

フーリエ領域上でのホログラフィック埋め込み

林克彦¹ 新保仁² 永田昌明¹

¹ 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

² 奈良先端科学技術大学院大学 情報科学研究科

¹{hayashi.katsuhiko, nagata.masaaki}@lab.ntt.co.jp

²shimbo@is.naist.jp

1 はじめに

Wordnet や Freebase のような知識データを使った研究が注目を集めて久しい。例えば、Freebase には (Barack Obama, presidentOf, USA) のような entity 間の relation を表した事実が集積されており、質問応答などのタスクに応用されてきた。一方で、それらの知識データはまだ構築途中であるため、必要な事実の欠落がしばしば指摘されている。そのため、データには存在しない事実を自動補完しながら質問応答などを行うことが実用上の課題となっている。

この課題を解決するため、各 entity と relation をユークリッド空間上のベクトルで表し、ベクトル間の演算を使って、ある事実が存在するかどうかの信頼度を自動推定するモデルが盛んに研究されている。これらのモデルではベクトル間の類似度を使って信頼度を計算するが、類似度計算には内積を使うのが一般的である。2つのベクトル \mathbf{x} , \mathbf{y} に対して、内積 $\mathbf{x} \cdot \mathbf{y}$ は次元ごとの要素の積を合計するが、知識データのように複雑な関係をモデル化する場合、異なる次元の要素間のインタラクションを考慮する構成的なベクトル演算が信頼度推定の精度向上には重要となる。

文献 [4] ではテンソル積 $\mathbf{x} \otimes \mathbf{y}$ を使った知識グラフデータの信頼度推定モデルを提案している (RESICAL モデル)。しかし、テンソル積には $O(n^2)$ の計算コストがかかるため、多くの研究者がより効率的な構成的ベクトル演算を検討してきた。その中でも文献 [3] で提案された相互相関演算に基づく手法 (HoIE モデル) は計算量を $O(n \log n)$ に削減でき、また、高い信頼度推定精度を達成することができるため、注目を集めている。本稿ではこのモデルの概略を説明すると共に、フーリエ領域上でのベクトル演算を使うことで、HoIE モデルの計算量を $O(n)$ に削減する手法を新たに提案する。また、提案手法は HoIE モデルと完全に等価であるため、モデル精度が落ちることはない。

本稿では i は虚数単位、 \mathbb{R} は実数の集合、 \mathbb{C} は複素数の集合とする。ある複素数 $z \in \mathbb{C}$ に対して、その共役複素数を \bar{z} と書く。

2 巡回畳み込みと相互相関

あるベクトル \mathbf{v} に対し、その j 番目の要素を $[\mathbf{v}]_j$ と表す。 $\mathbf{x} = [x_0, \dots, x_{n-1}] \in \mathbb{R}^n$ と $\mathbf{y} = [y_0, \dots, y_{n-1}] \in \mathbb{R}^n$ を n 次元ベクトルとし、 j は 0 から始まるとする。すると、 \mathbf{x} と \mathbf{y} の巡回畳み込み (circular convolution) $\mathbf{x} \circledast \mathbf{y}$ は次のように定義される。

$$[\mathbf{x} \circledast \mathbf{y}]_j = \sum_{k=0}^{n-1} x_{[(j-k) \bmod n]} y_k \quad (j = 0, \dots, n-1). \quad (1)$$

同様に、相互相関 (circular correlation) $\mathbf{x} \odot \mathbf{y}$ は

$$[\mathbf{x} \odot \mathbf{y}]_j = \sum_{k=0}^{n-1} x_{[(k-j) \bmod n]} y_k \quad (j = 0, \dots, n-1) \quad (2)$$

と定義される。巡回畳み込みは可換な操作である一方、相互相関は非可換な操作である。すなわち、 $\mathbf{x} \circledast \mathbf{y} = \mathbf{y} \circledast \mathbf{x}$ であるが、 $\mathbf{x} \odot \mathbf{y} \neq \mathbf{y} \odot \mathbf{x}$ となる。式 (1) と (2) から $\mathbf{x} \odot \mathbf{y} = \text{flip}(\mathbf{x}) \circledast \mathbf{y}$ は容易に示される。ここで $\text{flip}(\mathbf{x})$ は \mathbf{x} の要素を反転したベクトルを表す。

