構文と意味，copresheaf 意味論，AIの幾何とMagnitude Homology

平山 諒次 名古屋大学 多元数理科学研究科 Email: hirayama.ryoji.k8@s.mail.nagoya-u.ac.jp

サンサルガ

citrasyaikasvarūpasya yathā bhedanidarśanaiḥ |
nīlādibhiḥ samākhyānaṃ kriyate bhinnalakṣaṇaiḥ || 8 ||

tathaivaikasya vākyasya nirākāṅkṣasya sarvataḥ |
śabdāntaraiḥ samākhyānaṃ sākāṅkṣair anugamyate || 9 ||

たとえば、「極彩色」という一つの色がいかなるものであるかの説明は、「この部分は青、この部分は赤」といったように、その特徴がそれぞれ異なるものについて、それぞれの区分を指し示すことによってなされる。それとまさに同じように、［一つのまとまりとしては］全く他に対する期待を欠いている一つの文について、それがいかなるものであるかの説明は、互いに他に対する期待を有する個々別々の単語［の抽出］によってなされる。（赤松訳 1998, p. 45）

prefix 順序

モノイド M は一対象の圏と見ることができるが，別の圏としての見方がある．具体的には，対象を M として，$a, b \in M$ に対して $at = b$ となる $t \in M$ が存在するとき，射 $a \leq b$ が存在する，として preorder を定めることができる．これは $\mathbf{Bool}$-豊穣圏[a]である．

? 豊穣圏のベースを変えて，「他に対する期待」を量的に定められないだろうか．また，LLM はこの構造を持っているように見える．取り出せるだろうか．

✓ 可能．Bradley and Vigneaux (BV25) はその具体的なモデルを与えた．

LLM as an enriched category

有限集合 A を固定し，その元を *token* と呼ぶ．token 列がなす自由モノイド A^* に対し，$a \in A^*$ の長さを $|a|$ と書く．開始トークンと終端トークンをそれぞれ $\bot$，$\dagger$ と書き，$\widetilde{A} = A \sqcup \{\dagger\}$ と置く．

定義 (BV25, §2.1) cutoff $N \geq 1$ を固定する．集合

$$\mathrm{ob}\, L = \{\bot a \mid a \in A^*,\ |a| \leq N-1\} \sqcup \{\bot a\dagger \mid a \in A^*,\ |a| \leq N-2\}$$

に prefix の順序を入れた順序集合を $L = L_{\leq N}$ と書く．$\bot a$ の形の元を未完の文字列，$\bot a\dagger$ の形の元を完結した文字列，L の極大元を葉と呼び，葉でない元を In と書く．また，$[x, y] = \{z \in \mathrm{ob}\, L \mid x \leq z \leq y\}$ と置く．

定義 cutoff N の言語モデルとは，各 $x \in \mathrm{In}$ に $\widetilde{A}$ 上の確率分布 p_x を与える族 $p = (p_x)_{x \in \mathrm{In}}$ のことである．すなわち $p_x(c) \geq 0$ かつ $\sum_{c \in \widetilde{A}} p_x(c) = 1$．

定義 (BV25, Definition 2.8) $x, y \in \mathrm{ob}\, L$ に対し

$$\pi(y \mid x) = \begin{cases} 1 & (x = y) \\ 0 & (x \not\leq y) \\ \displaystyle\prod_{i=1}^{m} p_{z_{i-1}}(c_i) & (x < y) \end{cases}$$

と定める．第三の場合，$[x, y] = \{x = z_0 < z_1 < \cdots < z_m = y\}$ であり，$z_i = z_{i-1} c_i$（$c_i \in \widetilde{A}$）とする．

事実 (BV25, Proposition 2.16) $\mathrm{ob}\, \mathcal{L} = \mathrm{ob}\, L$，$\mathcal{L}(x, y) = \pi(y \mid x)$ と置くと，$\mathcal{L}$ は $[0, 1]$-豊穣圏である．

LLM は $[0, 1]$-豊穣圏！

LLM as a generalized metric space

定義 強モノイダル関手 $-\log : [0, 1] \to [0, \infty)$ により $\mathcal{L}$ に対応する一般化距離空間（$[0, \infty)$-豊穣圏）を $\mathcal{M}$ と書く．

$$\mathrm{ob}\, \mathcal{M} = \mathrm{ob}\, L, \qquad d(x, y) = -\log \pi(y \mid x).$$

辺 $x \to xc$ の長さを $-\log p_x(c) \in [0, \infty]$ とし，$p_x(c) > 0$ である辺を正の辺と呼ぶ．

LLM は一般化距離空間！

[a] ここで，$\mathbf{Bool}$ は二つの対象 $0, 1$ と一つの非自明な射 $0 \to 1$ を持ち，AND 演算をテンソル積とする対称モノイダル閉圏．

Magnitude Homology

以下 X を一般化距離空間とする．

定義 (LS21, Definition 4.2) $x, y, z \in X$ について，$d(x, y) + d(y, z) = d(x, z)$ が成り立つとき y は x と z の間にあるという．さらに $x \neq y$ かつ $y \neq z$ のとき真に間にあるという．

