

Deriving Boolean Structures from Distributional Vectors

Germán Kruszewski, Denis Paperno, Marco Baroni

TACL 2015, pages 375-388

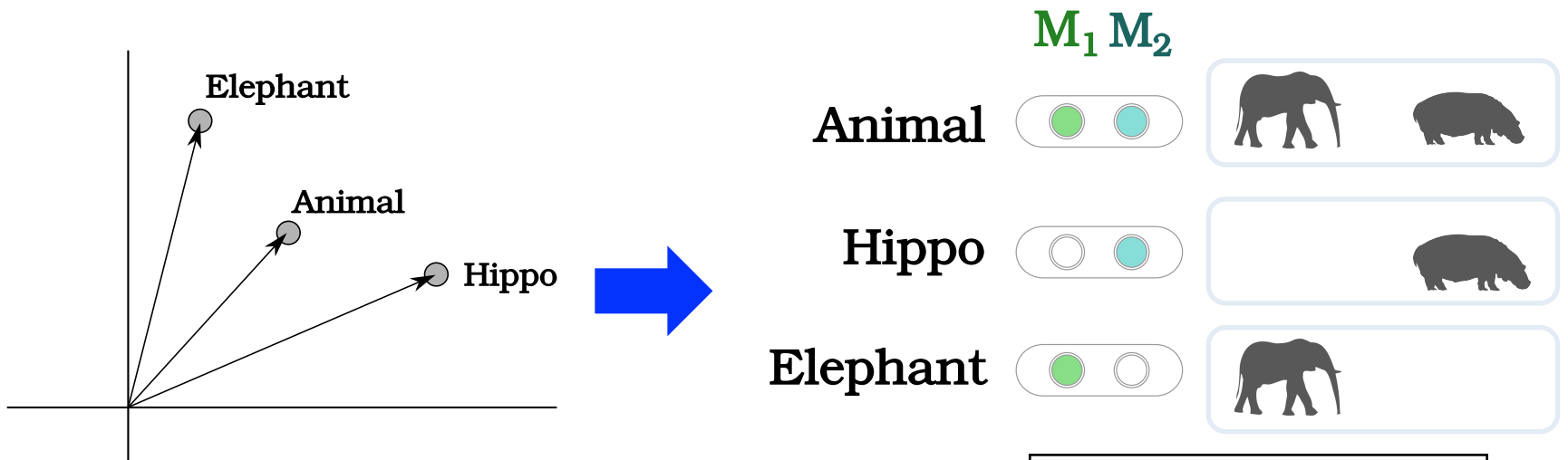
<http://aclweb.org/anthology/Q15-1027>

読み手： 東北大学 情報科学研究科 乾・岡崎研究室

修士1年 小林 颯介

要約

- 単語の Distributional (real-valued) vector を 単語や文の Boolean vector へ変換
 - 各次元の値の包含で Entailment が判定可能なように



- 含意認識性能ますます

各次元で
右が成り立つ
(論理包含)