式 (1) と (2) から明らかなように、巡回畳み込みと相互相関をナイーブに実装すると $O(n^2)$ の計算コストがかかる。しかしながら、時間領域における巡回畳み込みとフーリエ領域 (または、周波数領域) における要素積の同一性を利用することで、これらの計算を高速化できる。巡回畳み込みの場合、まず \mathbf{x} と \mathbf{y} の離散フーリエ変換 (DFT) を計算し、それらの結果に対して要素積を計算する。そして、その結果を逆離散フーリエ変換 (iDFT) することで巡回畳み込みを

$$\mathbf{x} \circledast \mathbf{y} = \text{DFT}^{-1}(\text{DFT}(\mathbf{x}) \odot \text{DFT}(\mathbf{y})) \quad (3)$$

と計算できる。ここで DFT と DFT^{-1} はそれぞれ DFT と iDFT を表す。DFT と iDFT は $O(n \log n)$ で計算で

きるため、巡回畳み込みの計算量は $O(n \log n)$ となる。相互相関についても同様に

$$\mathbf{x} \circledast \mathbf{y} = \text{DFT}^{-1}(\overline{\text{DFT}(\mathbf{x})} \circledast \text{DFT}(\mathbf{y})) \quad (4)$$

として計算することができる。

3 ホログラフィック埋め込み

3.1 知識グラフの埋め込み

ℓ 個の entity から成る集合 \mathcal{E} と 2 項 relation の集合 \mathcal{R} が与えられたとき、ある relation $r \in \mathcal{R}$, subject $j \in \mathcal{E}$, object $k \in \mathcal{E}$ に対して、entity j, k が relation r という関係にあるかどうかを予測する関数 $f: (r, s, o) \rightarrow \{-1, 1\}$ を設計したい。このような予測関数を設計するため、ここでは訓練データ集合 $\mathcal{D} \subset \mathcal{R} \times \mathcal{E} \times \mathcal{E}$ を持つと考える。このデータは 3 つ組 $(r, s, o) \in \mathcal{D}$ から成り、 $y = r(s, o)$ で (r, s, o) が事実として真か偽 (1 または -1) を表す。

このデータ \mathcal{D} を使って、未知の 3 つ組 $(r, s, o) \notin \mathcal{D}$ が事実として成り立つかどうかの真偽を高精度に予測する関数 f を設計する ($f(r, s, o) = r(s, o)$ となるように設計したい)。さらに一般には、後述する文献 [3] のホログラフィック埋め込みのように予測関数 f として真理値ではなく、 $r(s, o)$ の真偽を確率値として判定するようにモデル化する場合がある ($f: (r, s, o) \rightarrow [0, 1]$)。

3.2 Nickel16 のモデル (HoLE)

文献 [3] はホログラフィック縮退表現 (HRR: *holographic reduced representation*) [5] を知識グラフの埋め込みに応用した。この手法では entity に対するベクトル $\mathbf{e}_j, j = 1, \dots, \ell$, と relation に対するベクトル $\mathbf{w}_r, r = 1, \dots, |\mathcal{R}|$ を使って、 s と o 番目の entity が r 番目の relation で関連付けられる信頼度を

$$f(r, s, o) = \sigma(\mathbf{w}_r \cdot (\mathbf{e}_s \circledast \mathbf{e}_o)) \quad (5)$$

としてモデル化する。ここで $\sigma(\cdot)$ はシグモイド関数とする。節 2 で述べたように式 (5) の相互相関は式 (4) を用いて $O(n \log n)$ で実行できる。ただし、 n はベクトル $\mathbf{e}_j, \mathbf{w}_r$ の長さである。

3.3 フーリエ領域上での埋め込み

文献 [3] の手法は時間領域での相互相関をフーリエ領域での要素積に変換することで計算コストを下げているが、その結果を iDFT で再び元の時間領域に変換

表 1: 時間領域とフーリエ領域での対応表。

時間領域	\longleftrightarrow	フーリエ領域
$\mathbf{x} = \text{DFT}^{-1}(\boldsymbol{\xi})$	\longleftrightarrow	$\boldsymbol{\xi} = \text{DFT}(\mathbf{x})$
$\mathbf{y} = \text{DFT}^{-1}(\boldsymbol{\eta})$	\longleftrightarrow	$\boldsymbol{\eta} = \text{DFT}(\mathbf{y})$
$\alpha \mathbf{x}$	\longleftrightarrow	$\alpha \boldsymbol{\xi}$
$\mathbf{x} + \mathbf{y}$	\longleftrightarrow	$\boldsymbol{\xi} + \boldsymbol{\eta}$
$\mathbf{x} \cdot \mathbf{y}$	\longleftrightarrow	$\frac{1}{n} \boldsymbol{\xi} \cdot \boldsymbol{\eta}$
flip(\mathbf{x})	\longleftrightarrow	$\bar{\boldsymbol{\xi}}$
$\mathbf{x} \circledast \mathbf{y}$	\longleftrightarrow	$\boldsymbol{\xi} \circledast \boldsymbol{\eta}$
$\mathbf{x} \circledast \mathbf{y}$	\longleftrightarrow	$\bar{\boldsymbol{\xi}} \circledast \boldsymbol{\eta}$

