laskentaa.

## 4 Päätöspuut

Päätöspuu on kaavioesitys, jonka rakenne kuvaa aineiston jakautumista eri muuttujien mukaan. Puuta muodostettaessa määritellään, minkä muuttujan arvoja luokitellaan. Päätöspuu esittää sellaiset muuttujat ja näiden muuttujien arvoalueet, joiden avulla luokiteltavan muuttujan arvot voidaan parhaiten erottaa toisistaan. Päätöspuu muodostaa joukon päättelysääntöjä, joita voidaan käyttää uuden aineiston luokittelussa [Qui93].

Päätöspuu alkaa juurisolmusta (root node), joka sisältää kaikki havainnot käsiteltävästä aineistosta. Juurisolmun muuttujaa kutsutaan analyysin kohdemuuttujaksi (target). Puuta alaspäin tarkasteltaessa havainnot jakautuvat täydellisesti poissulkeviin joukkoihin (mutually exclusive subset). Valitusta puunkasvatusalgoritmista riippuen jako voidaan tehdä kahteen tai useampaan haaraan. Jokaiseen jakautumiskohtaan liittyy muuttuja, jonka perusteella jako on tehty sekä muuttujan arvoalue kullekin haaralle. Puun alimpia solmuja kutsutaan päätösolmuiksi (terminal node). Päätösolmut kuvaavat, miten aineisto jakautuu luokkiin soveltamalla sääntöjä juurisolmun ja jokaisen päätösolmun välillä. Jos puu on täydellinen, päätösolmut kuvaavat aineiston yksikäsitteisiin luokkiin.

Käytännössä puu ei yleensä ole täydellinen. Yksittäiset päätösolmut voivat kuvata aineiston useampaan eri luokkaan. Puun luokitteluvirhe voidaan kuvata esimerkiksi

prosentuaalisena riskiarviona (risk estimate), joka kertoo, mikä osa aineistosta luokitellaan väärään luokkaan.

Kaikkien menetelmien avulla saaduista puista on mahdollista laskea puun luokitteluvirhe. Luokitteluvirhe lasketaan joko jakamalla aineisto opetus- ja validointijoukkoon tai ristiinvalidoinnin (cross validation) avulla. [Bre84] Ristiinvalidoinnissa aineisto jaetaan haluttuun määrään osia, esimerkiksi kymmeneen. Kun validointi suoritetaan, aineistosta rakennetaan tässä tapauksessa 10 eri päätöspuuta siten, että jokainen kymmenys käytetään vuorollaan tarkistusjoukkona. Luokitteluvirhe lasketaan eri validoinneilla saatujen virheiden keskiarvona. Ristiinvalidointi antaa paremman kuvan puun luokitteluvirheestä kuin jako opetus- ja validointijoukkoon, mutta se on laskennallisesti vaativampi.

Päätöspuut, jotka kussakin solmussa jakavat data-avaruuden yhden muuttujan suhteen, tuottavat helposti tulkittavia malleja. Kirjoitelman puut ovat osittain juuri tämän muotoisia. Mallien puutteena on kuitenkin niiden rajoittunut kuvauskyky, koska niillä ei voida kuvata muita kuin attribuuttiakselien suuntaisia sääntöjä. Jos malliavaruutta laajennetaan sallimalla mutkikkaammat rajoittavat hyperpinnat, kasvaa mallien ennustuskyky vastaavasti mutta mallin tulkinnan helppous heikkenee.

Päätöspuu-mallien arviointi perustuu suurelta osin suurimman uskottavuuden käyttöön, missä laskettua uskottavuus arvoa rangaistaan mallin mutkikkuuden kasvaessa. Ahneessa (greedy) lähestymistavassa valmiista puusta karsitaan vähiten päätösfunktiota parantavat oksat tai alipuut ja siten rajoitetaan eksponentiaalista hakuavaruutta. Karsinnalla paitsi pienennetään mallia yritetään parantaa sen yleistyskykyä poistamalla siltä liika joustavuus.

