Innovative R&D by NTT

知識グラフの埋め込みとその応用

創情部 創言 G 林 克彦
hayashi.katsuhiko@lab.ntt.co.jp

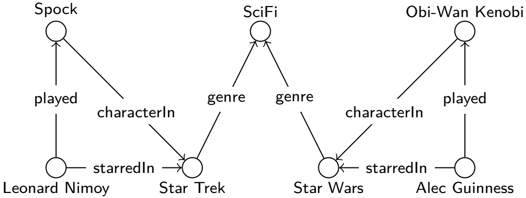
NTT コミュニケーション科学基礎研究所

目次

- 知識グラフとは?

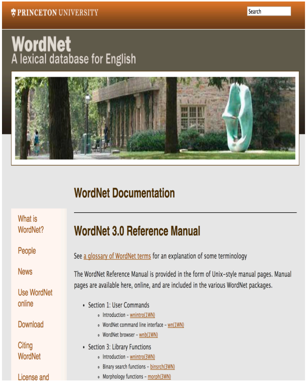
知識 (ナレッジ) グラフとは?

- 知識: Entity 間の Relation で記述
- 知識グラフ $\mathcal{G} = (V, E)$
  - V: 頂点の集合 (頂点 $=$ Entity)
  - $E \subseteq V \times \mathcal{R} \times V$: 辺の集合 (辺 $s \xrightarrow{r} o =$ Fact(s, r, o))
    - $\mathcal{R}$: Relation の集合
    - s を Subject, o を Object と呼ぶ



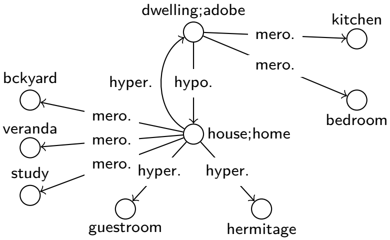
- 実例: WordNet, Freebase, DBPedia, Yago など [LSS+11]

WordNet [Fel98]

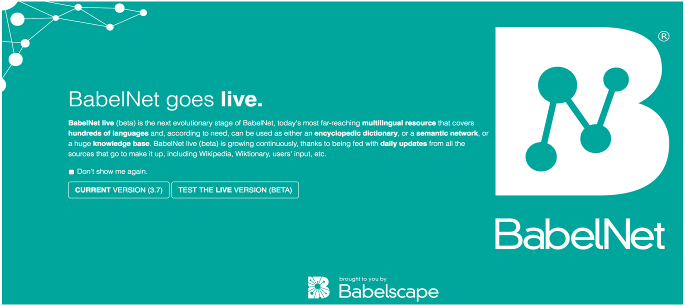


https://wordnet.princeton.edu/wordnet/documentation/

- 語彙シソーラスデータ
  - 同義語グループ (synset) の集合 (約 11 万 5 千種類)
  - 名詞，動詞，形容詞，副詞の約 15 万語を収録
  - synset 間に関係を定義
    - hypernym (上位語)，hyponym (下位語)，holonym (全体語)，meronym (部分語)



BabelNet [NP10]



http://babelnet.org/

- Wikipedia のページ (Entity) を WordNet へマッピング
  - Wikpedia のリンクや Google 翻訳などを利用して，271 言語に渡る大規模なシソーラスを構築

Freebase [BEP+08]

https://developers.google.com/freebase/

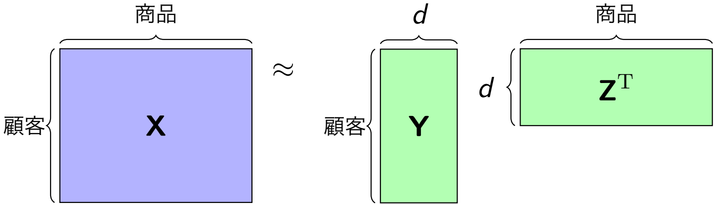
- <u>人間の一般知識を協力的に構築</u>
  - Google に吸収され，Knowledge Graph API で検索可能
  - Wikidata への統合が行われている [PTVS+16]

目次

- 知識グラフとは?
- 知識グラフに対する関係学習
  - 基本事項

関係学習の例 [鹿島 09, 石黒 16]

- 商品推薦システム
  - 顧客と商品との間の関係 (評価や購買) $\rightarrow$ 行列 $\mathbf{X}$ で表現
  - 2 つの行列 $\mathbf{Y}$ と $\mathbf{Z}$ に分解 (ランク d 行列分解)

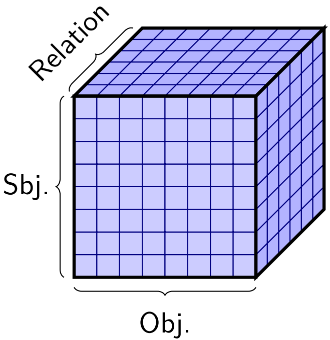


$*$ 2 乗誤差 $\arg\min_{\mathbf{Y},\mathbf{Z}} ||\mathbf{X} - \mathbf{Y}\mathbf{Z}^{\mathrm{T}}||_F^2$ を最小化して学習

- 顧客 i の商品 j への関係を予測
  - $\mathbf{y}_i$: $\mathbf{Y}$ の i 行ベクトル, $\mathbf{z}_j$: $\mathbf{Z}$ の j 行ベクトル

$$x_{i,j} \approx \mathbf{y}_i^{\mathrm{T}} \mathbf{z}_j$$

3 次隣接テンソルによる知識グラフ表現



- <u>Relation r に対する隣接行列 $\mathbf{X}_r$</u>

$$x_{s,r,o} = \begin{cases} 1 & (s,r,o) \in E \\ 0 & (s,r,o) \notin E \end{cases}$$

