圏論に基づく分散表現の加減算の定式化に向けて
Toward Formulating Addition and Subtraction of Word Embeddings Based on Category Theory

宮崎　祐 [1], 小林 隼人 [1,2], 菅原 晃平 [1], 山崎 朋哉 [1], 野口 正樹 [1]
Tasuku Miyazaki, Hayato Kobayashi, Kohei Sugawara, Tomoya Yamazaki, Masaki Noguchi
[1] ヤフー株式会社, [2] 理研 AIP
Yahoo Japan Corporation, RIKEN AIP
{tamiyaza, hakobaya, ksugawar, tomoyama, manoguch}@yahoo-corp.jp

概要

自然言語処理の分野では，大量の文書データを用いて単語の分散表現（実数ベクトル表現）を学習するニューラルネットワークのツール word2vec が様々な応用に活用されている．word2vec により学習された分散表現上では，単語の意味的な変換をベクトル演算として計算できることが知られており，その中でも単語のアナロジー（類推）変換を実現する計算はアナロジー操作と呼ばれている．本論文では，このアナロジー操作を可能にするベクトル空間が word2vec の学習処理の背後でどのように構築されていくのかについて，圏論を用いた定式化を試みる．具体的には，定式化の一つの試みとして，層の概念を用いてベクトル空間と内積関数空間との対応付けの方法を示し，この層に基づくホモトピー，導来圏，三角圏の安定化条件の導入により分散表現空間上の代数構造を捉える方法について議論する．

キーワード：自然言語処理, 分散表現, 認知科学, 圏論

1. はじめに

近年，様々な分野において圏論の積極的な活用が行われてきている [1, 2]．自然言語処理の分野においても，その言語の仕組みを圏論により把握する試みが行われている．文献 [3] においては 比喩理解の構造を圏論を活用して捉える試みが述べられており，文献 [4] においては，範疇文法の構造を圏論で捉える試みが述べられている．一方で，自然言語処理の分野では，大量の文書データを用いて単語の分散表現（実数ベクトル表現）を学習するニューラルネットワークのツール word2vec [5] が様々な応用に活用されている．この word2vec により学習された分散表現を用いると，単語の意味的な変換をベクトル演算として計算することができる．例えば，$\mathrm{Vec}(w)$ を単語 w の分散表現ベクトルとすると，Vec(“king”)−Vec(“man”)+Vec(“woman”) により計算されたベクトルは，単語 “king” の概念を男性から女性に変更した単語 “queen” の分散表現 Vec(“queen”) に近くなる性質を持っている．このような計算は,「単語 “man” に単語 “king” が対応するとき，単語 “woman” にはどの単語が対応するか」というアナロジー（類推・類比）の質問に答える事ができるため，アナロジー操作と呼ばれている．本論文では，このようなアナロジー操作を実現するベクトル空間が，学習時の処理によってどのように構築されていくのかを，圏論を活用することにより解明することを目標とする．具体的には，認知的視点における言語の意味の類似性が反映されたベクトル空間が，どのように構成されていくのかについて圏や層に基づいた定式化を試みる．

本論文の構成は以下の通りである．まず，2. 章で word2vec の処理の概要を説明し，3. 章で全体の定式化の方針を示す．4. 章で，層の概念を用いてベクトル空間を内積関数空間へと対応付ける方法を示す．次に，5. 章で，この層に基づくホモロジー，導来圏，三角圏の安定化条件の導入について議論する．最後に 6. 章で，まとめと今後の課題を述べる．

2. word2vec の処理概要

本論文では，word2vec で実装されているモデルの一つである skip-gram [5] に焦点を当てて定式化を検討する．skip-gram は,「単語の意味の類似性と，単語が出現する文脈の類似性には相関がある」という分布仮説に基づいており，学習データ中の各単語を実数ベクトル表現に対応付け，その実数ベクトル表現から文脈単語（例えば前後に出現する c 単語）を予測するようなモデルとなっている．具体的には，学習データを $w_1, \ldots, w_N$ としたとき，下記の対数尤度 L を最大化することで分散表現の学習が実現される．

$$L = \sum_{i=1}^{N} \sum_{j:|j|\le c, j\neq 0} \log p(w_{i+j}|w_i). \tag{1}$$

文脈単語の条件付き確率を表す関数 $p(\cdot)$ は，下記のように対数双線形モデルとして定義される．

$$p(w_O|w_I) = \frac{\exp(\tilde{\boldsymbol{v}}_{w_O} \cdot \boldsymbol{v}_{w_I})}{\sum_{w \in W} \exp(\tilde{\boldsymbol{v}}_w \cdot \boldsymbol{v}_{w_I})}. \tag{2}$$