## 5 Mainontasovellukset

Ongelmana Chickeringin ja Heckermanin artikkelissa esiteltiin Microsoft, jolla on paljon potentiaalisia asiakkaita. Tuotteena jota Microsoft yrittää tässä esimerkissä markkinoida on Microsoft Network (MSN) –tilaus. Potentiaalinen asiakaskunta koostuu Windows 95 :n rekisteröineistä käyttäjistä.

Markkinointistrategiaa suunniteltaessa mainostaja käy mahdollisesti läpi erilaisia tapoja lähestyä kuluttajaa. Tätä vaihetta kutsutaan markkinointitutkimukseksi. Markkinointitutkimusten tavoitteena on tiedon kerääminen ja asiakaspotentiaalnin kartoitus. Markkinoija tarvitsee tietoa kuluttajista, kulutustottumuksista sekä tuotteista ollakseen kilpailukykyinen ja vastatakseen entistä paremmin kuluttajien tarpeisiin. Markkinointitutkimus voidaan tehdä todellisen asiakaskunnan osajoukolla, jonka perusteella muodostetaan käsillä olevien markkinoiden malli. Tätä osajoukkoa esimerkissä vastaavat asiakaskunnan rekisteröityneet Windows 95 :n käyttäjät.

Mainonta voidaan kohdistaa suoraan kaikille asiakaskannassa oleville (mail-to-all). Tämä strategia voi olla kustannustehokas [Hug96], mutta muitakin vaihtoehtoja on mahdollista käyttää. Yksi näistä strategioista käyttää koneoppimistekniikoita asiakkaiden ryhmittelyyn. Oikean strategian löytämiseen vaikuttavat asiakkaiden ryhmittelytavan lisäksi eri markkinointitapoihin liittyvät kulut.

Chickeringin ja Heckermanin artikkelissa esitellyissä strategioissa käytetään erityisesti tapoja, joissa kustannusfunktiona on koko markkinaprosessin voitto. Hughesin menetelmässä vastaavasti haetaan suurinta mahdollista kiinnostuneiden asiakkaiden määrää.

## 6 Asiakasryhmien valinta

Määriteltyämme menetelmän, jolla asiakaskannan henkilöt jaetaan ryhmiin, on valittava kriteerit joiden avulla päätetään missä tilanteessa kutakin ryhmää lähestytään mainoksin. Kriteerinä esimerkissä käytetään tilauksen hintaa, joka on samalla esimerkin asiakkaiden ostopäätökseen vaikuttava muuttuja. Tämän vuoksi markkinoinnin tehokkuus voidaan laskea kokonaisuudessaan jäljempänä nähtävällä tavalla.

Testiasiakkaista koostuva testiasiakaskanta jaetaan ostopäätöksiensä mukaan neljään ryhmään. Asiakasjoukko always-buy ostaa tuotteen aina, oli häneen kohdistettu mainontaa tai ei. Joukko persuadable ostaa tuotteen jos häneen kohdistetaan mainontaa. Joukossa anti-persuadable asiakkaat vastaavasti eivät osta tuotetta, jos heille lähetetään mainos. Joukko never-buy ei osta tuotetta missään tilanteessa. Merkitään ryhmien sisältämien asiakkaiden määrää seuraavasti:  $N_{Alw}$ ,  $N_{Pers}$ ,  $N_{Anti}$  ja  $N_{Never}$ .

## 7 Mainonnasta kertyvä hyöty

Olkoon  $N$  kaikkien asiakkaiden määrä,  $c$  mainoksen lähettämisen hinta,  $r_u$  mainostamattoman myynnin voitto ja  $r_s$  mainostetun myynnin voitto. Odotettu voitto mainonnan tekemisestä kaikille ryhmille on

$$-c + \frac{(N_{Alw} + N_{Pers})}{N} \cdot r_s$$

Jos mainoksia ei lähetetä tulee voitto always-buy- ja anti-persuadable -ryhmien tekemistä ostoista. Voitto on tällöin

$$\frac{(N_{Alw} + N_{Anti})}{N} \cdot r_u$$

Valittaessa edellä mainittujen yksinkertaisten markkinointistrategioiden välillä on riippuvainen asiakasryhmien koosta ja mainonnan kuluista. Näinollen näistä strategioista valitaan se jonka kaavalla on suurempi arvo.