0	1	0	0	0	1	0	1
1	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0

$\Rightarrow$ <u>$x_{s,r,o} \approx f(s,r,o;\Theta)$ となる予測関数を設計したい</u>

様々な予測関数

モデル	予測関数 $f(s,r,o;\Theta)$	パラメータ Θ
CP 分解	$\langle \mathbf{w}_r, \mathbf{y}_s, \mathbf{z}_o \rangle$	$\mathbf{w}_r, \mathbf{y}_s, \mathbf{z}_o \in \mathbb{R}^d$
DistMult [YYH+14]	$\langle \mathbf{w}_r, \mathbf{y}_s, \mathbf{y}_o \rangle$	$\mathbf{w}_r, \mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d$
RESCAL [NT13]	$\mathbf{y}_s^{\mathrm{T}} \mathbf{W}_r \mathbf{y}_o$	$\mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d, \quad \mathbf{W}_r \in \mathbb{R}^{d \times d}$
HolE [NRP16]	$\mathbf{w}_r^{\mathrm{T}} (\mathbf{y}_s \star \mathbf{y}_o)$	$\mathbf{w}_r, \mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d$
ComplEx [TWR+16]	$Re(\langle \mathbf{w}_r, \mathbf{y}_s, \overline{\mathbf{y}_o} \rangle)$	$\mathbf{w}_r, \mathbf{y}_s, \mathbf{y}_o \in \mathbb{C}^d$
ANALOGY [LWY17]	$\mathbf{y}_s^{\mathrm{T}} \mathbf{W}_r^{mn} \mathbf{y}_o$	$\mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^{m+2n}$ $\mathbf{W}_r^{mn} \in \mathbb{R}^{(m+2n)\times(m+2n)}$
SE [BWCB11]	$-\lVert \mathbf{W}_{r,1} \mathbf{y}_s - \mathbf{W}_{r,2} \mathbf{y}_o \rVert_1$	$\mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d, \quad \mathbf{W}_{r,1}, \mathbf{W}_{r,2} \in \mathbb{R}^{d \times d}$
TransE [BUGD+13]	$-\lVert (\mathbf{y}_s + \mathbf{w}_r) - \mathbf{y}_o \rVert_2^2$	$\mathbf{w}_r, \mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d$
TransH [WZFC14]	$-\lVert (\mathbf{y}_s - \mathbf{w}_r^{\mathrm{T}} \mathbf{y}_s \mathbf{w}_r) + \mathbf{z}_r - (\mathbf{y}_o - \mathbf{w}_r^{\mathrm{T}} \mathbf{y}_o \mathbf{w}_r) \rVert_2^2$	$\mathbf{y}_s, \mathbf{y}_o, \mathbf{w}_r, \mathbf{z}_r \in \mathbb{R}^d$
TransR [LLS+15]	$-\lVert \mathbf{y}_s^{\mathrm{T}} \mathbf{W}_r + \mathbf{z}_r - \mathbf{y}_o^{\mathrm{T}} \mathbf{W}_r \rVert_2^2$	$\mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d, \quad \mathbf{z}_r \in \mathbb{R}^n, \quad \mathbf{W}_r \in \mathbb{R}^{d \times n}$
E-MRP [SCMN13]	$\mathbf{w}_r^{\mathrm{T}} tanh(\mathbf{Z}_r^{\mathrm{T}} \begin{bmatrix} \mathbf{y}_s \\ \mathbf{y}_o \end{bmatrix})$	$\mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d, \mathbf{w}_r \in \mathbb{R}^n, \mathbf{Z}_r \in \mathbb{R}^{2d \times n}$
ER-MRP [DGH+14]	$\mathbf{v}^{\mathrm{T}} tanh(\mathbf{Z}^{\mathrm{T}} \begin{bmatrix} \mathbf{y}_s \\ \mathbf{y}_o \\ \mathbf{w}_r \end{bmatrix})$	$\mathbf{w}_r, \mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d, \quad \mathbf{v} \in \mathbb{R}^n, \quad \mathbf{Z} \in \mathbb{R}^{3d \times n}$
NTN [SCMN13]	$\mathbf{v}_r^{\mathrm{T}} tanh(\mathbf{y}_s^{\mathrm{T}} \mathbf{W}_r^{[1\ldots n]} \mathbf{y}_o + \mathbf{Z}_r^{\mathrm{T}} \begin{bmatrix} \mathbf{y}_s \\ \mathbf{y}_o \end{bmatrix} + \mathbf{b}_r)$	$\mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d, \mathbf{v}_r, \mathbf{b}_r \in \mathbb{R}^n$ $\mathbf{Z}_r \in \mathbb{R}^{2d \times n}, \quad \mathbf{W}_r^{[1\ldots n]} \in \mathbb{R}^{d \times d \times n}$

$\Rightarrow$ 内積，距離，多層ニューラルネットなど様々

予測関数の学習

- 確率的生成モデル: 各 $x_{s,r,o}$ は Θ から互いに独立に生起

$$P(\mathbf{X}|\Theta) = \prod_{s=1}^{|V|}\prod_{o=1}^{|V|}\prod_{r=1}^{|\mathcal{R}|}\mathrm{Ber}(x_{s,r,o}|\sigma(f(s,r,o;\Theta)))$$

  * シグモイド関数: $\sigma(a) = \frac{1}{1+\exp(-a)}$
  * ベルヌーイ分布: $\mathrm{Ber}(x|p) = \begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$