定義 (BV25, §3.13, LS21, Definition 3.3) $k \geq 0$，$\ell \in [0, \infty)$ とする．隣接する点が異なり，$\sum_{i=0}^{k-1} d(x_i, x_{i+1}) = \ell$ を満たす点列 $(x_0, \ldots, x_k) \in X^{k+1}$ の集合を $G_{k,\ell}(X)$ と書き，$MC_{k,\ell}(X) = \mathbb{Z}[G_{k,\ell}(X)]$ と置く．$0 < i < k$ について，準同型 $\partial_i \colon MC_{k,\ell}(X) \to MC_{k-1,\ell}(X)$ を

$$\partial_i(x_0, \ldots, x_k) = \begin{cases} (x_0, \ldots, \widehat{x_i}, \ldots, x_k) & (\text{下記の条件を満たすとき}) \\ 0 & (\text{それ以外}) \end{cases}$$

により生成元上で定める．条件は，x_i が x_{i-1} と x_{i+1} の間にあり，かつ $x_{i-1} \neq x_{i+1}$ であることとする．$\partial = \sum_{i=1}^{k-1} (-1)^i \partial_i$ と置く（$k \leq 1$ ならば $\partial = 0$）．$(MC_{*,\ell}(X), \partial)$ を X の *magnitude complex* と呼ぶ．

定義 X が $d(x, y) = d(y, x) = 0 \implies x = y$ を満たす[a]とする．$k \geq 0$，$\ell \in [0, \infty)$ について，X の *magnitude homology* を

$$H_{k,\ell}(X) := H_k(MC_{*,\ell}(X)) = \frac{\ker(\partial \colon MC_{k,\ell}(X) \to MC_{k-1,\ell}(X))}{\mathrm{im}(\partial \colon MC_{k+1,\ell}(X) \to MC_{k,\ell}(X))}$$

で定める．ただし $MC_{-1,\ell}(X) = 0$ とする．

消滅定理（結果）

定義 $\ell \in [0, \infty)$ について，長さ ℓ の正の辺の集合を

$$E_\ell^+ = \{\, x \to xc \mid x \in \mathrm{In},\ c \in \widetilde{A},\ p_x(c) > 0,\ -\log p_x(c) = \ell \,\}$$

と置く．

定理 (HEX) 任意の $\ell \in [0, \infty)$ について次が成り立つ．

$$\begin{aligned} &H_{k,\ell}(\mathcal{M}) = 0 \qquad (k \geq 2), \\ &H_{1,\ell}(\mathcal{M}) \cong \mathbb{Z}[E_\ell^+], \\ &H_{0,0}(\mathcal{M}) \cong \mathbb{Z}[\mathrm{ob}\, L], \qquad H_{0,\ell}(\mathcal{M}) = 0 \quad (\ell > 0). \end{aligned}$$

$k = 0, 1$ の部分は一般的な結果 LS21, Theorems 4.1, 4.3 をこの空間に当てはめたもので，BV25, p. 22 が既に指摘している．同論文は高次のホモロジーが消えることを予想として述べていた．

? もっと非自明な幾何学はあるか？ また，$\mathcal{M}$ はどちらかというと構文の構造に見える．意味との対応は？

✓ Bradley, Terilla, and Vlassopoulos (BTV22) による copresheaf 意味論が使えそう．

? 意味の世界が cocomplete であることの意味は何か．「意味論の張り合わせ」はどのように行えるか．それぞれの構成要素の情報を見てそれらを集めることと，全体の情報を見ることの差異はこのようなモデルにおいてどのように表れてくるか．

参考文献

Bhartṛhari (1966). *Vākyapadīya of Bhartṛhari: With the Vṛtti and the Paddhati of Vṛṣabhadeva*. Ed. by K. A. Subramania Iyer. Vol. 1. Poona: Deccan College Postgraduate and Research Institute.

Bradley, Tai-Danae, John Terilla, and Yiannis Vlassopoulos (2022). "An Enriched Category Theory of Language: From Syntax to Semantics". In: *La Matematica* 1, pp. 551–580. arXiv: 2106.07890.

Bradley, Tai-Danae and Juan Pablo Vigneaux (2025). *The Magnitude of Categories of Texts Enriched by Language Models*. Version 2. arXiv: 2501.06662 [cs.CL]. URL: https://arxiv.org/abs/2501.06662v2.

Vanishing of Higher Magnitude Homology for Finite Prefix Models (2026). Contrib.: Ryoji Hirayama. URL: https://hexagonmath.org/2610.00445v2.

Leinster, Tom and Michael Shulman (2021). "Magnitude Homology of Enriched Categories and Metric Spaces". In: *Algebraic & Geometric Topology* 21, pp. 2175–2221. arXiv: 1711.00802.

バルトリハリ (1998). 古典インドの言語哲学（1）：ブラフマンとことば. Trans. by 赤松明彦. 東洋文庫 637. 平凡社.

[a] magnitude complex が鎖複体になる条件．