して内積計算を行っている。 $O(n \log n)$ は、小規模な知識ベースに対しては十分高速だが、より大規模な応用を考えると、さらなる高速化が望まれる。

本稿ではこのモデルにかかる全ての計算をフーリエ領域だけで行い、DFT/iDFT の計算コストを取り除くことで HoLE モデルの計算を $O(n)$ で行う。HoLE モデルの計算では、相互相関の他にスカラ積・加算 (確率的勾配硬化法を用いた学習に使用)、および内積計算 (ベクトル類似度計算) が用いられる。このため、同モデルをフーリエ領域で計算するためには、それぞれに該当する演算が存在し、かつ効率的に行える必要がある。

幸運な事に、フーリエ領域上での上記演算は元の時間領域上での各演算と対応がとれる。2 つの n 次元ベクトル $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ とそれらの DFT 結果をそれぞれ $\boldsymbol{\xi} = \text{DFT}(\mathbf{x}) \in \mathbb{C}^n$ と $\boldsymbol{\eta} = \text{DFT}(\mathbf{y}) \in \mathbb{C}^n$ と定義すると、表 1 のような対応が存在し、必要な演算はすべてフーリエ領域側で等価計算可能な事がわかる。

このうち、特に内積計算についてはパーセバルの定理から両領域上で式 (6) のように比例定数を除いて等しい。すなわち、ベクトル $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ とそれらを DFT した結果できるベクトル $\boldsymbol{\xi} = \text{DFT}(\mathbf{x}) \in \mathbb{C}^n$ と $\boldsymbol{\eta} = \text{DFT}(\mathbf{y}) \in \mathbb{C}^n$ の間に以下の等式が成り立つ。

$$\mathbf{x} \cdot \mathbf{y} = \frac{1}{n} \boldsymbol{\xi} \cdot \boldsymbol{\eta}. \quad (6)$$

これは $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ に対して、それらを DFT した結果が複素ベクトルである一方、その内積結果の虚数部は 0 になることを意味する。

3.3.1 空間効率及び iDFT の実数性に関する考察

フーリエ領域は n 次元の複素ベクトル空間となり、各要素は実数と虚部を持つため、ベクトルの空間使用量は元の時間領域の 2 倍になるように思われる。しかし、これは必ずしも正しくない。もし時間領域のベクトル

ルが実数ベクトルであれば、その DFT 結果は共役対称性を持つ。すなわち、 $\mathbf{x} \in \mathbb{R}^n$ に対して、その DFT 結果 $\boldsymbol{\xi} = \text{DFT}(\mathbf{x}) = [\xi_0, \dots, \xi_{n-1}] \in \mathbb{C}^n$ は全ての $j = 1, \dots, n-1$ に対して、 $\xi_j = \overline{\xi_{n-j}}$ となる。さらに、 ξ_0 と $\xi_{n/2}$ は実数となる¹。それゆえ、 $\boldsymbol{\xi}$ を表すには、最初の $\lfloor n/2 \rfloor + 1$ 要素、 $j = 0, \dots, \lfloor n/2 \rfloor$ に対する ξ_j のみを格納すれば良く、 ξ_0 は (もし n が偶数なら、 $\xi_{n/2}$ も) 実数となる。そして、残りの $j = \lfloor n/2 \rfloor + 1, \dots, n-1$ に対する ξ_j は不要となる。複素数は実部と虚部に実数を持つので、最終的に保持しなければならない浮動小数点数は n 個で元と同じに抑えられる。

上記の共役対称性が保たれれば、逆も成立する。もし、 $\boldsymbol{\xi} = [\xi_0, \dots, \xi_{n-1}] \in \mathbb{C}^n$ が全ての $j = 1, \dots, n-1$ に対して、 $\xi_j = \overline{\xi_{n-j}}$ が成り立ち、 $\xi_0, \xi_{n/2} \in \mathbb{R}$ となる場合、 $\text{DFT}^{-1}(\boldsymbol{\xi})$ は実数ベクトルになる。