- ロジスティック回帰学習: SGD などで最適化

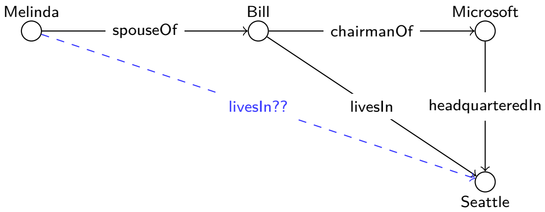
$$\begin{aligned}\hat{\Theta} &= \arg\min_{\Theta}\{-\log P(\mathbf{X}|\Theta)\} \\ &= \arg\min_{\Theta}\Big\{-\sum_{(s,r,o)} x_{s,r,o}\log\sigma(f(s,r,o;\Theta)) \\ &\quad +(1-x_{s,r,o})\log(1-\sigma(f(s,r,o;\Theta)))\Big\}\end{aligned}$$

負例サンプリング

- Closed world assumption
  - 存在しない Fact は負例
  - 知識グラフの不完全度が高い場合，危険な仮定
- Open world assumption
  - 存在しない Fact は未知
  - 知識グラフの不完全度が高い場合，有効な仮定
- Local-closed world assumption
  - $(s, r, o) \in E$ の場合，$\forall e \in V$ に対して，$(s, r, e) \notin E$ を負例

知識グラフ補完 (リンク予測)

- 知識グラフは不完全である



- $f(\mathrm{Melinda}, \mathrm{livesIn}, \mathrm{Seattle}; \Theta)$ によってリンク予測
  - 新しい Fact の発見
- 他にも
  - Entity/Relation の同義判定
  - Entity/Relation のクラスタリング
  - Entity/Relation のランキング

知識グラフ補完のベンチマーク

- WN18, FB15k データ[1]
  - WordNet と Freebase の一部を抽出 [BWU14, BUGD$^+$13]

- 知識グラフ補完タスク
  - テストデータの (s, r, o) に対して，$(?, r, o)$ 及び $(s, r, ?)$ を全ての Entity に対して計算
  - 平均逆順位 (Mean Reciprocal Rank)

$$\mathrm{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\mathrm{rank}_i}$$

where Q はクエリの集合，rank_i は正解の順位

または，Top@N ($\mathrm{rank}_i \leq N$ になる割合) で評価[2]

[1] https://everest.hds.utc.fr/doku.php?id=en:transe

[2] Filtered 設定: $(e, r, o) \in E$ となる e を除いて評価 (o に対しても同様)

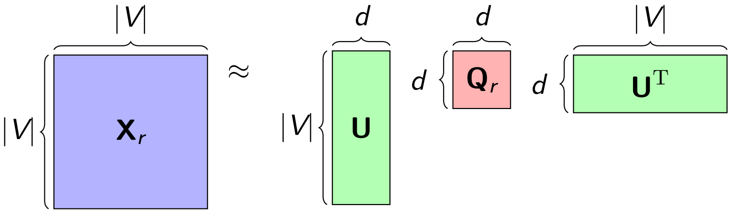
目次

- 知識グラフとは?
- 知識グラフに対する関係学習
  - 基本事項
  - 基本的な予測関数モデル

内積に基づく予測関数の基本形

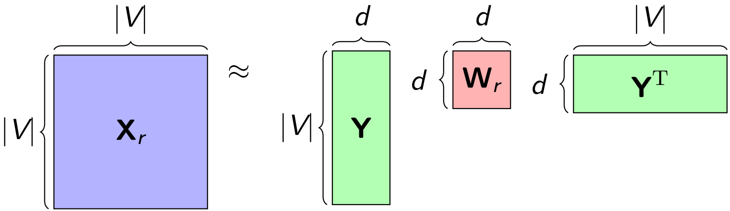
- $(s, r, o) \in E$ に対して
  - 予測関数: $f(s, r, o; \Theta) = \mathbf{u}_s^{\mathrm{T}} \mathbf{Q}_r \mathbf{u}_o$
  - パラメータ: $\Theta = \{\mathbf{Q}_r\}_{\forall r \in \mathcal{R}} \cup \{\mathbf{U}\}$

- Relation r の隣接行列 $\mathbf{X}_r$ を近似



RESCAL モデル [NTK11] (1/2)

- Entity の集合 V に対する行列: $\mathbf{Y} \in \mathbb{R}^{|V| \times d}$
- Relation r に対する実正方行列: $\mathbf{W}_r \in \mathbb{R}^{d \times d}$



$\Rightarrow$ 基本形そのもの

RESCAL モデル [NTK11] (2/2)

- テンソル積 $\otimes$: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$
  - 例: $\mathbf{x} = [x_0, x_1, x_2]$, $\mathbf{y} = [y_0, y_1, y_2]$

$$
\begin{aligned}
\mathbf{z} &= vec(\mathbf{x} \otimes \mathbf{y}) \\
&= [x_0 y_0, x_0 y_1, x_0 y_2, \\
&\quad x_1 y_0, x_1 y_1, x_1 y_2, \\
&\quad x_2 y_0, x_2 y_1, x_2 y_2]
\end{aligned}
$$

$$
\begin{array}{cccc}
 & x_0 & x_1 & x_2 \\
y_0 & z_0 & z_3 & z_6 \\
y_1 & z_1 & z_4 & z_7 \\
y_2 & z_2 & z_5 & z_8
\end{array}
$$

- 予測関数: 計算量 $O(d^2)$ $\Rightarrow$ 計算効率が悪い

$$f_{RESCAL}(s, r, o) = \mathbf{y}_s^{\mathrm{T}} \mathbf{W}_r \mathbf{y}_o = \mathbf{w}_r^{\mathrm{T}} vec(\mathbf{y}_s \otimes \mathbf{y}_o)$$

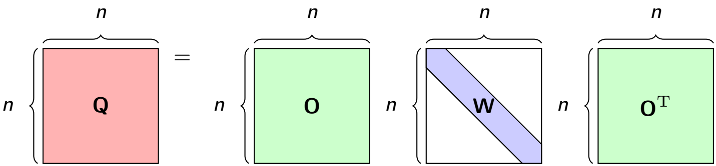
$\Rightarrow$ もっと計算効率の良い予測関数は設計できるか??