Olkoon  $M$  binäärimuuttuja, joka kuvaa onko asiakkaalle lähetetty mainos. Kun mainos on lähetetty  $M$  saa arvon  $m_1$  ja kun ei ole lähetetty  $M$  saa arvon  $m_0$ . Olkoon  $S$  binäärimuuttuja, joka kuvaa tilaako ( $s_1$ ) asiakas tuotteen vaiko ei ( $s_0$ ). Nyt mainonnasta saatava ELP (expected lift of profit) on

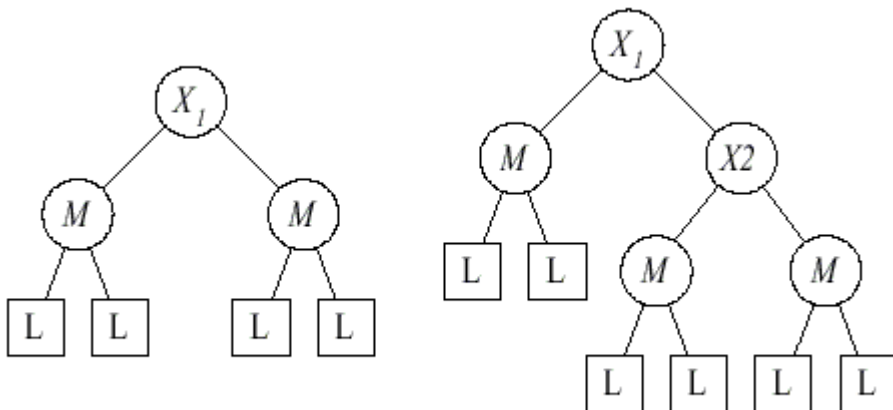
$$ELP = r_s \cdot p(S = s_1 | M = m_1) - r_u \cdot p(S = s_1 | M = m_0) - c$$

Malli, jota käytetään tämän kirjoitelman asiakasryhmien todennäköisyysjakauman kuvaamiseen on yksinkertainen päätöspuu. Muita mahdollisia mallinnustapoja olisivat voineet olla support vector machine (SVM)-tyyppiset asiakkaiden muuttujien vertailut, bayes-verkot tai yleiset lineaarimallit. Suuremmalla muuttujamäärällä SVM olisi saattanut antaa tarkempia asiakasryhmien määrittelyitä, mutta muuttujien määrän ollessa esimerkin kokoluokkaa, on lienee käteväntä käyttää päätöspuita.

Asiakasryhmille annettujen binaarimuuttujien lisäksi esitellään muuttujat  $X = \{X_1, \dots, X_n\}$ , jotka sisältävät muuta rekisteröintivaiheessa kerättyä tietoa asiakkaista ja erityisesti sen kuinka ELP eroaa asiakasryhmien osajoukkojen välillä. Näin voidaan tarkemmin tutkia asiakasryhmään kohdistuvan mainonnan tehokkuutta eri mainosbudjeteilla. Oletetaan että tiedossa ovat muuttujat  $S$ ,  $M$  ja  $X$ . Tällöin voidaan laskea

$$ELP = r_s \cdot p(S = s_1 | M = m_1, X_1 = x_1, \dots, X_n = x_n) - r_u \cdot p(S = s_1 | M = m_0, X_1 = x_1, \dots, X_n = x_n) - c$$

Päätöspuussa jokainen solmu  $I$  sisältää muuttujan  $X_j$  arvojen jakautumisen lapsisolmuille  $I$ . Juurisolmusta lehtisolmuun kulkemalla saadaan arvo muuttujan  $S$  todennäköisyysjakaumalle. Esimerkiksi kuvan 1 puulle todennäköisyysjakauma on  $p(S | M, X_1, X_2)$ .



kuva 1.

Erityisesti arvoilla  $\{X_1=1, X_2=2, M=m_0\}$  pätee  $p(S | X_1 = 1, X_2 = 2, M = m_0)$ , mikä saadaan laskettua myöhemmin esitettävällä algoritmilla kulkemalla puun juuresta lehtisolmuun isäsolmujen läpi.

