

Probabilistic FastText for Multi-SenseWord Embeddings

Ben Athiwaratkun*

Andrew Gordon Wilson

Anima Anandkumar

読み手：江原 遥

静岡理工科大学/産業技術総合研究所

2018/8/3

なぜこの論文を読むのか

[Vilnis and McCallum 14]

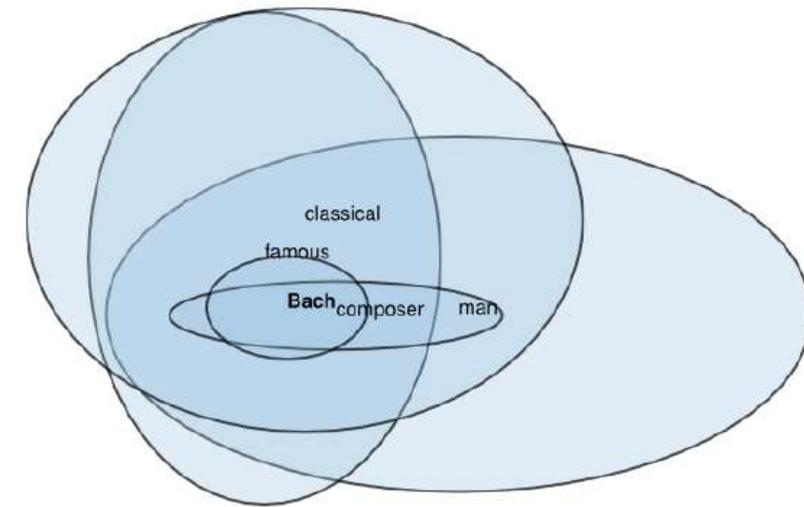
- SNLP2018の読むべき論文推薦で2位
分散表現：点→分布 の拡張

言語処理学会2014

ガウス分布による単語と句の意味の分布的表現

島岡 聖世 † 村岡 雅康 † 山本 風人 † 渡邊 陽太郎

† 岡崎 直観 †* 乾 健太郎



Ben Athiwaratkun*

Cornell Ph.D.
student

Andrew Gordon Wilson

Cornell

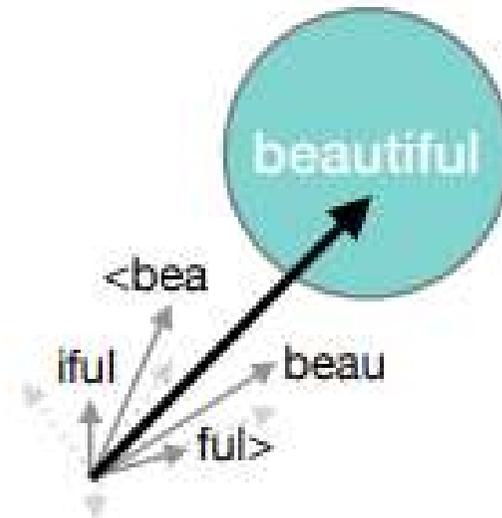
Anima Anandkumar

目的

- subwordも考慮したembeddingを作りたい
例：circumnavigation, dogz
fasttext [がこれをやっていた
- 同時に多義性も考慮したい
例：pop, rock

