Kausaalimallien kontrafaktuaalit

Paula Silvonen

Helsingin yliopisto

Kausaalimallien tutkimusseminaari

1.4.2003

1 Johdanto

"There are ever so many ways that a world might be; and one of these many ways is the way that this world is."
-David Lewis (1986)

Kontrafaktuaali tarkoittaa kirjaimellisesti faktojen vastaista. Kontrafaktuaalilause on tosi ehtolause, jonka etujäsen on tosiasioiden vastaisena epätosi, esimerkiksi ”Jos maahan olisi iskeytynyt vuonna 2000 jättimeteoriitti, ihmiselämä olisi käynyt mahdottomaksi”. Termillä kontrafaktuaalinen ajattelu viitataan tosiasiallisten tapahtumien vaihtoehtojen simulointiin.

Kontrafaktuaalianalyysin tutkiminen on yleistynyt rajusti vasta 1900-luvun loppupuolella, vaikka ensimmäisen kontrafaktuaalin eksplisiittisen määritelmän esitti jo Hume vuonna 1748: "We may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or, in other words, *where, if the first object had not been, the second never had existed*."
.
Lewis määritteli kontrafaktuaalilogiikan, joka perustuu lähimpien maailmojen käsitteelle. Lewisin mukaan lause, joka on muotoa ”Jos olisi *A*, niin olisi myös *B*”, on tosi maailmassa *w* vain siinä tapauksessa, että *B* on tosi *w*:n lähimmässä maailmassa, jossa *A* on tosi. Tässä mallissa oletetaan, että on olemassa mitta, jonka avulla maailmojen etäisyyttä toisistaan voidaan kuvata kaikille maailmoille *w* ja kaikille kuvauskielen lauseille *A*. Lewis ei kuitenkaan tällaista mittaa määritellyt.

Kausaalimallit tarjoavat formalismin kontrafaktuaalien esittämiseen matemaattisessa muodossa, jolloin kontrafaktuaalisiin kyselyihin voidaan johtaa probabilistisia vastauksia suoraan kausaalimallista mallin ominaisuuksien perusteella.

Tämän esityksen luvuissa 2-5 kuvataan kausaalimallien kontrafaktuaaliformalismia ja -analyysia; luvussa 6 esitetään yhteenveto.

2 Kausaalimallit ja rakenneyhtälöt

Joustavan rakenneyhtälön sisältävää kausaalimallia voidaan kuvata seuraavasti:

Määritelmä 1 Kausaalimalli on kolmikko

$M=<U,V,F>$,

jossa

i) U on joukko ulkoisia muuttujia, jotka määritellään mallin ulkopuolisten tekijöiden avulla
ii) V on joukko sisämuuttujia, jotka määritellään mallin muuttujien avulla
iii) F on joukko funktioita $\{f_1,f_2,...,f_n\}$, jossa jokainen f on kuvaus joukosta $U \cup (V \setminus V_i)$ muuttujaan V_i siten että F määrittelee kuvauksen joukosta U joukkoon V. (F:llä on yksikäsitteinen ratkaisu jokaiselle tilalle u joukossa U).

Määritelmä 2 (osamalli) Olkoon M kausaalimalli, X joukko muuttujia V:ssä, ja x realisaatio X:stä. Mallin M osamalli M_X on tällöin kausaalimalli

$M_X=<U,V,F_x>$,

jossa

$F_x=\{f_i:V_i \setminus X\} \cup \{X=x\}$

Toisin sanoen F_i muodostetaan poistamalla F:stä kaikki funktiot f_i, jotka vastaavat X:n jäseniä, ja korvaamalla ne joukolla funktioita $X=x$.

Määritelmä 3 (intervention vaikutus) Olkoon M kausaalimalli, X joukko muuttujia V:ssä, ja x realisaatio X:stä. Intervention $do(X=x)$ vaikutus malliin M on osamalli M_x.

Määritelmä 4 (potentiaalinen vaste) Olkoon Y muuttuja joukossa V, ja X V:n osajoukko. Y:n potentiaalinen vaste interventioon $do(X=x)$, merkitään $Y_x(u)$, on ratkaisu Y:lle yhtälöjoukosta F_x.