ここで，$\boldsymbol{v}_{w_I}$ は入力単語 w_I の分散表現を表す入力ベクトル，$\tilde{\boldsymbol{v}}_{w_O}$ は単語 w_O を予測するための出力ベクトルである．

モデルの学習は，確率的勾配法 (SGD) などを用いて上述の対数尤度 L を最大化するようにパラメータ $\boldsymbol{v}_{w_I}$，$\tilde{\boldsymbol{v}}_{w_O}$ を逐次更新することで実現される．1 ステップに焦点を当てると，学習データ中の各単語（とその文脈単語）について，対数尤度 L の偏微分に基づく誤差修正値により対応するベクトルの更新が行われることになる．

3. 定式化の方針

本論文における定式化のアイデアは，ベクトルの加減算によるアナロジー操作を，ベクトル空間上ではなく，対応する内積に基づく共起確率関数の空間上で捉えることにある．2. 章で説明した通り，skip-gram の学習過程では学習データ中の各単語とその文脈単語の内積による共起確率に基づき入出力ベクトルの更新が行われている．したがって，学習後の分散表現には内積に基づく共起確率情報が埋め込まれているはずであり，対応する共起確率空間上における代数構造が捉えられれば，アナロジー操作の意味付けに繋がると考えられる．

以降の章では，大きく分けて以下の二つのステップで定式化の検討を行う．

1. 層との対応付け： ベクトル空間を内積に基づく共起確率関数の空間に対応付けるために層の概念を利用する．具体的には，実際に skip-gram で学習されるベクトル空間から内積関数空間への対応付け（関手）を定義し，その対応付けが層となることを証明する．
2. 代数構造の抽出： 層を対象とする圏を考えることで，分散表現空間上の代数構造を捉える方法について議論する．具体的には，ホモロジー，導来圏，三角圏の安定化条件の概念について説明し，それらがどのように活用できるかについて示唆を与える．詳細な対応付けについては今後の課題とする．

4. 層との対応付け

本章では，学習されるベクトル空間に対して共起確率を表す内積関数空間への層による対応付けについて述べる．層とは，位相空間上の連続的に変化する構造を捉えるための数学的概念であり，大域的な空間を局所的に取り出す概念を指す前層のうち，貼り合わせの性質の良いものをいう．以降では，まず前層との対応付けを述べ，次に層との対応付けについて述べる．

4.1 前層

前層とは，大域的な空間を局所的に取り出す概念であり，下記のように定義される．なお，定義中の位相空間 (X, O_X) と前層 (P, ρ) について，以降の議論で位相 O_X や写像 ρ の存在が明らかな場合には省略して位相 X，前層 P などと表記する．

定義 1 (前層)．(X, O_X) を位相空間とする．開集合 $U \in O_X$ に対して集合 $P(U)$ が与えられ，2 つの開集合 $U, V \in O_X$ について，包含関係 $U \subset V$ にあるとき，写像 $\rho_{UV} : P(V) \to P(U)$ が与えられているとする．以下の条件が成り立つとき，組 (P, ρ) を X 上の前層という．

1. 任意の $U, V, W \in O_X$ について，$U \subset V \subset W$ のとき，$\rho_{UV} \circ \rho_{VW} = \rho_{UW}$ である．ここで，$\circ$ は合成射を表す．
2. 任意の $U \in O_X$ について，$\rho_{UU} = \mathrm{id}_{P(U)}$ である．ここで，id は恒等射を表す．

