

## 「大規模言語 Coherence モデル」

フーリエ位相による低損失構成モデリング：

複素指数ベクトルのフォン・ノイマンエントロピー

### “Large-scale Language Coherence Model” Low-Loss Compositional Modeling via Fourier Phases: Von Neumann Entropy of Complex Exponential Vectors

吹谷和雄 MathDoc.

2025.08.22

#### 【概要】

□大規模言語モデル (LLM) の内部表現は、高次元埋め込み空間における意味的な干渉や重ね合わせとして理解できる。しかし、従来の研究で用いられる Fourier Feature (座標位置をサイン・コサイン基底で拡張する方法) や Spectral Representation (信号を周波数領域に展開して表現する方法) は、主に「入力空間の座標的パターン」を捉えるために利用されており、言語意味の階層性や推論過程に直結する設計にはなっていない。

□本研究では、Embedding 空間を「波動的な意味干渉の場」と捉え、系列方向に対して DFT を適用することで、文中の意味的な相関や順序構造を周波数・位相の観点から解析する。従来の Fourier Feature が「外部入力の特徴マッピング」に留まるのに対し、本研究の立場は、LLM 内部に潜む表現そのものを量子統計的に再解釈し、Von Neumann エントロピーを用いて「意味の分散度」や「情報の周波数的集中度」を定量化する点に新規性がある。

この視点により、Attention による選択的強調を「位相空間における帯域選択」として統一的に説明でき、従来のスペクトル的アプローチでは得られなかった「意味と干渉の理論的基盤」を与えることが可能となる。

#### 離散フーリエ変換(DFT)

□トークン埋め込みベクトルを  $d$  次元ベクトルとし、文のトークン系列長を  $T$  とする。 ( $t = 0, \dots, T-1$ )

- ・文の埋め込み(ここでは one-hot)を縦ベクトル  $v(t) \in \mathbb{R}^d$  とし、文全体を

$$V^{(seq)} = [v(0)v(1)\dots v(T-1)] \in \mathbb{R}^{d \times T}$$

とする。

□系列方向の DFT を各次元ごとに施す。単位 ary のフーリエ行列を

$$(F_T)_{t,n} = \frac{1}{\sqrt{T}} e^{-i2\pi nt/T} \quad (t, n = 0, \dots, T-1)$$

とすると、周波数領域表現は、

$$\hat{V}^{(freq)} = V^{(seq)} F_T \in \mathbb{C}^{d \times T}$$

ここで  $\hat{V}_{:,n}^{(freq)}$  は、周波数  $n$  の「複素指数ベクトル(周波数成分の  $d$  次元ベクトル)を与

える。系列方向に DFT を施した結果を

$$v_n \in \mathbb{C}^d \quad (n = 0, 1, \dots, T-1)$$

すると、

文全体の周波数領域表現を

$$V^{(freq)} = [v_0, v_1, \dots, v_{T-1}] \in \mathbb{C}^{d \times T}$$

と書き換えられる。

□この流れの処理を、例文『猫が走る』

3次元 one-hot 埋め込みベクトルとして詳

しく説明する。

「猫」 「が」 「走る」

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

とする。

この例文は、

$$v(t) = \begin{bmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad t = 1, 2, 3$$

… ①

と行列で表現される。

□離散フーリエ変換：DFT

$$\hat{v}_k(n) = \sum_{t=0}^{N-1} v_k(t) e^{-iwn/N}$$

各行(各次元)ごとに“系列”を離散フーリエ変換(DFT)する。

(1)  $v_1(t)$  の系列 [1 0 0]  $v(t)$  の 1 行目

$$x[0] = 1, \quad x[1] = 0, \quad x[2] = 0$$

$$X[n] = 1 \cdot e^{-iwn \cdot \frac{0}{3}} + 0 + 0 = 1$$

$$\therefore X = [1 \quad 1 \quad 1] \quad \dots \textcircled{2}$$

(2)  $v_2(t)$  の系列 [0 1 0]  $v(t)$  の 2 行目

$$x[0] = 0, \quad x[1] = 1, \quad x[2] = 0$$

$$X[n] = 0 + 1 \cdot e^{-iwn \cdot \frac{1}{3}} + 0 = e^{-iwn/3}$$

$$\begin{aligned} \therefore X &= \begin{bmatrix} e^{-i0} & e^{-\frac{iw}{3}} & e^{-\frac{iw2}{3}} \end{bmatrix} \\ &= [1 \quad e^{-i\theta/3} \quad e^{-i2\theta/3}] \quad \dots \textcircled{3} \end{aligned}$$

