Luonnollisen kielen tilastollinen käsittely

T-61.281 (3 ov) L

Kevät 2003

Luennot: **Timo Honkela ja Krista Lagus**

Laskuharjoitukset: **Vesa Siivola**

Luentokalvot: Krista Lagus (päivityksiä: Timo Honkela)

8. N-grammi-kielimallit

8.1 Tilastollinen mallinnus

1. Otetaan dataa (generoitu tuntemattomasta tn-jakaumasta)
2. Tehdään estimaatti jakaumasta datan perusteella
3. Tehdään päätelmiä uudesta datasta jakaumaestimaatin perusteella

Mallinnuksen osatehtävät voidaan hahmottaa seuraavasti:

- Datan jakaminen ekvivalenssiluokkiin
- Hyvän tilastollisen estimaattorin löytäminen kullekin luokalle
- Useiden estimaattorien yhdistäminen

Tyypillinen oletus: **stationaarisuus**, eli että datan tn-jakauma ei muutu oleellisesti ajan myötä.

Tilastollisen kielimallin tehtävistä

Klassinen tehtävä: seuraavan sanan (tai kirjaimen) ennustaminen jo nähtyjen sanojen (tai kirjainten) perusteella ('Shannon game'). Esim. seuraavissa sovelluksissa:

- puheentunnistus
- optinen merkkientunnistus, käsinkirjoitettujen merkkien tunnistus
- kirjoitusvirheiden korjaus
- tilastollinen konekääntäminen

Estimointimenetelmät yleisiä, soveltuvat myös muihin tehtäviin (esim. WSD, word sense disambiguation, jäsentäminen)

8.2 N-grammimallit

N-grammimalli: ennustetaan sanaa w_n edellisten $n-1$ sanan perusteella:

$$P(w_n|w_1 w_2 \cdots w_{n-1}) \tag{1}$$

Kaava esiintyy myös muodossa $P(w_t|w_{t-(n-1)} w_{t-(n-2)} \cdots w_{t-1})$ jossa t viittaa sanan järjestysnumeroon (ajanhetkeen) koko aineistossa.

Esimerkki: aineistona tämän luennon kalvot, n=4:

		w_{t-3}	w_{t-2}	w_{t-1}	w_t		
...	sitä	enemmän	dataa	tarvitaan	**mallin**	estimointiin	...

Malleille käytettäviä nimiä

n=1	unigram
n=2	bigram
n=3	trigram
n=4	4-gram, fourgram

Yhteys ekvivalenssiluokkiin: n-grammimallissa jokainen $n-1$:n sanan pituinen historia saa oman ekvivalenssiluokkansa. Tämä tarkoittaa että historiat

joissa viimeiset 3 sanaa samoja käsitellään keskenään identtisinä tilanteina seuraavan sanan ennustamisen kannalta, eli niillä on yhteinen estimaatti.

Sama n-grammien ominaisuus toisesta näkökulmasta: malli olettaa että sana riippuu ainoastaan $(n-1)$ edeltävästä sanasta, mutta ei tätä kauempana olevista sanoista (ns. Markov-oletus).

Markov-malli: k:n asteen Markov-malli on malli joka asettaa kaikki k:n pituiset historiat samaan ekvivalenssiluokkaan. Ts. n-grammimalli on $n-1$:n asteen Markov-malli.

Esimerkkejä:
Sue swallowed the large green ____

Samppa Lajunen voitti kultaa ____

Parametrien määrän kasvu

	Malli	**Parametreja jos sanasto 20,000**
n=1	unigram	20000
n=2	bigram	$20000^2 = 400$ milj.
n=3	trigram	$20000^3 = 8$ miljardia
n=4	4-gram, fourgram	1.6×10^{17}

8.3 Piirteiden jakaminen ekvivalenssiluokkiin

- Piirteet (sekä jatkuva-arvoiset että diskreetit) voidaan jakaa ekvivalenssiluokkiin 'bins'
- Esim. jatkuva-arvoisen muuttujan 'ikä' jakaminen luokkiin 0-2; 3-5; 7-10; 11-15; 16-25; 26-35 jne
- Mitä useampia ekv.luokkia, sitä enemmän dataa tarvitaan mallin estimointiin, jotta tulokset *luotettavia* kullekin luokalle
- Toisaalta, jos luokkia on kovin vähän, ennustettavan kohdemuuttujan (esim. 'pituus') arvoa ei voida ennustaa kovin *tarkasti*.