sub-wordの
Gaussian Mixtureで分散表現を作ろう

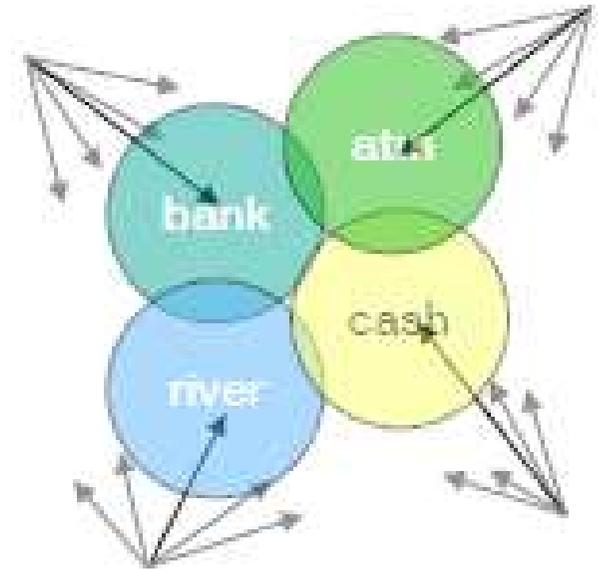
Mixtureのcomponent: 1つの意味



語の多義性→Gaussian Mixture

$$f(x) = \sum_{i=1}^K p_{w,i} \mathcal{N}(x; \mu_{w,i}, \Sigma_{w,i})$$

(すべての) 語にはK個の意味があると仮定
語 w の i 番目の意味の
分散表現の中心 $\mu_{w,i}$
共分散 $\Sigma_{w,i}$

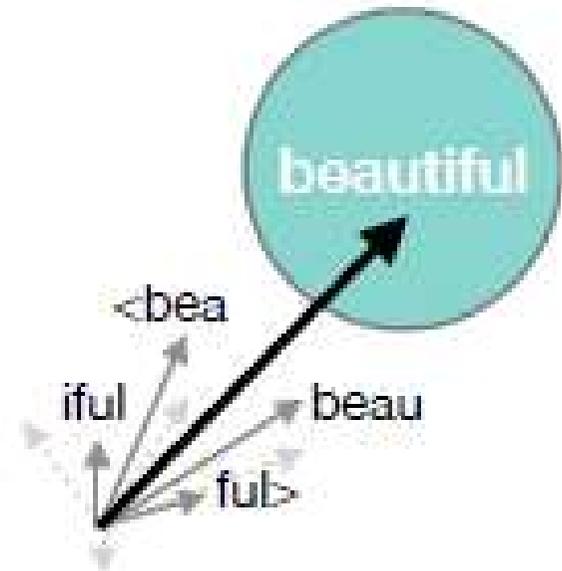


sub-wordも考慮して分散表現を作る

$$\mu_w = \frac{1}{|NG_w| + 1} \left(v_w + \sum_{g \in NG_w} z_g \right)$$

語 w の文字N-gram: NG_w
sub-word : z_g

- 3-grams: $\langle \text{be, bea, eau, aut, uti, tif, ful, ul} \rangle$
- 4-grams: $\langle \text{bea, beau ..., iful, ful} \rangle$



語が分布なら語同士の類似度は？

語と語の類似度は点なら典型的には：内積
分布でもやっぱり内積

$$\log \langle f, g \rangle_{L_2} = \log \int f(x)g(x) dx.$$

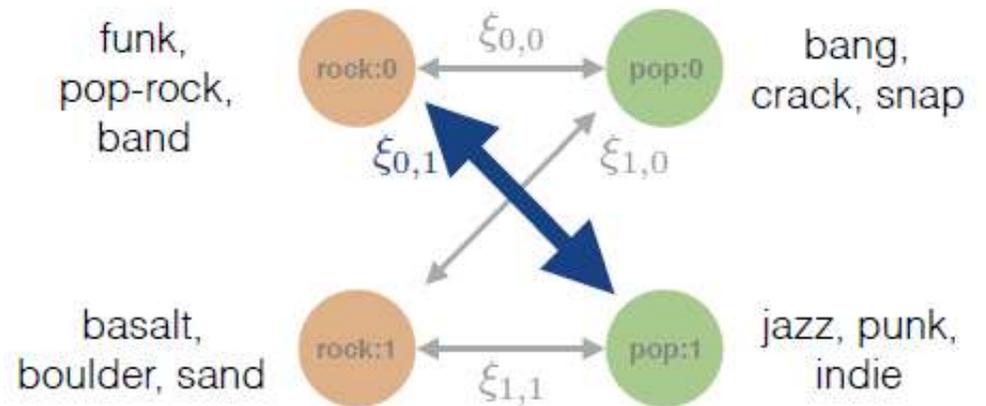
$$\begin{aligned} f(x) &= \sum_{i=1}^K p_i \mathcal{N}(x; \vec{\mu}_{f,i}, \Sigma_{f,i}) & \sum_i p_i &= 1 \\ g(x) &= \sum_{i=1}^K q_i \mathcal{N}(x; \vec{\mu}_{g,i}, \Sigma_{g,i}) & \sum_i q_i &= 1 \end{aligned}$$

$$E(f, g) = \log \sum_{j=1}^K \sum_{i=1}^K p_i q_j e^{\xi_{i,j}}$$

ξ の意味

$$E(f, g) = \log \sum_{j=1}^K \sum_{i=1}^K p_i q_j e^{\xi_{i,j}}$$

ξ 大 \rightarrow エネルギー大



$$\begin{aligned} \xi_{i,j} &\equiv \log \mathcal{N}(0; \vec{\mu}_{f,i} - \vec{\mu}_{g,j}, \Sigma_{f,i} + \Sigma_{g,j}) \\ &= -\frac{1}{2} \log \det(\Sigma_{f,i} + \Sigma_{g,j}) - \frac{D}{2} \log(2\pi) \\ &\quad - \frac{1}{2} (\vec{\mu}_{f,i} - \vec{\mu}_{g,j})^\top (\Sigma_{f,i} + \Sigma_{g,j})^{-1} (\vec{\mu}_{f,i} - \vec{\mu}_{g,j}) \end{aligned}$$