上記定義における位相空間 X として，skip-gram の学習過程における入出力ベクトル空間をとる．基本的には，入力ベクトル空間を X_1，出力ベクトル空間を X_2 として，直積空間 $X_1 \times X_2$ を考えればよいが，議論の複雑化を避けるため層との対応付けについては片方を固定してもう一方の空間のみを考える．具体的には，学習の 1 ステップにおいて入力単語が固定されることを想定し，ある入力ベクトル $x_1 \in X_1$ について，出力ベクトル空間 X_2 が前層，及び層をなすことを示す．これにより，入力空間 X_1 上の任意の点について層が対応付けられるため，これらの層を対象とした圏を構成することが可能となる（5. 章）．入出力ベクトルが n 次元とすると，位相空間 X は下記のように書ける．

$$X = X_2 = \mathcal{R}^n \tag{3}$$

次に，対応付け（関手と呼ばれる）P の構成について述べる．関手 P はベクトル空間を共起確率関数の空間に対応付ける役割を担う．2. 章で述べたように

skip-gram の学習過程では対数尤度を考えているため，式 (2) の確率は基本的に内積で表現することができる．ここで，入力ベクトルを固定すると正規化項が定数となることに注意する．したがって，以降は内積関数空間への対応付けのみを考える．具体的には，入力ベクトル $x_1 \in X_1$ が所与として関手 P を下記のように定義する．

$$P(U) = \{f : U \to \mathcal{R} \mid \forall u \in U, f(u) = x_1 \cdot u\} \tag{4}$$

次の定理は，ベクトル空間から内積関数空間への局所的な対応付けを示す．

定理 1. 式 (4) で定義された関手 P は，式 (3) で定義された X 上の前層である．

Proof. （条件 1）任意の $U, V, W \in O_X : U \subset V \subset W$ について，どんな $f \in P(W)$ についても，f が制限写像であることから，$\rho_{UV}(\rho_{VW}(f)) = \rho_{UW}(f)$ である．（条件 2）任意の $U \in O_X$ について，どんな $f \in P(U)$ についても，$\rho_{UU}(f) = f = id_{P(U)}(f)$ である． □

4.2 層

層とは，前層のうち貼り合わせの性質が良いものをいい，下記のように定義される．

定義 2 (層). (X, O_X) を位相空間，P を X 上の前層とする．P が層であるとは，任意の $U \in O_X$ と任意の開被覆 $U = \bigcup_{i \in I} V_i$ に対して，次の 2 つの条件が成り立つことを言う．

1. $f, g \in P(U)$ が「任意の $i \in I$ に対して $f|_{V_i} = g|_{V_i}$」を満たすならば $f = g$ である．
2. $(f_i)_{i \in I} \in \prod_{i \in I} P(V_i)$ が,「任意の $i, j \in I$ に対して $f_i|_{V_i \cap V_j} = f_j|_{V_i \cap V_j}$」を満たすならば，$f \in P(U)$ が存在して，任意の $i \in I$ に対して $f|_{V_i} = f_i$ となる．

ここで，$f|_V$ は関数 f の定義域を V に制限したもので，$U, V \in O_X : V \subset U$ が与えられたとき，$f \in P(U)$ について，$f|_V = p_{UV}(f)$ として定義される．

次の定理はベクトル空間から内積関数空間への局所的な対応付けが上手く張り合わせられることを示す．

定理 2. 式 (4) で定義された関手 P は，式 (3) で定義された X 上の層である．

Proof. 定理 1 より，P は X 上の前層である．（条件 1）内積関数 $f, g \in P(U)$ が与えられたとき、任意の $u \in U$ について，$u \in V_i$ のとき，$f|_{V_i} = g|_{V_i}$ が成り立つので $f(u) = g(u)$，かつ，$U = \bigcup_{i \in I} V_i$ なので $f = g$ である．（条件 2）仮定より，任意の開集合 V_i, V_j の交わりで $f_i = f_j$ なので，下記のように関数 $f \in P(U)$ を定義できる．

$$f(u) = f_i(u) \quad \text{if} \quad u \in V_i. \tag{5}$$

定義から、任意の $u \in U$、$i \in I$ に対して $f|_{V_i}(u) = f_i(u)$ である． □

5. 代数構造の抽出

本章では，ベクトル空間から内積関数空間への対応付けを表す層を用いて，分散表現空間上の代数構造を捉える方法について議論する．具体的には，ホモロジー，導来圏，三角圏の安定化条件の概念について説明し，それらがどのように活用できるかについて示唆を与える．