Esimerkki: ennustetaan seuraavaa sanaa
1. kolmen edellisen sanan sanaluokan (subst, verbi, adj, num jne) TAI
2. kolmen edellisen sanan perusteella

1. tapauksessa vähemmälläkin datalla jonkinlaiset estimaatit, kun taas
2. tapauksessa tarkempia estimaatteja mutta dataa tarvitaan paljon enemmän.

Joitain tapoja muodostaa ekvivalenssiluokkia

- Isojen ja pienten kirjainten käsittely samalla tavalla (esim. kaiken muuntaminen pieniksi kirjaimiksi)
- Sanojen muuntaminen perusmuotoon (saman sanan eri taivutusmuodot käsitellään ekvivalentteina)
- Ryhmittely sanaluokkatiedon mukaan (syntaktiselta rooliltaan samankaltaiset muodostavat ekv. luokan)
- Sanojen semanttinen ryhmittely (merkitykseltään samankaltaiset muodostavat ekv.luokan)

Kussakin vaihtoehdossa tarvitaan kuitenkin menetelmä jolla sanan ekvivalenssiluokka voidaan luotettavasti päätellä.

Lisäksi ekvivalenssiluokkien olisi hyvä olla sellaisia että niiden sisällä sanat todella käyttäytyvät samankaltaisesti, ts. tarkkuus säilytetään.

Historian huomioimisen eri tapoja

Edellä kuvattiin yksittäisten piirteiden ekvivalenssiluokkien laskemista. Eri tapoja ekvivalenssiluokkien muodostamiseen historian suhteen:

- Poimitaan historiasta tiettyjä piirteitä, mutta niiden sijainnilla ei ole väliä esim. malli: $P(w_t|\text{lauseen predikaatti}, w_{t-1})$
- Käsitellään sanajonon sijaan sanajoukkoa ('sanasäkkiä', bag-of-words), eli ei välitetä sanojen järjestyksestä:
  $P(w_n|w_1, w_2, \ldots, w_{n-1})$

8.4 N-grammimallin tilastollinen estimointi

Annettuna: joukko näytteitä jotka osuvat kuhunkin ekvivalenssiluokkaan (biniin). Bayesin kaavoista:

$$P(w_n|w_1 \cdots w_{n-1}) = \frac{P(w_1 \cdots w_n)}{P(w_1 \cdots w_{n-1})} \tag{2}$$

Mallin optimointi: maksimoidaan datan todennäköisyys (eli sanojen tn:ien tulo).

Notaatio:

N	Opetusnäytteiden lukumäärä
B	Ekv.luokkien (binien) lukumäärä
w_{1n}	n-grammi $w_1 \cdots w_n$
$C(w_1 \cdots w_n)$	ngrammin $w_1 \cdots w_n$ lukumäärä opetusdatassa
r	n-grammin lukumäärä
N_r	Niiden binien lukumäärä joissa on r näytettä
h	historia (edeltävä sanajono)

Maximum likelihood-estimaatti (MLE)

$$P_{\mathsf{MLE}}(w_1 \cdots w_n) = \frac{C(w_1 \cdots w_n)}{N} \tag{3}$$

$$P_{\mathsf{MLE}}(w_n | w_1 \cdots w_{n-1}) = \frac{C(w_1 \cdots w_n)}{C(w_1 \cdots w_{n-1})} \tag{4}$$

- MLE-estimointi johtaa parametrien valintaan siten että opetusdatan todennäköisyys maksimoituu.
  (Huom: tämä pätee vain tietyin oletuksin, kuten että näytteet, esim. tri-grammien sanakolmikot, oletetaan riippumattomiksi toisistaan. Tämä taas ei pidä paikkaansa mm. overlapin takia.)
- Koko tn-massa jaetaan opetusdatassa esiintyneiden tapausten kesken, niiden frekvenssien suhteessa.
- Antaa siis tn=0 tapaukselle jota ei nähty opetusdatassa, eli ei jätä lainkaan tn-massaa aiemmin näkemättömille sanoille.