(3)  $v_3(t)$  の系列 [0 0 1]  $v(t)$  の 3 行目

$$x[0] = 0, \quad x[1] = 0, \quad x[2] = 1$$

$$X[n] = 0 + 1 \cdot e^{-iwn \cdot 2/3} = e^{-iwn \cdot 2/3}$$

$$\begin{aligned} \therefore X &= \begin{bmatrix} e^{i0} & e^{-\frac{iw2}{3}} & e^{-\frac{iw4}{3} = -\frac{iw}{3}} \end{bmatrix} \\ &= [1 \quad e^{-i2\theta/3} \quad e^{-i\theta/3}] \quad \dots \textcircled{4} \end{aligned}$$

したがって、①と②、③、④から

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-i\theta} & e^{-i2\theta} \\ 1 & e^{-i2\theta} & e^{-i\theta} \end{bmatrix} \quad \dots \textcircled{5}$$

$$\text{where, } \theta = \frac{w}{3} = 2\pi/3$$

となる。

### Von Neumann Entropy(VNE)

□複素指数ベクトルの「散らばり(不確実さ)を量るために、量子統計で使う密度演算子  $\rho$  を周波数成分から構成し、

フォン・ノイマン・エントロピー

$$\text{VNE: } S(p) = -\text{Tr}(p \ln p)$$

を計算する。

・周波数  $n$  の複素指数ベクトル

$\hat{v}_n := \hat{v}_{:,n} \in \mathbb{C}^d$  に対し、情報量パワーを

$$E_n = \|\hat{v}_n\|_2^2 = \sum_{j=1}^d |\hat{v}_{j,n}^{(freq)}|^2 \quad (\geq 0)$$

と定義する。

全パワーを  $E_T = \sum_{n=0}^{T-1} E_n$  とする。

確率分布として、

$$p_n = \frac{E_n}{E_T}, \quad \sum_n p_n = 1$$

を得る。

これは周波数基底で対角な密度演算子に相当する。

□スペクトル対角形(周波数基底)  $\{|n\rangle\}_{n=0}^{T-1}$  を取り、

$$\rho_{spec} = \sum_{n=0}^{T-1} p_n |n\rangle \langle n| \in \mathbb{C}^{T \times T}$$

このとき、VNE は、

$$S(\rho_{spec}) = - \sum_{n=0}^{T-1} p_n \ln p_n$$

これは、周波数パワー分布のシャノン・エントロピーに一致する。

□幾何的(射影の混合)では、各周波数ベクトルを正規化

$$|\phi_n\rangle = \frac{\hat{v}_n}{\|\hat{v}_n\|} \quad (\hat{v}_n \neq 0)$$

として、

$$\rho_{geom} = \sum_{n: E_n > 0} p_n |\phi_n\rangle \langle \phi_n| \in \mathbb{C}^{d \times d}$$

・ここでも固有値 $\{p_n\}$ (非自明な重複を除く)になり、VNE は同じ値に一致する( $\rho_{geom}$ の台となる部分空間に限る)。したがって、VNE は「どの周波数に情報が分散しているか」を測る量になる。

□例文で説明すると、

⑤式の各周波数の情報量パワーは、

$$E_n = \|\hat{v}_n\|_2^2 = \frac{1}{3} \cdot 3 = 1 \quad (n = 0, 1, 2)$$

よって、 $E_t = 3$ なので、

$$p_0 = p_1 = p_2 = \frac{1}{3}$$

従って、密度演算子の対角形は、

$$\rho_{spec} = \text{diag}\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