上で議論したように、 $\boldsymbol{\xi}$ の最初の $\lfloor n/2 \rfloor + 1$ 要素のみをメモリに持ち、学習時にこれらの値を任意に更新したとしても (ただし、 ξ_0 と $\xi_{n/2}$ は実数に保たなければならない)、結果から対称な共役複素要素を作り出し、それを iDFT すれば、実数ベクトルが保証できる。したがって、内積結果が実数になる条件を保つことができる。

3.3.2 Plate のホログラフィック縮退表現との関係

Plate はベクトル空間上の連想記憶モデルとしてホログラフィック縮退表現 (HRR) を提案している [5]。

全ての非ゼロの複素数 z は $z = re^{i\theta}$ という形で一意に表すことができる。ここで $r = |z| \in [0, \infty)$ は振幅、 $\theta \in [0, 2\pi)$ は位相を表す。Plate はフーリエ領域において複素数の振幅成分を無視し、位相のみに焦点を当て、全てのベクトル要素が $e^{i\theta} = \cos\theta + i\sin\theta$ となる複素数に限定されたモデルを考えた。Plate は実数ベクトルに基づく通常の HRR よりも分散が小さく、空間効率が良いモデルだと主張している一方、パラメータが位相に限定されるため、ニューラルネットワークなどの学習の枠組みに組み込むことが容易ではないことについても言及している。

4 実験: 知識グラフ補完タスク

Wordnet データ (WN18) と Freebase (FB15k)² 上の知識グラフ補完タスクで各手法を比較する。WN18 は訓練データに 141,442 個、開発データに 5,000 個、テス

¹もし n が奇数ならば、 ξ_0 のみ実数となる。

²<https://everest.hds.utc.fr/doku.php?id=en:transe>

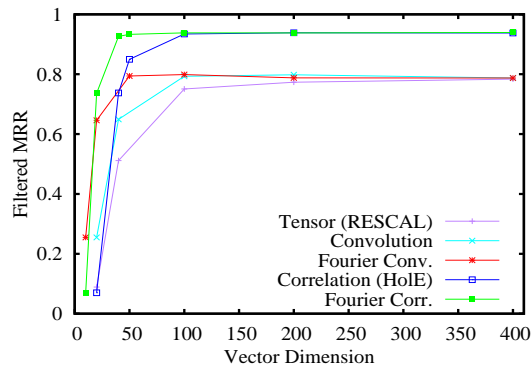


図 1: WN18 に対するベクトルの次元数と MRR の関係。

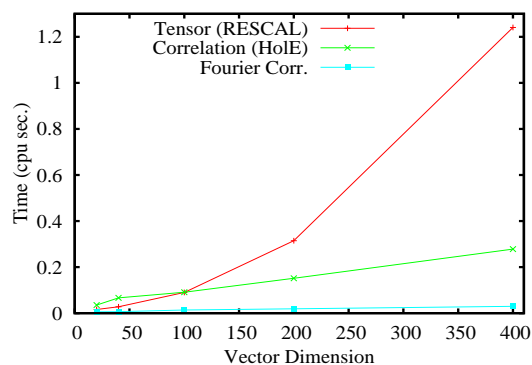


図 2: ベクトルの次元数と信頼度スコア計算にかかる時間の関係。

トデータに 5,000 個の 3 つ組を持ち、合計で 40,943 種類の entity と 18 種類の relation が存在する。FB15k は訓練データに 483,142 個、開発データに 50,000 個、テストデータに 59,071 個の 3 つ組を持ち、合計で 14,951 種類の entity と 1,345 種類の relation が存在する。評価は従来研究に従って行う。各テストサンプル (r, s, o) に対して、object o (または subject s) を全ての entity e で置き換えて、 (r, s, e) (または (r, e, o)) のスコアを計算する。そして、そのスコアに基づいて降順にソートし、平均逆順位 (MRR: mean reciprocal rank) とトップ N に正解が含まれる割合 (“Hits @ N ”) を求めて、“raw” または “filtered” の設定で評価する。“filtered” では訓練、開発、テストデータ上の $e \neq o$ (または $e \neq s$) となる (r, s, e) (または (r, e, o)) を全て削除して評価し、“raw” ではそのような削除を行わず評価する。

4.1 実験結果・分析

学習は L2 正則化付きのロジスティック回帰モデル

$$\min_0 \sum_{(r,s,o) \in \mathcal{D}} \log\{1 + \exp(-r(s, o) f(r, s, o))\} + \lambda \|\theta\|_2^2.$$

表 2: 知識グラフ補完タスク精度の比較: † は文献 [2] からの引用.