目次

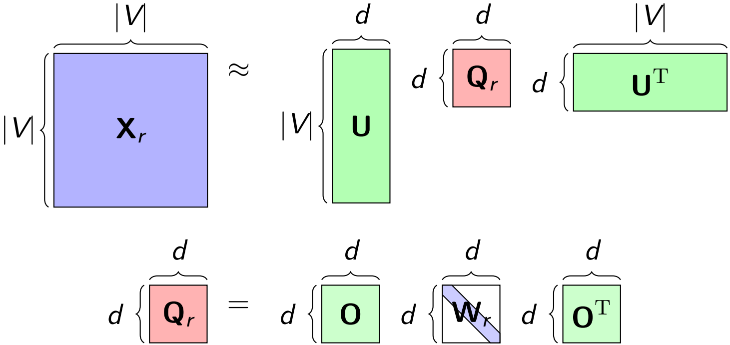
直交行列と対角行列を使った分解

定義 (実正方行列の直交対角化)

実正方行列 $\mathbf{Q} \in \mathbb{R}^{n \times n}$ が対角行列 $\mathbf{W} \in \mathbb{R}^{n \times n}$ と直交行列 $\mathbf{O} \in \mathbb{R}^{n \times n}$ $(\mathbf{O}\mathbf{O}^{\mathrm{T}} = \mathbf{O}^{\mathrm{T}}\mathbf{O} = \mathbf{I})$ を使って $\mathbf{Q} = \mathbf{O}\mathbf{W}\mathbf{O}^{\mathrm{T}}$ と書けるとき， $\mathbf{Q}$ は直交行列により対角化可能である



DistMult モデル [YYH+14]



- 予測関数: $O(d)$ で計算可能

$$f_{DistMult}(s, r, o) = \mathbf{u}_s^{\mathrm{T}} \mathbf{O} \mathbf{W}_r \mathbf{O}^{\mathrm{T}} \mathbf{u}_o = \mathbf{y}_s^{\mathrm{T}} \mathbf{W}_r \mathbf{y}_o = \langle \mathbf{w}_r, \mathbf{y}_s, \mathbf{y}_o \rangle$$

$*$ $\mathbf{w}_r = diag(\mathbf{W}_r)$ で $\mathbf{W}_r$ の対角成分をベクトル化

DistMult モデルの問題点

定理 (スペクトル定理 ([柴田 13] を参考))

実正方行列 $\mathbf{Q}$ が直交行列により対角化可能である必要十分条件は $\mathbf{Q}$ が実対称行列の場合である $(\mathbf{Q} = \mathbf{Q}^{\mathrm{T}}, q_{ij} = q_{ji})$

- $f_{DistMult}(s, r, o) = f_{DistMult}(o, r, s)$
  - 対称な Relation: *is_identical*, *is_married_to*
- 知識グラフには非対称な Relation が存在
  - 1 対多の Relation: *is_parent_of*
  - 多対 1 の Relation: *is_child_of*

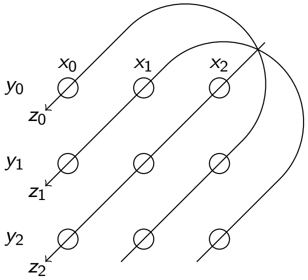
$\Rightarrow$ 非対称 + 計算効率の良い予測関数は設計できるか??

目次

- 知識グラフとは?
- 知識グラフに対する関係学習
  - 基本事項
  - 基本的な予測関数モデル
  - 実対称行列に限定した予測関数
  - 実正規行列に限定した予測関数

巡回畳み込みと相互相関 (1/2)

- 巡回畳み込み $* : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$



$$\mathbf{z} = \mathbf{x} * \mathbf{y}$$

$$z_0 = x_0 y_0 + x_1 y_2 + x_2 y_1$$

$$z_1 = x_0 y_1 + x_1 y_0 + x_2 y_2$$

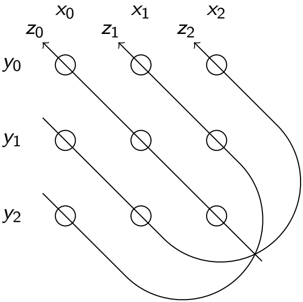
$$z_2 = x_0 y_2 + x_1 y_1 + x_2 y_0$$

- 性質
  - 畳み込みは可換 $\mathbf{x} * \mathbf{y} = \mathbf{y} * \mathbf{x}$
  - 高速フーリエ変換により $O(d \log d)$

$$\mathbf{x} * \mathbf{y} = \mathfrak{F}^{-1}(\mathfrak{F}(\mathbf{x}) \odot \mathfrak{F}(\mathbf{y}))$$

巡回畳み込みと相互相関 (2/2)

- 巡回相互相関 $\star : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$



$$\mathbf{z} = \mathbf{x} \star \mathbf{y}$$

$$z_0 = x_0 y_0 + x_1 y_1 + x_2 y_2$$

$$z_1 = x_0 y_2 + x_1 y_0 + x_2 y_1$$

$$z_2 = x_0 y_1 + x_1 y_2 + x_2 y_0$$

- 性質
  - 相互相関は非可換 $\mathbf{x} \star \mathbf{y} \neq \mathbf{y} \star \mathbf{x}$
  - 高速フーリエ変換により $O(d \log d)$

$$\mathbf{x} \star \mathbf{y} = \mathfrak{F}^{-1}(\overline{\mathfrak{F}(\mathbf{x})} \odot \mathfrak{F}(\mathbf{y}))$$

HolE モデル [NRP16]