Määritelmä 5 (kontrafaktuaali) Olkoon Y muuttuja joukossa V, ja X V:n osajoukko. Kontrafaktuaalinen lause ”Arvo, jonka Y olisi saanut, jos X olisi ollut x” tulkitaan merkitsemällä samoin kuin potentiaalinen vaste $Y_x(u)$.

i) Jos $Y=V_i$, ja $X=V \setminus Y$, $Y_x(u)=f_i(pa_i,u)$ jossa pa_i on funktion $X=x$ projektio PA_i:lle. Näin jokainen M:n funktio f_i voi saada kontrafaktuaalisen tulkinnan; se määrittelee V_i:n potentiaalisen vasteen kaikkien muiden V:n muuttujien hypoteettiselle manipulaatiolle.
ii) Jos Y sisältyy joukkoon X ja $X=x \Rightarrow Y=y$, silloin $Y_x(u)=y$. Näin manipuloidun muuttujan potentiaalinen vaste on sama kuin manipulaatiossa asetettu arvo.

Määritelmä 6 (probabilistinen kausaalimalli) Probabilistinen kausaalimalli on pari

$$<M,P(u)>,$$

jossa M on kausaalimalli ja $P(u)$ on U:n yli määritelty todennäköisyysfunktio.

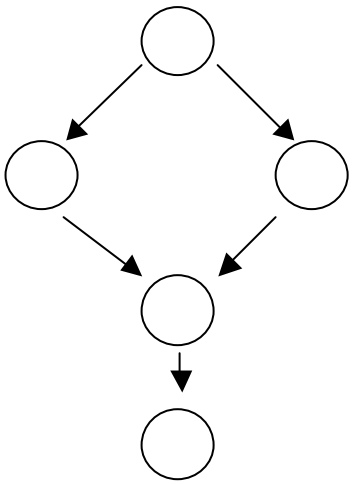
Koska jokainen sisämuuttuja on U:n funktio, $P(u)$ määrittelee todennäköisyysjakauman sisämuuttujien yli. Näin ollen jokaiselle muuttujajoukolle $Y \subseteq V$, saadaan

$$P(y) = \sum_{\{u|Y(u)=y\}} P(u)$$

Kontrafaktuaalilauseiden todennäköisyys määritellään samoin osamallin M_X indusoiman funktion $Y_X(u)$ avulla:

$$P(Y_X = y) = \sum_{\{u|Y_x(u)=y\}} P(u)$$

3 Esimerkki



Kuva 1 Viiden muuttujan kausaalimalli

Kuva 1 määrittelee kausaaliset suhteet vuodenajan, sateen, sadettajan käytön, jalkakäytävän märkyyden ja liukkauden välillä. Vuodenajan arvo on yksi joukosta (talvi, kevät, kesä, syksy) ja muut muuttujat saavat arvon tosi tai epätosi.

Malli voidaan kuvata viidellä funktiolla:

$x_1 = u_1$
$x_2 = f_2\ (x_1, u_2)$
$x_3 = f_3\ (x_1, u_3)$
$x_4 = f_4\ (x_3, x_2, u_4)$
$x_5 = f_5(x_4, u_5)$

Virhetermit (disturbances) $U_1, ..., U_5$ on jätetty pois kuvasta, mutta niiden voidaan ajatella kuvaavan kaikkia mahdollisia mallin ulkopuolisia (toisistaan riippumattomia) virhetekijöitä.

Funktioita ja niiden virhetekijöitä voitaisiin kuvata Boolen algebralla seuraavasti:

$x_2 = [(X_1 = talvi) \vee (X_1 = syksy)\ V\ ab_2] \wedge \neg\ ab'_2$
$x_3 = [(X_1 = kesä) \vee (X_1 = kevät)\ V\ ab_3] \wedge \neg\ ab'_3$
$x_4 = (x_2 \vee x_3 \vee ab_4) \wedge \neg\ ab'_4$
$x_5 = (x_4 \vee ab_5) \wedge \neg\ ab'_5$

joissa funktiot x_i kuvaavat tilannetta $X_i = tosi$ ja virhetermit ab_i ja $ab´_i$ säännöttömyyksiä. Esimerkiksi ab_4 voisi kuvata tilannetta, jossa vettä on kaatunut jalkakäytävälle (eli jalkakäytävä on märkä vaikka ei sada eikä sadettaja ole päällä).