Klassisissa päätöspuu-malleissa arviointi perustuu suurelta osin suurimman uskottavuuden käyttöön, milloin  $p(S | M, X)$ . Tämä ei ole kuitenkaan ole välttämättä paras ratkaisu kirjoitelman ongelmaan, jossa voiton maksimointi on tavoitteena. Seuraavassa esitellään ratkaisua tilanteeseen.

Jos muuttuja  $M$  asetetaan aina lehtisolmujen isäsolmuksi (kuva2), voidaan mainonnan helpoimmin muutettavissa oleva markkinointistrategian osa, mainoksen lähettämisen päätös jättää laskennassa viimeiseksi. Näin sisaruslehtien rajaamat asiakasryhmät voidaan ottaa tehokkaan markkinoinnin käsittelyyn samanlaisissa olosuhteissa, jolloin kirjoittajien Chickering ja Heckerman mukaan varmistutaan mainonnan tasaisesta vaikutuksesta asiakkaisiin.

Kirjoittajien mukaan binaarimuuttujan  $M$  sijoittaminen lehtisolmujen isäsolmuksi suuntaa toivottavasti algoritmit löytämään päätöspuita, joiden ELP:n ennustuskkyky on tavallisia puita parempi.

Olkoon  $M'$  lehtisolmun binaarinen isäsolmu. Kirjoittajat huomauttavat että samainen solmujen vaihto saattaa vaikuttaa myös negatiivisesti, koska binaarisolmuna viimeinen solmu  $M'$  vaikuttaa voimakkaasti laskentatulokseen.

Tämä ongelma poistuu kun polun  $p(S | M', X)$  arvo lasketaan solmun  $M'$  kanssa ja ilman. Jos ELP:n arvo paranee tuossa muutoksessa, poistetaan solmu  $M'$ .

## 8 Kokeelliset tulokset

Tässä osiossa esittelen tulokset kokeista, joissa käytettiin kahta ahnetta puunkasvatusalgoritmia ja niiden tekemiin puihin sovellettiin edellisessä osiossa esiteltyjä tekniikoita. Ensimmäinen algoritmi FORCE etsii puita, joiden lehtisolmujen isäsolmu on  $M$ . Sen jälkeen FORCE poistaa solmut  $M$ , jos edellisessä osiossa kuvattu ehto toteutuu. Toinen algoritmi, nimeltään NORMAL, etsii yksinkertaisesti suurimman ELP:in arvon puusta ilman  $M$  solmun jakamisen tutkimista.

Kokeen kulku oli seuraavanlainen. Koeryhmä, joka koostui satunnaisotoksesta Windows 95 :n rekisteröineistä käyttäjistä jaettiin kahteen ryhmään. Ensimmäiseen ryhmään kuului noin yhdeksänkymmentä prosenttia koko koeryhmästä. Heille lähetettiin mainos MSN:n tilaamisesta. Lopuille kymmenelle prosentille koeryhmästä ei lähetetty mainosta. Koeryhmän tilaustoimintoja tarkkailtiin tämän jälkeen määrätyn ajan verran.

Koeryhmän käyttäytymistä tarkkailemalla kerätty materiaali jaettiin seitsemänkymmentäprosenttisesti puiden opetusaineistoon ja loput testaamiseen. Puut kasvatettiin jakaumalle  $p(S | M, X)$  käyttäen algoritmeja FORCE ja NORMAL.

(Molempien algoritmien tekemiä puita arvioitiin ahneilla Bayes-tulosfunktioilla)

Algoritmien toimivuutta arvioitiin koeryhmän testiosajoukon henkilöiden aiheuttaman voiton avulla. Jokaiselle koeryhmän henkilölle laskettiin kaavan 2 avulla hänelle lähetetyn mainoksen tuotto. ELP:n muutoksen mukaan tehtiin päätös mainoksen lähettämisestä ko. henkilölle. Positiivinen ELP:n muutos johti mainoksen lähettämiseen.