- Koska yleisesti sanajonon tn lasketaan kertomalla kunkin sanan tn, yksikin nolla saa koko sanajonon tn:n nollaksi.
- Esimerkki datan harvuudesta: ensimmäisten 1.5 miljoonan sanan jälkeen (IBM laser patent text corpus) 23% myöhemmistä trigrammeista oli ennennäkemättömiä.
- MLE ei kovin hyödyllinen estimaatti harvalle datalle, kuten n-grammeille.
- Tarvitaan siis systemaattinen tapa jolla huomioidaan ennennäkemättömi sanojen ja ennennäkemättömien n-grammien tn:t. Tätä kutsutaan mm. nimellä *tasoitus* eli *smoothing*

Taulukko 6.3: MLE-estimaatteja Austenin kirjoista eräälle lauseelle eri n-grammeilla.

In person	*she*		*was*		*inferior*		*to*	
1-gram	$P(\cdot)$		$P(\cdot)$		$P(\cdot)$		$P(\cdot)$	
1	the	0.034	the	0.034	the	0.034	the	0.034
2	to	0.032	to	0.032	to	0.032	**to**	**0.032**
3	and	0.030	and	0.030	and	0.030		
4	of	0.029	of	0.029	of	0.029		
...								
8	was	0.015	**was**	**0.015**	was	0.015		
...								
13	**she**	**0.011**			she	0.011		
...								
254					both	0.0005		
...								
435					sisters	0.0003		
...								
1701					**inferior**	**0.00005**		
2-gram	$P(\cdot \mid person)$		$P(\cdot \mid she)$		$P(\cdot \mid was)$		$P(\cdot \mid inferior)$	
1	and	0.099	had	0.141	not	0.065	**to**	**0.212**
2	who	0.099	**was**	**0.122**	a	0.052		
3	to	0.076			the	0.033		
4	in	0.045			to	0.031		
...								
23	**she**	**0.009**						
...								

Laplacen laki eli 'yhden lisäys'

Annetaan hiukan tn-massaa näkemättömille tapauksille lisäämällä jokaiseen lukuun 1:

$$P_{\mathsf{LAP}}(w_1 \cdots w_n) = \frac{C(w_1 \cdots w_n) + 1}{N + B} \tag{5}$$

- Vastaa Bayesin estimaattia priorilla, että kaikki tapahtumat ovat yhtä todennäköisiä, ja tähän prioriin uskotaan aivan kuin olisi nähty yksi näyte joka lajia.
- Esim. 44 milj. sanan AP newswire-korpus, sanaston koko 400,653 sanaa, jolloin bigrammeja 1.6×10^{11}, eli $N =$44 milj., $B = 1.6 \times 10^{11}$
- Jos data on hyvin harvaa, antaa liiaksi tn-massaa ennen näkemättömille tapauksille (tässä 46.5% tn-massasta).
- Ts. uskotaan tasajakauma-prioriin liian vahvasti verrattuna datan määrää
- Kannattaisiko 1:n sijaan uskoa että ollaan nähty esim. 0.0001 jokaista näytettä?

Odotetun frekvenssin estimaatteja seuraavassa taulukossa:

$r = f_{MLE}$	$f_{empirical}$	f_{Lap}	f_{del}	f_{GT}	N_r	T_r
0	0.000027	0.000137	0.000037	0.000027	74 671 100 000	2 019 187
1	0.448	0.000274	0.396	0.446	2 018 046	903 206
2	1.25	0.000411	1.24	1.26	449 721	564 153
3	2.24	0.000548	2.23	2.24	188 933	424 015
4	3.23	0.000685	3.22	3.24	105 668	341 099
5	4.21	0.000822	4.22	4.22	68 379	287 776
6	5.23	0.000959	5.20	5.19	48 190	251 951
7	6.21	0.00109	6.21	6.21	35 709	221 693
8	7.21	0.00123	7.18	7.24	27 710	199 779
9	8.26	0.00137	8.18	8.25	22 280	183 971

Table 6.4 Estimated frequencies for the AP data from Church and Gale (1991a). The first five columns show the estimated frequency calculated for a bigram that actually appeared r times in the training data according to different estimators: r is the maximum likelihood estimate, $f_{empirical}$ uses validation on the test set, f_{Lap} is the 'add one' method, f_{del} is deleted interpolation (two-way cross validation, using the training data), and f_{GT} is the Good-Turing estimate. The last two columns give the frequencies of frequencies and how often bigrams of a certain frequency occurred in further text.