下位語		上位語
1か0	←	1
0	←	0

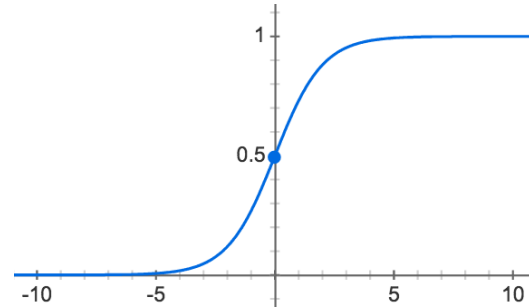
変換

線型写像と活性化関数で二値に変換

$$M_{\Theta}(x) = g(Wx + b)$$

$$g(x) = \frac{1}{1 + e^{-x}}$$

シグモイド関数



二値ベクトル
値域 $[0, 1]$
ただしほぼ0か1
疎



M_{Θ}

元ベクトル
値域 $(-\infty, \infty)$ とか $[0, \infty)$
密



M_{Θ}



変換の学習

p→q (正例)なら h=1
 q→p (負例)なら h=0

$$J(\Theta) = \frac{1}{2} \sum_{k=1}^m (h_{\Theta}(p_k, q_k) - y_k)^2$$

+ L2正則化
 + 変換後ベクトルの sparseにする項

$$h_{\Theta} = \min_i w_{\Theta_i}$$

全て1なら1、そうでなければ0

$$w_{\Theta_i} = \max(1 - r_{\Theta_i}, s_{\Theta_i})$$

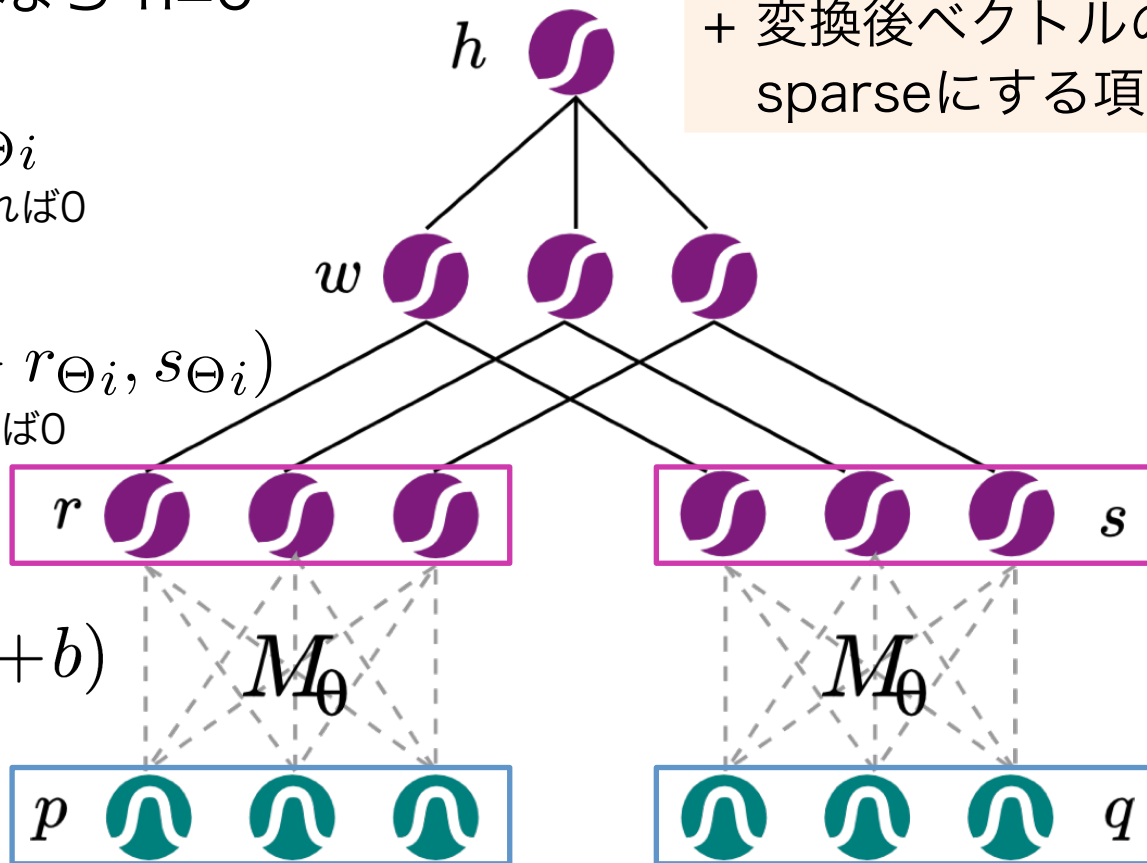
r ≤ s なら1、そうでなければ0

$$M_{\Theta}(x) = g(Wx + b)$$

実際は近似

$$\max(x, y) \approx \frac{\log(e^{Lx} + e^{Ly})}{L}$$

$$\min(w_{\theta}) \approx -\log\left(\frac{\sum_i e^{-Lw_{\theta_i}}}{L}\right)$$



実際(テスト時)の判定

$$BI(u, v) = \frac{\sum_i \text{rnd}(M_{\Theta}(u)_i) \text{rnd}(M_{\Theta}(v)_i)}{\sum_i \text{rnd}(M_{\Theta}(u)_i)}$$