実際には対角成分だけ考えて...

$$\xi_{i,j} = -\frac{\alpha}{2} \cdot \|\mu_{f,i} - \mu_{g,j}\|^2,$$

訓練

positive context pair: w, c

negative context pair: w, n

マージン最大化

w の分布 : f

c の分布 : g

n の分布 : n

$$L(f, g) = \max [0, m - E(f, g) + E(f, n)]$$

訓練に使う語は、高頻度語に集中してサンプルされるのを防ぐために、

$$P(w) = 1 - \sqrt{t/f(w)}$$

$f(w)$: w の頻度

$U(w)$: w の uni-gram 確率

負例は $P_n(w) \propto U(w)^{3/4}$

$$K=2, t=10^{-5}$$

結果：PFT-GMが提案

D	50				300				
	W2G	W2GM	PFT-G	PFT-GM	FASTTEXT	W2G	W2GM	PFT-G	PFT-GM
SL-999	29.35	29.31	27.34	34.13	38.03	38.84	39.62	35.85	39.60
WS-353	71.53	73.47	67.17	71.10	73.88	78.25	79.38	73.75	76.11
MEN-3K	72.58	73.55	70.61	73.90	76.37	78.40	78.76	77.78	79.65
MC-30	76.48	79.08	73.54	79.75	81.20	82.42	84.58	81.90	80.93
RG-65	73.30	74.51	70.43	78.19	79.98	80.34	80.95	77.57	79.81
YP-130	41.96	45.07	37.10	40.91	53.33	46.40	47.12	48.52	54.93
MT-287	64.79	66.60	63.96	67.65	67.93	67.74	69.65	66.41	69.44
MT-771	60.86	60.82	60.40	63.86	66.89	70.10	70.36	67.18	69.68
RW-2K	28.78	28.62	44.05	42.78	48.09	35.49	42.73	50.37	49.36
AVG.	42.32	42.76	44.35	46.47	49.28	47.71	49.54	49.86	51.10

Table 2: Spearman’s Correlation $\rho \times 100$ on Word Similarity Datasets.

結果 2

多義語に着目した
評価用データセット
でもよくなっているよ

Model	Dim	$\rho \times 100$
HUANG AVGSIM	50	62.8
TIAN MAXSIM	50	63.6
W2GM MAXSIM	50	62.7
NEELAKANTAN AVGSIM	50	64.2
PFT-GM MAXSIM	50	63.7
CHEN-M AVGSIM	200	66.2
W2GM MAXSIM	200	65.5
NEELAKANTAN AVGSIM	300	67.2
W2GM MAXSIM	300	66.5
PFT-GM MAXSIM	300	67.2