Lidstonen laki, Jeffreys-Perksin laki

$$P_{\text{Lid}}(w_1 \cdots w_n) = \frac{C(w_1 \cdots w_n) + \lambda}{N + B\lambda} \tag{6}$$

Voidaan osoittaa että ylläoleva tarkoittaa lineaarista interpolointia tasajakauma priorin ja MLE-estimaatin välillä. Asetetaan $\mu = N/(N + B\lambda)$:

$$P_{\text{Lid}}(w_1 \cdots w_n) = \mu \frac{C(w_1 \cdots w_n)}{N} + (1 - \mu)\frac{1}{B} \tag{7}$$

- Jeffreysin priori: $\lambda = 1/2$, eli lisätään jokaiseen lukumäärään $1/2$ (vastaa sitä että olisi nähty puolikas näyte jokaista lajia). Käytetään myös nimeä *Expected Likelihood Estimation* (ELE)
- On valittava λ:n arvo tavalla tai toisella
- Alhaisilla frekvensseillä tämäkään ei kovin hyvin vastaa todellista jakaumaa

Good-Turing -estimaattori

Ks. frekvenssien frekvenssi-histogrammeja taulukossa 6.7.

$$P_{GT}(w_1 \cdots w_n) = \frac{r^*}{N}, \text{ jossa } r^* = \frac{(r+1)S(r+1)}{S(r)} \tag{8}$$

ja $S(r)$ on odotusarvo N_r:lle, tai vaihtoehtoisesti, arvo joka on saatu sovittamalla jokin tasainen käyrä frekvenssien frekvensseille: $N_r = S(r)$

Simple Good-Turing -estimaattori: Valitaan käyräksi potenssifunktio: $S(r) = ar^b$ jossa parametrit a ja b sovitetaan frekvenssien frekvenssi-histogrammin mukaan.

Melko hyvä estimaattori, yleisesti käytössä.

	Bigrams				Trigrams		
r	N_r	r	N_r	r	N_r	r	N_r
1	138741	28	90	1	404211	28	35
2	25413	29	120	2	32514	29	32
3	10531	30	86	3	10056	30	25
4	5997	31	98	4	4780	31	18
5	3565	32	99	5	2491	32	19
6	2486		...	6	1571		...
7	1754	1264	1	7	1088	189	1
8	1342	1366	1	8	749	202	1
9	1106	1917	1	9	582	214	1
10	896	2233	1	10	432	366	1
	...	2507	1		...	378	1

Table 6.7 Extracts from the frequencies of frequencies distribution for bigrams and trigrams in the Austen corpus.

Muita tasoitusmenetelmiä

Termi 'discounting' viittaa siihen, että nähtyjen n-grammien tn:iä alennetaan ja tätä massaa jaetaan ennen näkemättömille.

- Absoluuttinen alennus (absolute discounting): Kaikista nähdyistä n-grammeista vähennetään vakio-tnmassa σ joka jaetaan tasan näkemättö n-grammien kesken.
- Lineaarinen alennus (linear discounting): Skaalataan nähtyjen n-grammie tn:ät vakiolla joka on hiukan pienempi kuin 1, ja saatu tnmassa jaetaan tasan ei-nähtyjen kesken. Ei kovin hyvä, koska 'rankaisee' frekventtejä enemmän—kuitenkin niiden estimaatit ovat parempia.
- Witten-Bell discounting: Arvioidaan yllättävien asioiden näkemisen tn-massa sen perusteella kuinka tavallista yllättävien asioiden näkeminen on ollut tähän mennessä: $\sum_{i:C(i)=0} p_i = \frac{T}{N+T}$ jossa T on tähän mennessä nähtyjen binien määrä.

Pohjimmiltaan menetelmien eroissa on kyse on siitä minkälaisia oletuksia tehdään tapauksista, joita ei ole nähty, ja niiden suhteesta tapauksiin, joita on nähty.

Huom: Esim. CMU Statistical Language Toolkit toteuttaa useita eri discountin ja tasoitusmenetelmiä n-grammeille.