Päätöksen vaikutusta verrattiin kokeen jälkeen päätöspuun malliin. Jos henkilön toiminta oli arvattavissa päätöspuun mukaan, hänen tuottama ELP lisättiin markkinoinnin kokonaisvoittoon. Jos henkilön toiminta oli mallin vastainen, poistettiin henkilö testiaineistosta ja markkinointibudjetista poistettiin henkilöön mahdollisesti kohdistuneen mainonnan tappio. Lopuksi laskettiin mainosbudjetti yhtä koehenkilöä kohden.

Kirjoittajat Chickering ja Heckerman havaitsivat että mainonnan suuntaaminen päätöspuun avulla kasvattaa mainonnasta saatavaa hyötyä lineaarisesti suhteessa käyttökertoihin. Kaikille asiakkaille suunnattu mainonta taas päinvastoin menettää hyötyään mainoskertojen lisääntyessä.

Algoritmien FORCE ja NORMAL väliset erot ELP:ssä olivat pieniä. FORCE oli kirjoittajien tuloksien mukaan hieman parempi kuin NORMAL.

Kirjoittajat ovat testauksessaan perehtyneet perusmarkkinointistrategioiden eri muotoihin ja niiden vaikutukseen markkinointibudjettiin. He ovat tuoneet esille selvän eron, joka syntyy joukkomainonnan, suoramainonnan ja henkilökohtaisen mainonnan välille. Käytännössä markkinointi ei kuitenkaan ole erillinen osa organisaatiota eikä sen budjetointia voi kohdistaa vain koeryhmien ominaisuuksien mukaan. Markkinointi ei vaikuta pelkästään potentiaaliin asiakkaisiin, vaan myös myytävän tuotteen imagoon. Myös koko yrityksen taloudellinen tila ja tulevaisuudennäkymät vaikuttavat vahvasti valittaviin markkinointilinjoihin. Kirjoittajien esittämä tapa kuvata markkinoiden käyttäytymistä on varmasti hyvä pohja markkinasuunnittelulle, mutta ei ainoa asia jonka perusteella päätökset tulisi tehdä [Lah98].

## 9 Yhteenveto

Markkinoitavien tuotteiden mainonnan suhteellinen osuus koko tuotantoketjun kuluista voidaan nähdä kapitalistisen talouden tunnusmerkkinä. Tuotteiden arvo määräytyy niihin liitetyn mielikuvan mukaan ja tätä mielikuvaa pidetään yllä tarvittavin keinoin. Tuottajan mielenkiinto kohdistuu kulujen minimointiin, minkä takia mainonnan osuus kuluissa on monesti tarkkailun alaisena.

Chickering ja Heckerman esittelevät kirjoitelmassaan kaksi tapaa vähentää mainontaan käytettyjä kuluja. Asiakkaat segmentoidaan ja mainonta suunnataan ryhmiin, joissa siitä on suurin hyöty.

Tässä kirjoitelmassa olen esitellyt Chickeringin ja Heckermanin tapoja tehostaa mainontaa. Lisäksi selvensin mainonnan pohjana olevia kaupallisia teorioita ja niiden yhteyttä päätöspuu-malliin.

## Viitteet

- [Bre84] Breiman, L., Friedman, J., Olshen, R., and Stone, C. Classification and Regression Trees. Wadsworth & Brooks, Monterey, CA, 1984.
- [Chi20] Chickering, D., Heckerman, D. A Decision Theoretic Approach to Targeted Advertising. Microsoft Research, Redmond WA, 2000.
- [Hug96] Hughes, A. The Complete Database Marketer: Second-Generation Strategies and Techniques for tapping the Power of Your Customer Database. Irwin Professional, Chicago, Ill, 1996.
- [Jen96] Jensen, V. An introduction to Bayesian Networks. UCL Press, 1996.
- [Koh20] Kohonen T. Self-Organizing Maps. Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York, 2001.
- [Lah98] Lahtinen J., Isoviita A. Markkinoinnin suunnittelu. Gummerus Kirjapaino Oy, Jyväskylä, 1998.
- [Qui93] Quinlan J.R. Programs for Machine Learning. Morgan Kaufman Publishers, Inc., CA, 1993.