1. M_{θ} で変換

$u=r, v=s$

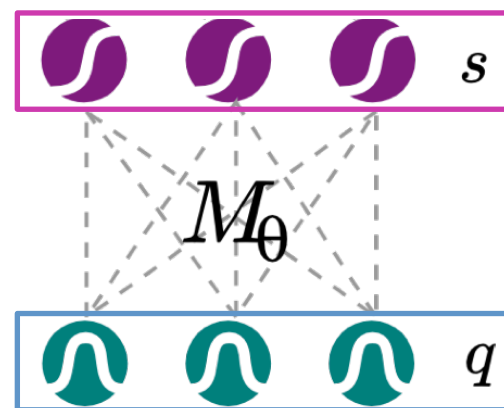
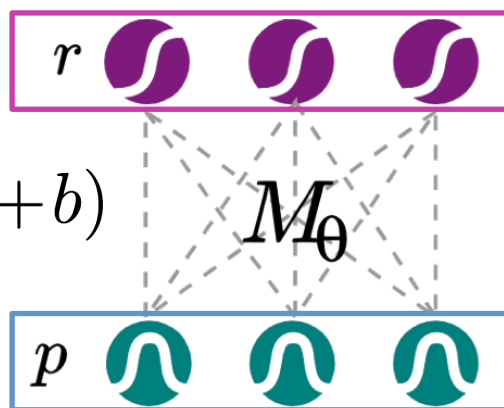
2. 0.5を閾値に二値化

$$\text{rnd}(x) = \mathbb{1}[x > 0.5]$$

3. BIスコア (本論文考案)

を求めて、ある閾値で判定

$$M_{\Theta}(x) = g(Wx + b)$$



実際(テスト時)の判定

$$BI(u, v) = \frac{\text{rとsが共に1である次元の数}}{\text{rが1である次元の数}}$$

1. M_θ で変換

$$u=r, v=s$$

rが1である次元の数

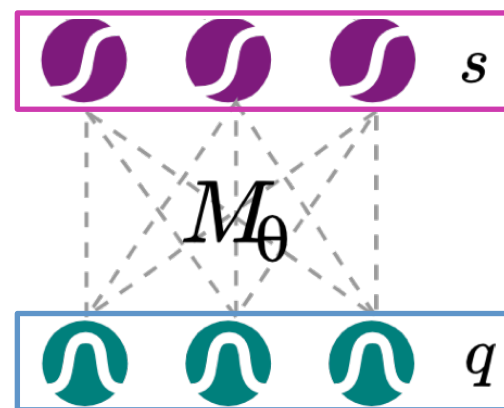
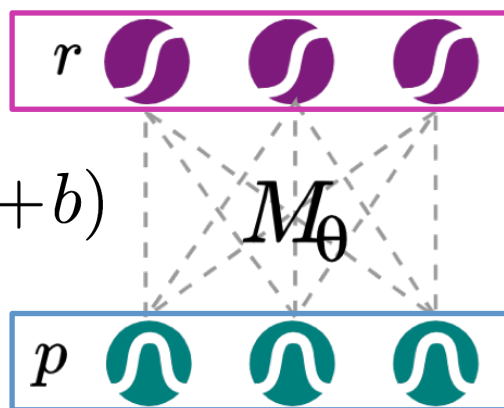
2. 0.5を閾値に二値化

$$\text{rnd}(x) = \mathbb{1}[x > 0.5]$$

3. BIスコア (本論文考案)

を求めて、ある閾値で判定

$$M_\Theta(x) = g(Wx + b)$$



単語ペア 実験設定

ベクトル

1. C-BOW (次元数: 400 → 1000)
2. PPMI-SVD (次元数: 300 → 100)
3. TypeDM (次元数: 400? → 500)

(Baroni, 2010) 係り受けパスの種類数 から生成

判定手法

1. BDSM (提案手法)
2. 元のベクトルで SVM-2次カーネル
3. 元のベクトルから直接スコアを算出する手法5種

e.g.,

$$weedsPrec(u, v) = \frac{\sum_i \mathbb{1}[v_i > 0] \cdot u_i}{\sum_i u_i}$$

単語ペア データセット

データ	負例	特徴
LEDS-core 911:911	1/3:下,上位語 2/3: ランダム	類似度見るだけでも割と楽
LEDS-dir 911:911	下,上位語	各単語(個別)の抽象度を見るだけでいいかも
BLESS-coord 1236:3526	同格 (coordinate)	むずかしい
BLESS-mero 1236:2943	部分-全体	