- 相互相関に基づく予測関数
  - $s, o \in V$ を d 次元ベクトル $\mathbf{y}_s, \mathbf{y}_o \in \mathbb{R}^d$
  - $r \in \mathcal{R}$ を d 次元ベクトル $\mathbf{w}_r \in \mathbb{R}^d$

$$f_{HolE}(s, r, o) = \mathbf{w}_r^{\mathrm{T}}(\mathbf{y}_s \star \mathbf{y}_o)$$

- HolE モデルの利点
  - 計算量 $O(d \log d)$, Relation r のベクトル次元数 $O(d)$
  - 相互相関は非可換な演算
    - $f_{HolE}(s, r, o) \neq f_{HolE}(o, r, s)$
    - $\mathbf{X}_r$ を非対称な行列として表現可能

正規行列について

定義 (正規行列)

複素正方行列 $\mathbf{Q} \in \mathbb{C}^{n \times n}$ が $\mathbf{Q}^*\mathbf{Q} = \mathbf{QQ}^*$ のとき，$\mathbf{Q}$ を正規行列と呼ぶ．ただし，$\mathbf{Q}^* = \overline{\mathbf{Q}}^{\mathrm{T}}$ は共役転置

- 正規行列の例:
  - 実対称行列: $\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}$
  - 歪対称行列 (Skew-symmetric): $\mathbf{Q}^{\mathrm{T}} = -\mathbf{Q}$ ($q_{ij} = -q_{ji}$)
  - エルミート行列 (Hermitian): $\mathbf{Q}^* = \mathbf{Q}$
  - 歪エルミート行列 (Skew-Hermitian): $\mathbf{Q}^* = -\mathbf{Q}$

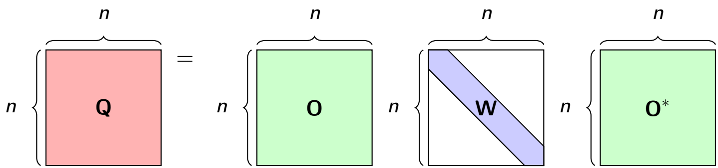
$\Rightarrow$ 歪対称 (反対称) 性が扱える

ユニタリ行列と対角行列を使った分解

Innovative R&D by NTT

定義 (複素正方行列のユニタリ対角化)

複素正方行列 $\mathbf{Q} \in \mathbb{C}^{n \times n}$ が対角行列 $\mathbf{W} \in \mathbb{C}^{n \times n}$ とユニタリ行列 $\mathbf{O} \in \mathbb{C}^{n \times n}$ $(\mathbf{O}\mathbf{O}^* = \mathbf{O}^*\mathbf{O} = \mathbf{I})$ を使って $\mathbf{Q} = \mathbf{O}\mathbf{W}\mathbf{O}^*$ と書けるとき, $\mathbf{O}$ はユニタリ行列により対角化可能である



定理 (スペクトル定理 ([藤原 96] を参考))

複素正方行列 $\mathbf{Q}$ がユニタリ行列により対角化可能である必要十分条件は $\mathbf{Q}$ が正規行列の場合である

ComplEx モデル [TWR+16] (1/3)

系 (実正規行列のユニタリ対角化 [TDW+17])

実正規行列 $\mathbf{Q} \in \mathbb{R}^{d \times d}$ に対して，$\mathbf{Q} = Re(\mathbf{OWO}^*)$ となるユニタリ行列 $\mathbf{O} \in \mathbb{C}^{d \times d}$ と対角行列 $\mathbf{W} \in \mathbb{C}^{d \times d}$ が存在する

- 基本形への統合を考える: 複素数のままでは扱いにくい

$$
\begin{aligned}
\mathbf{O} &= \begin{pmatrix} a_{1,1} + b_{1,1} i & \cdots & a_{1,d} + b_{1,d} i \\ \vdots & \ddots & \vdots \\ a_{d,1} + b_{d,1} i & \cdots & a_{d,d} + b_{d,d} i \end{pmatrix} \in \mathbb{C}^{d \times d} \\
\mathbf{O}' &= \begin{pmatrix} b_{1,1} & a_{1,1} & \cdots & b_{1,d} & a_{1,d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{d,1} & a_{d,1} & \cdots & b_{d,d} & a_{d,d} \end{pmatrix} \in \mathbb{R}^{d \times 2d}
\end{aligned}
$$

ComplEx モデル [TWR+16] (2/3)

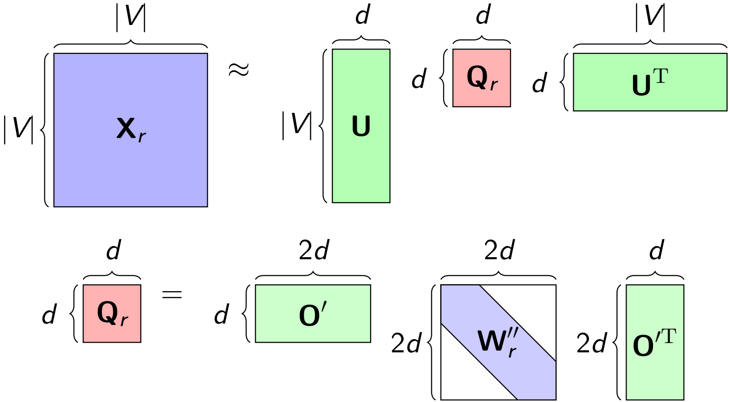
$$\mathbf{W} = \begin{pmatrix} a_{1,1} + b_{1,1} i & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{n,n} + b_{n,n} i \end{pmatrix} \in \mathbb{C}^{d \times d}$$

$$\mathbf{W}'' = \begin{pmatrix} a_{1,1} & -b_{1,1} & \cdots & 0 & 0 \\ b_{1,1} & a_{1,1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{d,d} & -b_{d,d} \\ 0 & 0 & \cdots & b_{d,d} & a_{d,d} \end{pmatrix} \in \mathbb{R}^{2d \times 2d}$$