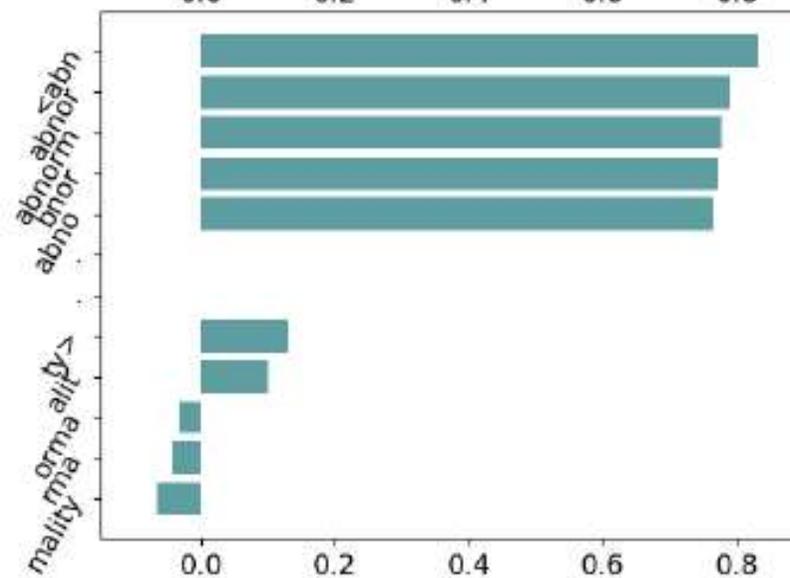
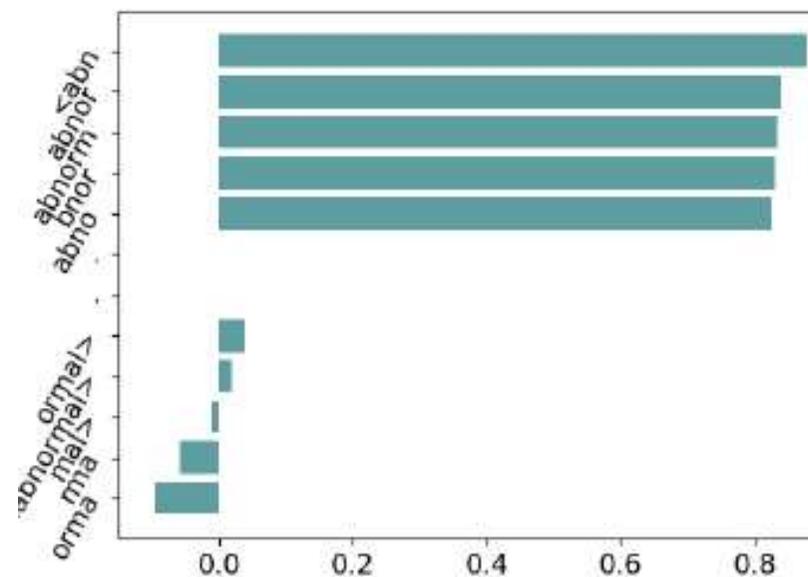
Table 3: Spearman’s Correlation $\rho \times 100$ on word similarity dataset SCWS.

結果 3

Word	Co.	Nearest Neighbors
rock	0	rock:0, rocks:0, rocky:0, mudrock:0, rockscape:0, boulders:0 , coutcrops:0,
rock	1	rock:1, punk:0, punk-rock:0, indie:0, pop-rock:0, pop-punk:0, indie-rock:0, band:1
bank	0	bank:0, banks:0, banker:0, bankers:0, bankcard:0, Citibank:0, debits:0
bank	1	bank:1, banks:1, river:0, riverbank:0, embanking:0, banks:0, confluence:1
star	0	stars:0, stellar:0, nebula:0, starspot:0, stars.:0, stellas:0, constellation:1
star	1	star:1, stars:1, star-star:0, 5-stars:0, movie-star:0, mega-star:0, super-star:0
cell	0	cell:0, cellular:0, acellular:0, lymphocytes:0, T-cells:0, cytes:0, leukocytes:0
cell	1	cell:1, cells:1, cellular:0, cellular-phone:0, cellphone:0, transcellular:0
left	0	left:0, right:1, left-hand:0, right-left:0, left-right-left:0, right-hand:0, leftwards:0
left	1	left:1, leaving:0, leavings:0, remained:0, leave:1, enmained:0, leaving-age:0, sadly-departed:0

Word	Nearest Neighbors
rock	rock, rock-y, rockn, rock-, rock-funk, rock/, lava-rock, nu-rock, rock-pop, rock/ice, coral-rock
bank	bank-, bank/, bank-account, bank., banky, bank-to-bank, banking, Bank, bank/cash, banks.**
star	movie-stars, star-planet, G-star, star-dust, big-star, starsailor, 31-star, star-lit, Star, starsign, pop-stars
cell	cellular, tumour-cell, in-cell, cell/tumour, 11-cell, T-cell, sperm-cell, 2-cells, Cell-to-cell
left	left, left/joined, leaving, left,right, right, left)and, leftsided, lefted, leftside

sub-wordが
どれだけ
きいてる？



感想

分布で表現するとパラメタが増えすぎてうまくいかないのので
できる限りパラメタを減らしている感

- 語については共分散行列を考慮するがsub-wordについては平均ベクトルのみ
- 共分散行列の対角成分だけ考える
- K=2決め打ち

K=2決め打ちは、「現実（にデータから考慮できる多義性）はそんなものだ、現実を見ろ」と言われている感

μ については評価しているが Σ については評価していない

- Σ は μ をよくするために使われているだけ。本当は、 Σ に entailmentが表れるようにしたいですね…