学習は

- ・ LEDSは
10-交差検定
- ・ BLESSは
LEDS-core
(重複正例除く)

	<i>Positive</i>	<i>Negative</i>
LEDS	elephant → animal	ape → book
LEDS-dir		animal → elephant
BLESS-coord	elephant → herbivore	elephant → hippo
BLESS-mero		elephant → trunk

単語ペア 結果

- 提案手法はSVMに
勝ったり負けたり

- 難しい
BLESS-mero
では大差で勝利

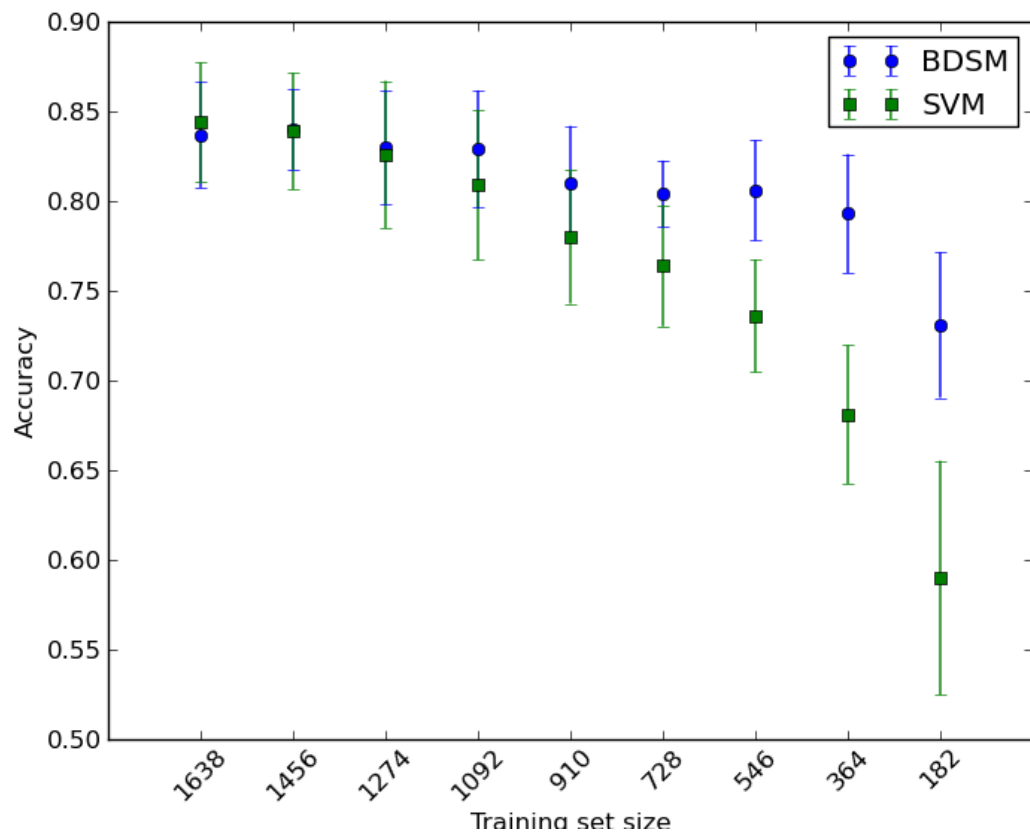
- 元ベクトルから
スコア算出系は弱い

<i>model</i>	LEDS		BLESS	
	<i>core</i>	<i>dir</i>	<i>coord</i>	<i>mero</i>
<i>count</i>				
clarkeDE	77	63	27	36
weedsPrec	79	75	27	33
cosWeeds	79	63	26	35
invCL	77	63	27	36
balAPinc	79	66	26	36
SVM (count)	84	90	55	57
BDSM (count)	83	87	53	55
<i>predict</i>				
SVM (predict)	71	85	70	55
BDSM (predict)	80	79	76	68
<i>TypeDM</i>				
SVM (TypeDM)	78	83	56	60
BDSM (TypeDM)	83	71	31	59