- 一旦，実行列で表現

$$Re(\mathbf{OWO}^*) = \mathbf{O}'\mathbf{W}''\mathbf{O}'^{\mathrm{T}}$$

ComplEx モデル [TWR$^+$16] (3/3)



- 予測関数: $O(d)$ で計算可能

$$
\begin{aligned}
f_{Comp.}(s, r, o) &= \mathbf{u}_s^{\mathrm{T}} \mathbf{O}' \mathbf{W}''_r \mathbf{O}'^{\mathrm{T}} \mathbf{u}_o = \mathbf{y}'^{\mathrm{T}}_s \mathbf{W}''_r \mathbf{y}'_o \\
&= Re(\mathbf{y}_s^{\mathrm{T}} \mathbf{W}_r \overline{\mathbf{y}_o}) = Re(\langle \mathbf{w}_r, \mathbf{y}_s, \overline{\mathbf{y}_o} \rangle)
\end{aligned}
$$

HolE と ComplEx の等価性 [HS17]

- HolE は周波数領域で計算可能
  - $\omega_r = \mathfrak{F}(\mathbf{w}_r),\quad \psi_s = \mathfrak{F}(\mathbf{y}_s),\quad \psi_o = \mathfrak{F}(\mathbf{y}_o)$
  - 内積/相互相関の対応

$$\begin{aligned} f_{HolE}(s,r,o) &= \mathbf{w}_r^{\mathrm{T}}(\mathbf{y}_s \star \mathbf{y}_o) \\ &= \frac{1}{d}\omega_r^{\mathrm{T}}(\overline{\psi_s} \odot \psi_o) \end{aligned}$$

演算	時間		周波数
スカラ積	$\alpha\mathbf{x}$	$\longleftrightarrow$	$\alpha\,\mathfrak{F}(\mathbf{x})$
和	$\mathbf{x}+\mathbf{y}$	$\longleftrightarrow$	$\mathfrak{F}(\mathbf{x})+\mathfrak{F}(\mathbf{y})$
畳み込み	$\mathbf{x}*\mathbf{y}$	$\longleftrightarrow$	$\mathfrak{F}(\mathbf{x})\odot\mathfrak{F}(\mathbf{y})$
相互相関	$\mathbf{x}\star\mathbf{y}$	$\longleftrightarrow$	$\overline{\mathfrak{F}(\mathbf{x})}\odot\mathfrak{F}(\mathbf{y})$
内積	$\mathbf{x}\cdot\mathbf{y}$	$=$	$\frac{1}{d}\,\mathfrak{F}(\mathbf{x})\cdot\mathfrak{F}(\mathbf{y})$

- 共役対称性の議論を経て

$$f_{HolE}(s,r,o) = \frac{2}{d} f_{Comp.}(s,r,o)$$

ANALOGY モデル [LWY17] (1/2)

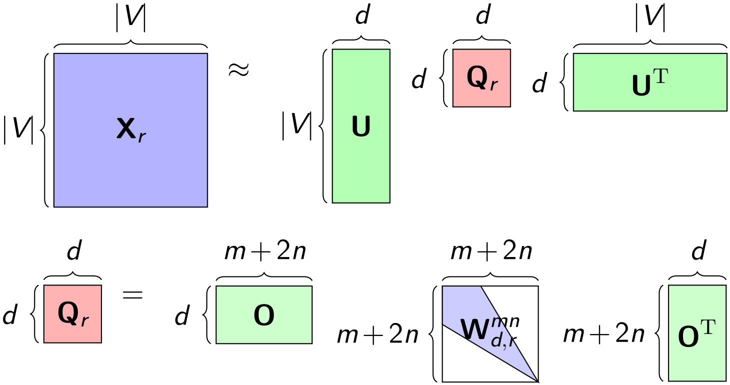
小定理 (実正規行列の直交ブロック対角化 [LWY17])

実正規行列 $\mathbf{Q}$ は直交行列 $\mathbf{O}$ とブロック対角行列 $\mathbf{W}$ を使って $\mathbf{O}\mathbf{W}\mathbf{O}^{\mathrm{T}}$ と書ける．ただし，$\mathbf{W}$ の各ブロックは *(1)* 実数，または，*(2)* $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ をとる

- 対角行列 $\mathbf{W}_d^{mn} \in \mathbb{R}^{(m+2n)\times(m+2n)}$: $d = m + n$
  - m 個の実数と n 個の 2×2 ブロック

$$
\mathbf{W}_d^{mn} = \left(\begin{array}{ccccc|ccc}
a_{1,1} & -b_{1,1} & \cdots & 0 & 0 & 0 & \cdots & 0 \\
b_{1,1} & a_{1,1} & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & a_{n,n} & -b_{n,n} & 0 & \cdots & 0 \\
0 & 0 & \cdots & b_{n,n} & a_{n,n} & 0 & \cdots & 0 \\
\hline
0 & 0 & \cdots & 0 & 0 & c_1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & c_m
\end{array}\right)
$$

ANALOGY モデル [LWY17] (2/2)



- DistMult と ComplEx の線形結合

$$f_{ANALOGY}(s,r,o) = \underbrace{f_{DistMult}(s,r,o)}_{m} + \underbrace{f_{Comp.}(s,r,o)}_{2n}$$

目次

まとめ

モデル	計算時間	非対称性	次元数
RESCAL	$O(d^2)$	✓	d^2
DistMult	$O(d)$	✗	d
HolE	$O(d \log d)$	✓	d
ComplEx	$O(d)$	✓	$2d$
ANALOGY	$O(d)$	✓	$m+2n$ $(d=m+n)$