8.5 Estimaattorien yhdistäminen

- Tähän asti tarkasteltu tilannetta jossa pyritään estimoimaan identtinen tn esim. kaikille 3-grammeille joita ei ole nähty.
- Kuitenkin jos 3-grammin osat (esim. 2-grammit) ovat frekventtejä, eikö niistä kerättyä tietoa kannattaisi käyttää 3-grammin tn:n estimoinnissa?
- Motivaationa estimaattien tasoitus (smoothing) tai yleisemmin eri informaationlähteiden yhdistäminen.

Lineaarinen interpolointi

(yleisemmin nimellä äärelliset mikstuurimallit tai sum of experts)

Lasketaan painotettu keskiarvo eri pituisten kontekstien antamista estimaateista:

$$P_{li}(w_n|w_{n-2}w_{n-1}) = \lambda_1 P_1(w_n) + \lambda_2 P_2(w_n|w_{n-1}) + \lambda_3 P_3(w_n|w_{n-2}w_{n-1}) \quad (9)$$

($0 \leq \lambda_1 \leq 1$ ja $\sum_i \lambda_1 = 1$)

Parametrit λ voidaan asettaa käsin tai optimoida datan avulla.

Yleinen lineaarinen interpolointi

Edellä parametrit λ eivät riippuneet sanoista joiden kohdalla niitä sovelletaan, eli parametri on vakio vaikkapa kaikille bi-grammeille.

Yleisemmin ne voidaan kuitenkin asettaa riippumaan historiasta:

$$P_{li}(w|h) = \sum_i \lambda_i(h) P_i(w|h) \tag{10}$$

($0 \leq \lambda_1 \leq 1$ ja $\sum_i \lambda_1 = 1$) ja optimoida esim. EM-algoritmilla. Kuitenkin, jos jokaiselle historialle on oma λ ollaan taas datan harvuusongelmassa, ja joudutaan soveltamaan jotain tasoitusta, historioiden ekvivalenssiluokkia tms.

Perääntyminen (backing off)

- Periaate: Katsotaan aina spesifeintä mallia joka antaa 'riittävän luotettavaa' informaatiota tämänhetkisestä kontekstista.
- Eli peräännytään pitkien kontekstien käytöstä yhä lyhempiin: Päätetään uskoa estimaattia jos se perustuu vähintään k näytteeseen (k esim. 1 tai 2)
- Kritiikkiä: Uuden opetusdatan lisääminen voi vaikuttaa voimakkaasti tn:iin kun se aiheuttaa muutoksia useiden sanojen kohdalla niille sovellettavissa n-grammipituuksissa
- Kuitenkin mallit yksinkertaisia ja toimivat melko hyvin, joten yleisesti käytössä.
- back-off -malli on erikoistapaus yleisestä lineaarisesta interpoloinnista: $\lambda_i(h) = 1$ kun k:n arvo riittävän suuri, 0 muulloin.
- Lähestymistapa muistuttaa Kohosen Dynamically Expanding Context (DEC) -algoritmia.

Back-off-mallien käyttöesimerkki:

	$P(she\|h)$	$P(was\|h)$	$P(inferior\|h)$	$P(to\|h)$	$P(both\|h)$	$P(sisters\|h)$
Unigram	0.011	0.015	0.00005	0.032	0.0005	0.0003
Bigram	0.00529	0.1219	0.0000159	0.183	0.000449	0.00372
n used	2	2	1	2	2	2
Trigram	0.00529	0.0741	0.0000162	0.183	0.000384	0.00323
n used	2	3	1	2	2	2

Table 6.11 Probability estimates of the test clause according t
guage models. The unigram estimate is our previous MLE unigram
other two estimates are back-off language models. The last col
overall probability estimate given to the clause by the model.

8.6 Mallien estimoinnista yleisesti

Seuraava koskee mitä tahansa menetelmien vertailua, ei pelkästään n-grammej tai kielimalleja.

Held-out estimation

Tavallisesti data jaetaan ennen menetelmien kehittämistä kolmeen osaan

- **Opetusjoukko:** data jolla malli opetetaan
- **Validointijoukko:** opetusjoukosta riippumaton data, jonka avulla valitaan mallin opetuksessa käytettävät parametrit (esim. edellisen kalvon λ)
- **Testijoukko:** edellisistä riippumaton, satunnaisesti valittu datajoukko (kooltaan esim. 10% opetusdatasta), jolla lopullisen mallin hyvyys mitataan.