3.1 Interventio ja kontrafaktuaali

Interventio $do(X_3 = tosi)$ (eli sadettajan laittaminen päälle) voidaan kuvata korvaamalla $x_3 = f_3\,(x_3,u_3)$ funktiolla $x_3 = tosi$. Tuloksena saatu osamalli $M_{x3\,=\,tosi}$ sisältää kaiken informaation, joka tarvitaan muiden muuttujien intervention jälkeisten arvojen laskemiseksi. Intervention perusteella voisimme päätellä tällöin, että jalkakäytävä on intervention jälkeen märkä ja liukas, koska $x_3 = tosi$ vaikuttaa x_4:n ja sen välityksellä x_5:n arvoihin.

Kontrafaktuaalista päättelyä, kun löydettäisiin sadettaja käynnissä, voidaan kuvata lisäämällä yhtälöryhmään $x_3 = tosi$ poistamatta alkuperäistä x_3:n yhtälöä, jolloin voimme päätellä, että todennäköisesti vuodenaika on kesä tai kevät, ei sada, ja jalkakäytävä on märkä ja liukas.

Onkin tärkeää huomata ero tekemisen ja havaitsemisen välillä; sadettajan laittaminen päälle ei automaattisesti lopeta sadetta, mutta jos havaitsemme sadettajan olevan päällä, voimme kausaalimallimme perusteella päätellä, että tietyllä todennäköisyydellä ulkona ei sada.

4 Kompositio, tehokkuus, palautuvuus

Kontrafaktuaaleilla on kolme ominaisuutta, jotka pätevät kaikille kausaalimalleille - kompositio, tehokkuus ja palautuvuus.

Kompositio

Jokaiselle kahdelle erilliselle muuttujalle Y ja W, ja jokaiselle muuttujajoukolle X kausaalimallissa, saadaan

$W_x(u) = w \Rightarrow Y_{xw}(u) = Y_x(u)$,

eli jos pakotamme muuttujalle W arvon, jonka se olisi saanut ilman interventiota, interventio ei vaikuta mallin muihin muuttujiin. Samoin tyhjää interventiota (null action) voidaan merkitä $Y_{Æ}(u) = Y(u)$.

Komposition seurauksena saadaan

(Konsistenssi) Kaikille kausaalimallin muuttujille Y ja X, $X(u) = x \Rightarrow Y(u) = Y_x(u)$.

Tehokkuus

Kaikille muuttujille X ja W, $X_{xw}(u) = x$

Tehokkuus määrittelee intervention vaikutuksen manipuloituun muuttujaan itseensä; jos pakotamme muuttujalle X arvon x, X saa arvon x :)

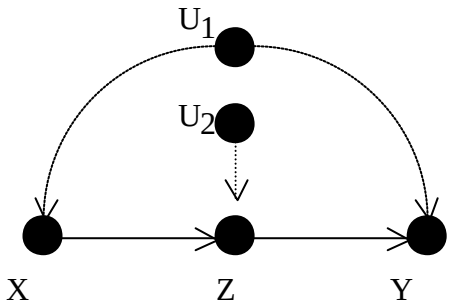
Palautuvuus

Mille tahansa muuttujalle Y ja W, ja jokaiselle muuttujajoukolle X,
$(Y_{xw}(u) = y) \,\&\, (W_{xy}(u) = w) \Rightarrow Y_x(u) = y$.

Jos X:lle pakotetaan arvo x, joka johtaa Y:n arvoon y, ja jos Y:n arvon pakottaminen y:ksi johtaa X:n arvoon x, tällöin X:llä ja Y:llä on arvot x ja y ilman interventiotakin.

Todistukset komposition, tehokkuuden ja palautuvuuden pitävyydestä kaikille kausaalimalleille löytyvät lähteestä Galles ja Pearl.