- 距離や多層ニューラルネットに基づく予測関数
  - 日本語だと資料 [海野 15] がわかりやすい

メモ: 直交行列の同一性について

定理 (文献 [LWY17] より)

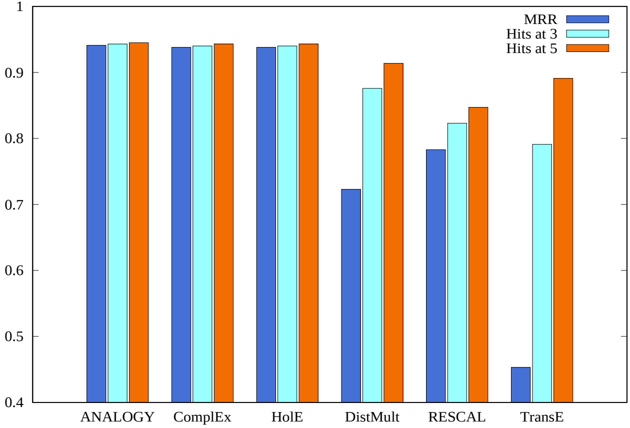
実正規行列の集合 $Q_1, Q_2, \ldots$ が可換族

$$Q_i Q_j = Q_j Q_i \ \forall i, j$$

を形成するとき，共通の直交基底を共有しなければならない

- $\forall r, r' \in \mathcal{R}$ に対して，$Q_r, Q_{r'}$ が可換族となる制約を考えているため，対角化の際に共通の直交 (ユニタリ) 行列 O が使える

WN18 上で Filtered 設定による再実験



⇒ データの量や性質，ハイパーパラメータの設定，学習/最適化法との兼ね合いでもあることに注意 [KBK17]

目次

- 知識グラフとは?
- 知識グラフに対する関係学習
  - 基本事項
  - 基本的な関係予測モデル
  - 実対称行列に限定した予測関数
  - 実正規行列に限定した予測関数
  - まとめ
- 応用

推論規則の抽出 [YYH+14]

- 推論規則

$$bornInCity(a,b) \wedge cityOfCountry(b,c) \Rightarrow nationality(a,c)$$

  - 新しい Fact の生成
  - 知識グラフのコンパクトな格納
  - 高度な推論システム

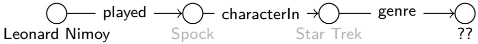
- 長さ k の規則を抽出 (実験では $k=2,3$ に限定)
  - Relation r に対し，$\mathcal{X}_r$ と $\mathcal{Y}_r$ を Subject，Object 候補の集合
  - 開始 Relation 集合 $\mathcal{S} = \{s : \mathcal{X}_s \cap \mathcal{X}_r \neq \emptyset\}$
  - 終端 Relation 集合 $\mathcal{O} = \{t : \mathcal{Y}_t \cap \mathcal{Y}_r \neq \emptyset\}$
  - $\mathcal{S}$ と $\mathcal{O}$ から可能な Relation path $r_1, \ldots, r_k$ を列挙
  - $\mathbf{y}_a^{\mathrm{T}}\mathbf{W}_{r_1} \approx \mathbf{y}_b$ 及び $\mathbf{y}_b^{\mathrm{T}}\mathbf{W}_{r_2} \approx \mathbf{y}_c$ から $\mathbf{y}_a^{\mathrm{T}}\mathbf{W}_{r_1}\mathbf{W}_{r_2} \approx \mathbf{y}_c$

$$dist(\mathbf{W}_r, \mathbf{W}_{r_1} \ldots \mathbf{W}_{r_k})$$

  で閾値を越える推論規則を抽出

パスクエリ応答処理 [GML15] (1/2)

- パスクエリ: $[\![q]\!] =: s/r_1/r_2/\ldots/r_k$



- 解の集合: 離散的な定義

  初期化 $[\![s]\!] =: \{s\}$

  再帰関数 $[\![q/r]\!] =: \{o : \exists s \in [\![q]\!], (s, r, o) \in E\}$

- パスクエリ応答モデル: ベクトルによる定義

  ベクトル初期化 $[\![s]\!]_V =: \mathbf{e}_s$

  ベクトル遷移関数 $[\![q/r]\!]_V =: \mathcal{T}_r([\![q]\!]_V)$

  ベクトル評価関数 $f(q, o) = \mathcal{M}([\![q]\!]_V, [\![o]\!]_V)$

$\Rightarrow \mathcal{T}$ と $\mathcal{M}$ はどんな関数?

パスクエリ応答処理 [GML15] (2/2)

- DistMult or RESCAL モデル

  遷移関数 $\mathcal{T}_r(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{W}_r$

  評価関数 $\mathcal{M}(\mathbf{x}, \mathbf{y}) = \mathbf{x}^{\mathrm{T}}\mathbf{y}$

  パスクエリ予測関数 $f(s/r_1/r_2/\ldots/r_k, o) = \mathbf{y}_s^{\mathrm{T}}\mathbf{W}_{r_1}\ldots\mathbf{W}_{r_n}\mathbf{y}_o$

- TransE モデル

  遷移関数 $\mathcal{T}_r(\mathbf{x}) = \mathbf{x} + \mathbf{w}_r$

  評価関数 $\mathcal{M}(\mathbf{x}, \mathbf{y}) = -||\mathbf{x} - \mathbf{y}||$

  パスクエリ予測関数

  $f(s/r_1/r_2/\ldots/r_k, o) = -||\mathbf{y}_s + \mathbf{w}_{r_1} + \cdots + \mathbf{w}_{r_k} - \mathbf{y}_o||$

オープンドメイン質問応答 [BWU14]