5.1 ホモロジー

ホモロジーは，空間を分類するために使われる位相幾何学における基本的な概念の一つであり，位相空間や群にアーベル群や加群の列を対応させる手続きとして定義される．このホモロジーを用いることで，ベクトル空間上の各単語の変化系列がどのような関係を持っているかを抽出することが可能となり，次節で述べる導来圏の構成要素としても使われる．より具体的には，局所的に見た関数空間の変化を把握することができ，変化しないものを見つけ出すことにより，アナロジー操作の代数構造に繋がるものを見出す役割を果たす [7]. 単体複体によるホモロジーを活用した分散表現の手法としては，文献 [6] に詳細に分析されている．本研究では下記に定義されるホモロジーの利用を想定している．

定義 3 (ホモロジー). R を環とする．R 上の加群のなす圏 Mod_R はアーベル圏となる．R 加群の複体 $M^\bullet = \{M^n, d^n\}_{n \in \boldsymbol{Z}}$ は，R 加群 M^n と準同型 ${d^n}_{M^\bullet} : M^n \to M^{n+1}$ の組であって

$${d^n}_{M^\bullet} \circ {d^{n-1}}_{M^\bullet} = 0 \tag{6}$$

を満たし，下記で表される．

$$H^n(M^\bullet) = \mathrm{Ker}\ {d^n}_{M^\bullet} / \mathrm{Im}\ {d^{n-1}}_{M^\bullet}. \tag{7}$$

ここで，$\mathrm{Ker}\ d$，$\mathrm{Im}\ d$ は d の核と像を表す．

5.2 導来圏

前節で述べたホモロジーの特徴を維持しつつ，アナロジー操作の代数構造を満たすベクトル空間を構築し

たい．この代数構造（加群の圏）を作り上げるための手法として，導来圏がある．導来圏は，次節で説明する三角圏の性質をもち，この三角圏に Bridgeland の安定化条件 [8] を付与することで導来圏に座標系を定めることができる．すなわち，各単語がその属性として共起性関係を保持する抽象化されたベクトル空間としての代数構造を定めることができる．この代数構造では，分散表現の学習過程で学習データ中の単語の共起性パターンが集約されていき，相対的に同値構造を持つ空間が構成される．このような相対的な構造を持つ空間に，ホモトピー同値を導入し，単体複体の同値関係をまとめることで代数展開可能な空間にする．以下に，導来圏の定義を示す．

定義 4 (導来圏)**.** 導来圏 $D(R)$ の対象は，加群の複体のなす圏 $\mathrm{Com}(R)$ とする．$\mathrm{Com}(R)$ の写像 $f^\bullet : M^\bullet \to N^\bullet$ が与えられたとき

$$H^n(f^\bullet) : H^n(M^\bullet) \to H^n(N^\bullet) \tag{8}$$

$$\mathrm{Hom}_{D(R)}(M^\bullet, N^\bullet) = (Qis)^{-1}\mathrm{Hom}_{K(R)}(M^\bullet, N^\bullet) \tag{9}$$

ここで，$\mathrm{Hom}_{D(R)}(M^\bullet, N^\bullet)$ は，$D(R)$ 上における $M^\bullet$ から $N^\bullet$ への射を表わす．Qis は，式 (8) の $H^n(f^\bullet)$ がすべての n に関して同型写像となるとき，その写像全体を表わす．$K(R)$ は，加群の複体のなす圏 $\mathrm{Com}(R)$ をホモトピー同値による同値関係でまとめた圏を表わす．

5.3 三角圏の安定化

前節で述べた導来圏は，三角圏の性質を持つ．三角圏の定義は，下記のようになる [9].