5 Kontrafaktuaalianalyysi vs. graafinen malli



Kuva 2 Kausaalimalli tupakoinnin vaikutuksesta keuhkosyövän syntyyn

Tarkastellaan kuvan 2 tilannetta, jossa on esitetty tupakoinnin, tervan kerääntymisen keuhkoihin, ja keuhkosyövän välisiä kausaalisia suhteita. Kausaalimallia voidaan kuvata seuraavasti:

V={X (Tupakointi), Y (Keuhkosyöpä), Z (tervaa keuhkoissa)}
U={U_1,U_2}, U_1 _||_ U_2
x=f_1 (u_1)
z=f_2 (x,u_2)
y=f_3 (z,u_1)

Mallin ulkoiselle muuttujalle U_1 voidaan ajatella tulkinta ”keuhkosyöpää ja nikotiininhimoa aiheuttava genotyyppi”.

Graafinen malli sisältää mm. seuraavat oletukset:

i) Tupakoinnin vaikutus keuhkosyöpään aineellistuu kokonaan tervan kerääntymisessä keuhkoihin
ii) Vaikka genotyyppi vaikuttaisi syövän syntyyn, se ei vaikuta tervan kerääntymiseen muuten kuin tupakoinnin välityksellä

Kontrafaktuaaloanalyysin avulla voimme kuvitella tilannetta, jossa *X, Y* ja *Z* on mitattu yhtäaikaisesti suurelta, satunnaisesti valitusta populaation osasta. Tästä datasta haluamme päätellä keuhkosyöpäriskin kahden hypoteettisen vaihtoehdon avulla, tupakoinnin ja tupakoimattomuuden. Formaalisti, haluamme johtaa todennäköisyyden tilanteelle *Y=y*, kun *do(X=x), P(Y = y | do(x)) = P(Yx = y)*, pohjautuen yhteisjakaumalle *P(x,y,z)* ja graafisen mallin sisältämille oletuksille. Jotta tämä saavutettaisiin, oletukset on käännettävä kontrafaktuaalien kielelle. Tämä voidaan tehdä kahden säännön avulla:

Sääntö 1 Poissulkevat rajoitteet Jokaiselle muuttujalle *Y*, jolla on vanhemmat PA_Y, ja jokaiselle PA_Y:stä eroavalle muuttujajoukolle *Z*, saadaan

$$Y_{paY}(u) = Y_{paYz}(u)$$

Sääntö 2 Itsenäisyysrajoitteet Jos $Z_1, ..., Z_k$ on mikä tahansa *Y*:hyn vain ulkomuuttujia sisältävällä polulla kytkeytymätön solmujoukko *V*:ssä, saadaan

$$Y_{paY} \perp\!\!\!\perp \{Z_{1paZ1}, ..., Z_{kpaZk}\}$$

Sääntö 1 kuvaa sitä, että interventio *Y*:lle jää vaille vaikutusta, jos *Y*:n syyt pysyvät vakioina. Sääntö 2 samaistaa muuttujien *U* riippumattomuuden ja vastaavien *V*:n kontrafaktuaalien riippumattomuuden, jos niiden vanhemmat pidetään muuttumattomina.

Soveltamalla ylläolevia sääntöjä saadaan kausaaliverkon sisältämiksi oletuksiksi

$$Z_x(u) = Z_{yx}(u)$$
$$X_y(u) = X_{zy}(u) = X_z(u) = X(u)$$
$$Y_z(u) = Y_{zx}(u)$$
$$Z_x \perp\!\!\!\perp \{Y_z, X\}$$

Kolme ensimmäistä yhtälöä on saatu soveltamalla sääntöä 1 ($PA_X = \{Æ\}$, $PA_Y = \{Z\}$, $PA_Z = \{X\}$), viimeinen säännöstä 2 (koska U_1 ja U_2 eivät ole kausaalisessa suhteessa keskenään).