- 質問パタンと Fact から質問文を生成

  q: what are afraid of cats?
  t: $(rats, afraidOf, cats)$

KB Fact	質問パタン	KB Fact	質問パタン
$(?, r, o)$	who r o ?	$(?, r, o)$	what is o's r ?
$(?, r, o)$	what r o ?	$(s, r, ?)$	who is r by s ?
$(s, r, ?)$	who does s r ?	$(s, r\text{-}in, ?)$	when did s r ?
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- 質問文と Fact を近付けるように学習

$$f(q, t) = (\mathbf{W}^{\mathrm{T}}\mathbf{q})^{\mathrm{T}}(\mathbf{V}^{\mathrm{T}}\mathbf{t})$$

$\mathbf{q} \in \{0,1\}^{n_1}$, $\mathbf{t} \in \{0,1\}^{n_2}$: q と t の 2 値素性ベクトル
$\mathbf{W} \in \mathbb{R}^{n_1 \times d}$, $\mathbf{V} \in \mathbb{R}^{n_2 \times d}$: 埋め込み行列

目次

- 知識グラフとは?
- 知識グラフに対する関係学習
  - 基本事項
  - 基本的な関係予測モデル
  - 実対称行列に限定した予測関数
  - 実正規行列に限定した予測関数
  - まとめ
- 応用
- 今後の課題

今後の課題

- ベンチマークの見直しや人工データの整備
  - モデルの表現力，パラメータ数，最適化方法などの要因を切り分けにくい
- 汎用的なクエリ応答処理系との統合
  - RESCAL + 関係データベース [KNT14]
  - ComplEx + 一階述語論理 (Deep Prolog) [Roc17]

参考文献 I

[BEP+08] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD*, pp. 1247–1250. AcM, 2008.

[Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[BUGD+13] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pp. 2787–2795, 2013.

参考文献 II

[BWCB11] Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. Learning structured embeddings of knowledge bases. In *Proceedings of the 25th AAAI*, 2011.

[BWU14] Antoine Bordes, Jason Weston, and Nicolas Usunier. Open question answering with weakly supervised embedding models. In *Proceedings of ECML-PKDD'14*, pp. 165–180, 2014.

[DGH+14] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD*, pp. 601–610, 2014.

参考文献 III

[Fel98] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA, 1998.

[GML15] Kelvin Guu, John Miller, and Percy Liang. Traversing knowledge graphs in vector space. In *Proceedings of the 2015 EMNLP*, pp. 318–327, 2015.

[HS17] Katsuhiko Hayashi and Masashi Shimbo. On the equivalence of holographic and complex embeddings for link prediction. *arXiv preprint arXiv:1702.05563*, 2017.

[KBK17] Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst. Knowledge base completion: Baselines strike back. *arXiv preprint arXiv:1705.10744*, 2017.

参考文献 IV

[KNT14] Denis Krompaß, Maximilian Nickel, and Volker Tresp. Querying factorized probabilistic triple databases. In *International Semantic Web Conference*, pp. 114–129. Springer, 2014.

[LLS+15] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the 29th AAAI*, pp. 2181–2187, 2015.

[LSS+11] Hady W. Lauw, Ralf Schenkel, Fabian Suchanek, Martin Theobald, and Gerhard Weikum. Semantic knowledge bases from web sources. `http://ijcai-11.iiia.csic.es/files/proceedings/IJCAI-T5.pdf`, 2011.

参考文献 V

[LWY17] Hanxiao Liu, Yuexin Wu, and Yiming Yang. Analogical inference for multi-relational embeddings. *arXiv preprint arXiv:1705.02426*, 2017.

[NMTG16] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, Vol. 104, No. 1, pp. 11–33, 2016.

[NP10] Roberto Navigli and Simone Paolo Ponzetto. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th ACL*, pp. 216–225. Association for Computational Linguistics, 2010.

参考文献 VI

[NRP16] Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. Holographic embeddings of knowledge graphs. In *Proceedings of the 30th AAAI*, pp. 1955–1961, 2016.

[NT13] Maximilian Nickel and Volker Tresp. Logistic tensor factorization for multi-relational data. *arXiv preprint arXiv:1306.2084*, 2013.

[NTK11] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th ICML*, pp. 809–816, 2011.

参考文献 VII

[PTVS+16] Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. From freebase to wikidata: The great migration. In *Proceedings of the 25th WWW*, pp. 1419–1428, 2016.

[Roc17] Tim Rocktäschel. Deep prolog: End-to-end differentiable proving in knowledge bases. 2017.

[SCMN13] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*, pp. 926–934, 2013.

参考文献 VIII

[TDW+17] Théo Trouillon, Christopher R Dance, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Knowledge graph completion via complex tensor factorization. *arXiv preprint arXiv:1702.06879*, 2017.

[TWR+16] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *Proceedings of the 33rd ICML*, pp. 2071–2080, 2016.

[WZFC14] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the 28th AAAI*, 2014.

参考文献 IX

[YYH+14] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *CoRR*, Vol. abs/1412.6575, , 2014.

[海野 15] 海野裕也. 「知識」の deep learning. `https://www.slideshare.net/unnonouno/deep-learning-48974928`, 2015.

[鹿島 09] 鹿島久嗣. 関係データの機械学習. `www.geocities.co.jp/kashi_pong/relationalLearningTensors.pdf`, 2009.

[柴田 13] 柴田正和. 科学者・技術者のための基礎線形代数と固有値問題; pp.163–167. 森北出版, 2013.

参考文献 X

[石黒 16] 石黒勝彦, 林浩平. 関係データ学習. MLP 機械学習プロフェッショナルシリーズ. 講談社, 2016.

[藤原 96] 藤原毅夫. 線形代数 (理工系の基礎数学 2). 岩波書店, 1996.