定義 5 (三角圏)**.** τ が三角圏であるとは，τ が加法圏で

- シフト関手と呼ばれる自己同値 $[1] : \tau \to \tau$ が存在する．
- τ における完全三角圏の六つ組み (X, Y, Z, u, v, w) の族が与えられて，X, Y, Z は τ の対象で，$u : X \to Y$，$v : Y \to Z$，$w : Z \to X[1]$ は τ における射で，下記の可換図式が存在し，いくつかの公理を満たすものである．

$$\begin{tikzcd}
X \arrow[r, "u"] \arrow[d, "f"'] & Y \arrow[r, "v"] \arrow[d, "g"'] & Z \arrow[r, "w"] \arrow[d, "h"] & X[1] \arrow[d, "f[1]"] \\
X' \arrow[r, "u'"] & Y' \arrow[r, "v'"] & Z' \arrow[r, "w'"] & X'[1]
\end{tikzcd}$$

τ を三角圏とし $K(\tau)$ をグロタンディーク群とする．(Z, P) が τ 上の安定化条件とは以下の性質を満たすものである．$Z : K(\tau) \to \mathcal{G}$ は群準同型で，各実数 $\phi \in \mathcal{R}$ に対して $P(\phi) \subset \tau$ が与えられて，以下の公理を満たす．

1. $E \in P(\phi) \setminus \{0\}$ なら，正の実数 $m(E)$ が存在して $Z(E) = m(E)\exp(i\pi\phi)$ と書ける．
2. 任意の $\phi \in \mathcal{R}$ に対して $P(\phi + 1) = P(\phi)[1]$ である．
3. $\phi_1 > \phi_2$ で $A_j \in P(\phi_j)$ $(j = 1, 2)$ なら $\mathrm{Hom}_\gamma(A_1, A_2) = 0$.
4. 任意の対象 $E \in \tau \setminus \{0\}$ に対して有限個の実数列 $\phi_1 > \phi_2 > \cdots > \phi_n$ と対象 $0 = E_0, E_1, \ldots, E_{n-1}, E_n = E$ と対象 $A_j \in P(\phi)$ $(j = 1, \ldots, \mathrm{n})$ と完全三角圏 $(E_{j-1}, E_j, A_j, u_{j-1}, u_j, w_j)$ $(j = 1, \ldots, \mathrm{n})$ が存在する．

ここで，使用しているグロタンディーク群は，この導来圏の対象であるベクトル空間全体の変化をとらえるための基底を入れるのに用いている．これは加減算の構造を捉える役割を果たしており，最終的にアナロジー操作の定式化に繋がることが期待される．

6. おわりに

本論文では，分散表現上のアナロジー操作の代数構造の解明を目標に，層の概念を用いて学習時のベクトル空間を内積関数空間に対応付ける方法を示した．また，層を対象とする圏を考えることで分散表現空間上の代数構造を捉える方法として，ホモロジー，導来圏，三角圏の安定化の概念について説明し，それらの活用方法について議論した．今後は，本論文で示した方針に従い詳細な定式化を進める予定である．

文献

[1] 西郷甲矢人,（2018）“自然知能と圏論” 人工知能, Vol. 33, No. 5.

[2] 圏論の歩き方委員会, (2015) “圏論の歩き方”, 日本評論社.

[3] 布山美慕, 西郷甲矢人,（2018）“比喩理解における意味構造の対応付け：不定化した自然変換の探索として”, 32 回人工知能学会 2018.

[4] 尾崎竜史, 一杉裕志, “範疇文法の構文解析についての圏論的な視点”, 情報処理学会研究報告.

[5] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J., (2013) “Distributed Representations of-Words and Phrases and their Compositionality”, In Advances in Neural Information Processing Systems.

[6] Tadas Temčinas, (2018) “Local Homology of Word Embeddings”, arxiv preprint arXiv:1810.10136.

[7] Genki Kusano, Kenji Fukumizu, Yasuaki Hiraoka, (2016) “Persistence weighted Gaussian kernel for topological data”, arxiv preprint arXiv:1601.01741.

[8] Tom Bridgeland, (2007) “Stability conditions on triangulated categories”, Annals of Mathematics, 166, 317–345.

[9] 稲葉道明, 2013）“三角圏上における stability とモヂュライ”, 数学 65 巻 2 号　2013 年 4 月.

[10] Kawin Ethayarajh, David Duvenaud, Graeme Hirst, (2019) “Towards Understanding Linear Word Analogies”, arxiv preprint arXiv:1810.04882, ACL 2019.