Data	Model	Raw				Filtered			
		MRR	Hits @ N			MRR	Hits @ N		
			3	5	10		3	5	10
WN18	†TransE	0.331	54.3	67.9	80.2	0.453	79.1	89.1	93.6
	†KALE-Joint	0.478	60.9	71.1	80.5	0.662	85.5	90.1	93.0
	Tensor (RESCAL)	0.564	63.4	70.3	77.2	0.783	82.3	84.7	86.9
	Correlation (HoIE)	0.635	70.9	77.4	82.7	0.938	94.0	94.3	94.3
	Fourier Corr.	0.635	71.0	77.4	82.7	0.938	94.0	94.3	94.3
FB15k	Correlation (HoIE)	0.224	24.71	32.83	44.73	0.595	66.21	69.64	73.30
	Fourier Corr.	0.224	24.71	32.83	44.73	0.595	66.21	69.64	73.30

を確率的勾配降下法で最小化して行った。θ は entity と relation のベクトル集合である。各モデルのパラメータは $n \in \{10, 20, 40, 50, 100, 200, 400\}$, $\lambda \in \{0.0, 0.05, 0.1, 0.5\}$, $\eta \in \{0.01, 0.02, 0.05, 0.1\}$ からグリッド探索して決定した (η は学習率)。各モデルに対する正例当たりの負例生成数は 5 で固定した。ベクトルの初期化については様々な手法を比較した結果、文献 [1] に従って、一様分布 $U[-\frac{\sqrt{6}}{\sqrt{2n}}, \frac{\sqrt{6}}{\sqrt{2n}}]$ からランダム生成した。フーリエ領域でのモデルでも同じ方法で生成した実数ベクトルを DFT して初期化した。また、以下の結果はベクトルの初期値と訓練データのシャッフルをランダムに変更してモデルを 5 つ作り、その結果を平均している。

図 1 に各手法のベクトル次元数と MRR 精度の関係を示す。相互相関 (Corr.) を使った手法に比べて、テンソルや畳み込み (Conv.) によるモデルは精度が悪い。これはこのタスク特有の問題に起因するもので、ある entity が subject と object で使われる場合を区別するために非可換な演算を必要とするためである。値の発散対策などにより若干の違いは出たが、本稿で示したように HoIE とフーリエ領域でのモデルは完全に等価である。そのため、フーリエ領域でのモデルのプロットをベクトルサイズが倍の位置にシフトすると時間領域でのモデルとほぼ一致する。

一方で、HoIE が $O(n \log n)$ の計算量に対し、フーリエ領域では $O(n)$ で計算できる。この計算時間の違いを実験的に示すため、WN18 の開発データ 5,000 組に対する計算時間を計測し、図 2 にプロットした。次元数が小さいときに DFT/iDFT の計算コストが理論上の計算コストを上回っているが、各手法の計算時間の差は実験的にも明らかである。また、フーリエ領域でのモデルは次元数が半分の際に HoIE と同等のモデル

になるため、提案法の方が明らかに効率的である。

表 2 に従来手法との比較を示す。HoIE は現時点で報告されている本タスクの最高精度とほぼ同等であることを明記しておく。FB15k データにおいても HoIE とフーリエ領域でのモデルの間で精度誤差はほとんど生じなかった。

5 まとめ

本稿では知識グラフ補完タスクにおいて現在最も高い精度を達成している HoIE モデルをフーリエ領域上で計算する方法を考案し、モデルを等価に保ったまま、その計算時間を $O(n \log n)$ から $O(n)$ に改善した。一方、文献 [6] で提案されているモデルも $O(n)$ で計算可能な手法となっており、本稿で提案したモデルとの関係については別の論文でまとめる予定である。

参考文献

- [1] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Aistats*, Vol. 9, pp. 249–256, 2010.
- [2] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In *EMNLP*, pp. 1488–1498, 2016.
- [3] Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. Holographic embeddings of knowledge graphs. In *AAAI*, 2016.
- [4] Maximilian Nickel, Volker Tresp, and Hans-Peter Krieger. A three-way model for collective learning on multi-relational data. In *ICML*, pp. 809–816, 2011.
- [5] Tony Plate. *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*. CSLI Publications., 2003.
- [6] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *ICML*, pp. 2071–2080, 2016.