Testijoukko on pidettävä kokonaan syrjässä menetelmien kehittämisen ai-

kana! Jos testijoukko pääsee vaikuttamaan menetelmänkehitykseen (vaikka vain alitajuisesti), se ei ole enää soveltuva menetelmän testaamiseen.

Kuitenkin usein menetelmänkehitys on syklinen prosessi jossa välillä muutetaan menetelmää ja sitten taas testataan. Siksi voi olla erikseen:
1. **kehittely-testijoukko**, jolla vertaillaan menetelmän eri variantteja
2. **lopullinen testijoukko** jolla tuotetaan julkaistavat tulokset, ja jota ei ole käytetty mihinkään ennen tätä.

Vaihtoehdot testijoukon (ja validointijoukon) valintaan:
1. täysin satunnainen valinta (satunnaisia lyhyitä tekstinpätkiä)
2. pitkiä yhtenäisiä pätkiä (esim. ajallisesti myöhempiä osia datasta)

2-tapa vastaa paremmin mallin käyttötilannetta: se myös antaa realistisemmat, yleensä hiukan huonommat tulokset johtuen siitä, että harvat ilmiöt ovat täysin stationaarisia.

Eri menetelmien vertailusta

Pelkkiä keskiarvotuloksia vertaamalla ei voi tietää ovatko havaitut erot menetelmissä merkitseviä.

Eräs ratkaisu: Mitataan lisäksi tulosten varianssi eri datajoukoilla, ja testataan erojen tilastollinen merkitsevyys esim. t-testillä.

	System 1	System 2
scores	71, 61, 55, 60, 68, 49, 42, 72, 76, 55, 64	42, 55, 75, 45, 54, 51 55, 36, 58, 55, 67
total	609	526
n	11	11
mean $\bar{x}_i$	55.4	47.8
$s_i^2 = \sum(x_{ij} - \bar{x}_i)^2$	1,375.4	1,228.8
df	10	10

$$\text{Pooled } s^2 = \frac{1375.4+1228.8}{10+10} \approx 130.2$$

$$t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{2s^2}{n}}} = \frac{55.4-47.8}{\sqrt{\frac{2\cdot 130.2}{11}}} \approx 1.56$$

Table 6.6 Using the t test for comparing the performance of two systems. Since we calculate the mean for each data set, the denominator in the calculation of variance and the number of degrees of freedom is $(11 - 1) + (11 - 1) = 20$. The data do not provide clear support for the superiority of system 1. Despite the clear difference in mean scores, the sample variance is too high to draw any definitive conclusions.

Ristiinvalidointi (cross-validation)

- Jaetaan data K:hon osajoukkoon, joista 1 kerrallaan on testidata, muut opetusdataa. Toistetaan siten että kukin osajoukko on vuorollaan testidata. K välillä 2. . . N, jossa N datan määrä.
- Hyöty: Kaikki datat vaikuttavat sekä mallin opetukseen että sen testaukseen, data siis hyödynnetään mahdollisimman tarkasti (tärkeää etenkin kun dataa on vähän).
- useita eri variantteja (deleted estimation, leave-one-out-estimation)

Sekä ristiinvalidoinnin että held-out-estimoinnin avulla voidaan valita mallien parametrejä, ja siis esim. tasoittaa tn-estimaatteja.

8.7 N-grammimallin kritiikkiä

N-grammien ongelmia kielimallina:

- Eivät huomioi pidemmän tähtäimen riippuvuuksia sanojen välillä
- Sanajono yhdessä järjestyksessä ei kontribuoi saman sanajoukon tn:ään jossain toisessa järjestyksessä
- Tasoitusongelmat voi myös nähdä mallin rakenteellisena ongelmana
- Riippuvuudet estimoidaan sanojen välillä suoraan. Intuitiivisesti järkevämmältä tuntuisi että olisivat osaksi joidenkin latenttien muuttujien, kuten käsitteiden ja/tai sanaluokkien tms välillä.
- Kuitenkin: n-grammimalli yhdistää syntaktiset ja semanttiset ja kollokationaaliset lyhyen kontekstin riippuvuudet käytännössä yllättävänkin hyvin toimivalla tavalla, etenkin/ainakin englannille.
- Mallin optimointiin ja tasoitusmenetelmien parantamiseen on käytetty hyvin paljon resursseja. On mahdollista, että on juututtu lokaaliin minimiin malliperheiden suhteen.