Näiden oletusten ja komposition ja tehokkuuden avulla voidaan päätellä seuraavaa:

Tupakoinnin kausaalinen vaikutus tervan kerääntymiseen, $P(Z_X = z)$:

$$P(Z_X = z) = P(Z_X = z \mid x) \quad \text{(oletus)}$$
$$= P(Z = z \mid x) \quad \text{(kompositio)}$$
$$= P(z \mid x)$$

Tervan kausaalinen vaikutus syövän syntyyn $P(Y_Z = y)$:

$$P(Y_Z = y) = \sum_x P(Y_z = y \mid x)P(x)$$

ja koska $Y_Z \perp\!\!\!\perp Z_X \mid X$, (oletus)

$$P(Y_Z = y \mid x) = P(Y_Z = y \mid x, Z_X = z) \quad \text{(oletus)}$$
$$= P(Y_Z = y \mid x, z) \quad \text{(kompositio)}$$
$$= P(y \mid x, z) \quad \text{(kompositio)}$$

$$\Rightarrow P(Y_Z = y) = \sum_x P(y \mid x, z)P(x)$$

Tupakoinnin kausaalinen vaikutus syövän syntyyn $P(Y_X = y)$:

Kaikille muuttujille Z,

$Y_X(u) = Y_{XZ}(u)$, *jos* $Z_X(u) = z$ (kompositio)

Koska $Y_{XZ}(u) = Y_Z(u)$, (oletus)

$Y_X(u) = Y_{XZ}(u) = Y_Z(u)$, jossa $z_X = Z_X(u)$

Tällöin

$$P(Y_X = y) = P(Y_{ZX} = y)$$

$$= \sum_z P(Y_{ZX} = y \mid Z_X = z)\, P(Z_X = z)$$

$$= \sum_z P(Y_Z = y \mid Z_X = z)\, P(Z_X = z) \quad \text{(kompositio)}$$

$$= \sum_z P(Y_{ZX} = y)\, P(Z_X = z) \quad \text{(oletus)}$$

Sijoittamalla saamme

$$P(Y_X = y) = \sum_z P(z \mid x) \sum_{x'} P(y \mid z, x')\, P(x'),$$

joka on siis todennäköisyys tupakoinnin kausaaliselle vaikutukselle syövän syntymisessä.

Kontrafaktuaalinen $P(Y_X = y)$ on identifioituva, koska se voidaan pelkistää havainnoitujen muuttujien todennäköisyyksiksi. On todistettu (Galles ja Pearl), että kaikki identifioituvat kontrafaktuaalit voidaan pelkistää vastaavalla tavalla käyttämällä vain kompositiota ja tehokkuutta.

6 Yhteenveto ja johtopäätökset

Tutkimusten mukaan (Sloman, Lagnado) useimmat ihmiset käyttävät rationaalista kontrafaktuaalista päättelyä. Tieteellisiä sovellutuksia kontrafaktuaaliselle päättelylle löytyy esimerkiksi sosiaalipsykologian, ekonometrian, ja monien muiden alojen piiristä.

Kontrafaktuaalikyselyjen evaluointi on hyödyllistä monissa päättelytehtävissä. Esimerkki tällaisesta tilanteesta on vastuun määrittäminen (esimerkiksi ”Jos et olisi tullut puhumaan minulle Limeksen bileissä, emme olisi tavanneet, jolloin emme olisi koskaan ostaneet yhdessä asuntoa; siispä olet vastuussa asuntolainastamme”). Balke ja Pearl esittävät formaalin notaation, semantiikan ja päättelyalgoritmit, joiden avulla kontrafaktuaalisia kyselyjä voidaan evaluoida probabilistisesti.

Tässä paperissa keskityttiin kontrafaktuaalien aksiomaattiseen määrittelyyn, ja rakenneyhtälöiden ja kontrafaktuaalien yhtenevyyteen. Nämä tekniikat yhdistettynä graafisiin kausaalimalleihin antavat työkaluja kausaalisten ja kontrafaktuaalisten päättelyjen matemaattiseen ratkomiseen.

Lähteet

Balke, Alexander and Pearl, Judea, Probabilistic Evaluation of Counterfactual Queries.

Galles, David and Pearl, Judea, An Axiomatic Characterization of Causal Counterfactuals.

Lewis, David, Counterfactuals. Cambridge, Harvard University Press, 1973.

Menzies, Peter, Counterfactual Theories of Causation, *The Stanford Encyclopedia of Philosophy (Spring 2001 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2001/entries/causation-counterfactual/>.