単語ペア 結果

BDSMは少量のデータでも学習効果が高い

1638→364 (2/9) にしても性能は 83%→約79%



文含意 実験設定

データ: SICK (Marelli et al., 2014), 訓練:4500, テスト:4927


文の構成法

1. 全単語ベクトル総和
2. Practical lexical function model (Paperno et al., 2014)

<i>Positive</i>	<i>Negative</i>
A man is slowly trekking in the woods → The man is hiking in the woods	A group of scouts are camping in the grass → A group of scouts are hiking through the grass

比較手法

1. 全部含意
2. 全部非含意
3. 単語の重なり数でスコア
4. BDSM (提案手法)
5. SVM



dogs	\vec{dogs}
run	$r\vec{un}, \overset{\square}{run}$
dogs run	$r\vec{un} + \overset{\square}{run} \times \vec{dog}$
house	\vec{house}
big	$\vec{big}, \overset{\square}{big}$
big house	$\vec{big} + \overset{\square}{big} \times \vec{house}$

(Paperno et al., 2014)より引用

文含意 結果

- BDSMはSVMに勝利

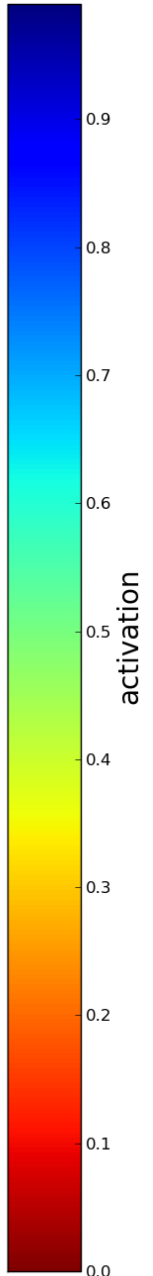
- 単なる総和で文構成した方が良い結果 (語順考慮なし)

- 単語重なり数 (WO) でも結構匹敵する性能になっていてなんだかなあ

<i>model</i>	P	R	F1	A
Sycophantic	29	100	45	29
Majority	0	0	0	71
WO	40	86	55	60
SVM (add)	47	54	51	70
BDSM (add)	48	74	58	69
SVM (plf)	39	45	42	64
BDSM (plf)	44	71	55	66
SVM(add) + WO	44	82	58	65
BDSM(add) + WO	48	80	60	69
SVM(plf) + WO	42	76	54	63
BDSM(plf) + WO	42	77	54	63

議論・考察

- 活性化率 (1値の次元の率)
 - word2vec上での car, newspaperの隣接30単語
- media, press, volvoなどはそれ自体は教師なし
- **活性化率は抽象度と相関あり**
 - WordNetでの深さ
 - 抽象度辞書 (Brysbaert et al., 2013)



議論・考察

- thing, anythingなどは全次元1になったが、benefit, achieveなど怪しい単語も全次元1だった (計768語)
 - そもそも次元数少ないから表現力が厳しい
 - そもそも訓練データが具体名詞ばかり
- 外延的な解釈が“そこそこ”可能
 - **類似度が表示対象の集合の重なり**の量に対応元のベクトルに比べて
 - co-hyponym は類似度が低くなる
 - 人名同士 は類似度が低くなる
 - synonym は類似度が高くなる
 - hypo/hypernym まで類似度が低くなってしまった

まとめ

- 分布意味論由来のベクトルを
外延的な意味成分の二値ベクトルへと
変換するニューラルネットの学習法を提案した
 - 意味の含意関係認識において
より**少ない教師データで効率的に学習可能**
 - **単なる類似度では解けない精密な識別能力が**
求められるタスクで**SVMを上回った**

おまけ 個人的感想

- コンセプトとモデルがシンプルに直結していて納得感
 - B1スコアはちょっと直感的でない
- 定量評価の実験には多少のもやもや
 - 勝ったり負けたり の割にその原因究明は薄かった
- 定性的な観察の結果はなかなか興味深い
 - 各次元個別での解釈は可能かも知りたかった
- WordNetまるごとなどでしっかり学習したらエンコード性能はかなり高まるのだろうか