で、これから VNE は、

$$S(\rho_{spec}) = - \sum_{n=0}^2 \frac{1}{3} \ln \frac{1}{3} = \ln 3$$

となる。

幾何的には、

$$\rho_{geom} = \sum_{n=0}^2 \frac{1}{3} |\phi_n\rangle \langle \phi_n|$$

$$\text{where, } |\phi_n\rangle = \hat{v}_n$$

となる。

□これは固有値集合 $\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$ を持つので、

上記の通り、VNE は「 $S = \ln 3$ 」となる。最大エントロピーから、この極端な均一な例文では「まず低周波( $n=0$ )で文の主題(文全体)を捉え、その後に  $n=1, 2$  を順次加えて文局所的な次予測(語順)を捉える」という自然な方針が導かれる。

### 周波数選択の原理

□VNE による周波数選択の意味合い。

・ $S(\rho)$ が大きい：情報量が多く周波数に広く分散。低周波(粗い位相)を優先し、大域的トピックや要約、主題抽出を行う。

・ $S(\rho)$ が小さい：情報量が少数の周波数に集中。高周波(細かい位相)を追加して語順や構文、推論の局所整合性を高める。

実装としては、

所望の帯域 $\Omega \subset \{0, \dots, T-1\}$ を選び、

$$\rho_\Omega = \sum_{n \in \Omega} \frac{p_n}{\sum_{m \in \Omega} p_m} |n\rangle \langle n|$$

の VNE の最大化(または所定の閾値を満たす)するように $\Omega$ を束縛しつつ、Band-pass 逆 DFT で系列側に戻す：

$$V_\Omega = \hat{V}_{:, \Omega} F_T^H(:, \Omega)$$

最大化ではなく、「なるべくエントロピーを崩さず周波数を選ぶ」という方針にすれば、KL ダイバージェンス最小化

$$\min_{\Omega} D_{KL}(p \parallel \tilde{p}_\Omega)$$

$$\tilde{p}_\Omega(n) = \begin{cases} \frac{p_n}{\sum_{m \in \Omega} p_m}, & n \in \Omega \\ 0, & \text{otherwise} \end{cases}$$

と定式化する。

※この構成は、Attention の“選択的強調”を位相空間で再解釈しつつ、周波数選択をエントロピー保存的に最適化する道具立てになる。

## MultiHeadAttention との統合

□ヘッドごとに  $\hat{V}^{(h)} = V^{(h)}F_T$  を計算し、重み  $w_h$  によって合成：

$$E_n = \sum_{h=1}^H w_h \|\hat{V}_{:,n}^{(h)}\|_2^2$$

$$p_n = \frac{E_n}{\sum_m E_m}$$

残りは同様に行う。これにより「アテンション=周波数選択」とみなせ、VNE が情報分散の要約になる。

## 理論的性質

□数学的な命題を示します。

- ・命題 1 (ユニタリ不変性)：トークン軸に沿う任意のユニタリ変換は  $S$  を不変に保つ。
- ・命題 2 (上限)：  $0 \leq S \leq \ln r$ 、ただし  $r = |\{n: E_n > 0\}|$ 。一様分布のとき上限に等しい。
- ・命題 3 (等価性)：  $\rho_{geom}$  と  $\rho_{spec}$  は、同じ固有値を持ち、VNE が一致する。

## 例文の詳細

□フーリエ行列  $F_T$  から、  $w = e^{-i2\pi/3}$  とする

$$\text{と、 } F_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & w & w^2 \\ 1 & w^2 & w \end{bmatrix}$$

となり、計算結果例は、

$$F_3 \approx \begin{bmatrix} 0.577 & 0.577 & 0.577 \\ 0.577 & -0.289 - 0.5i & -0.289 + 0.5i \\ 0.577 & -0.289 + 0.5i & -0.289 - 0.5i \end{bmatrix}$$

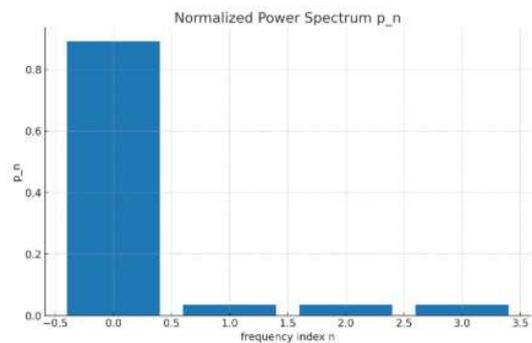
というユニタリフーリエ行列になる。

ユニタリ性のチェックは、  $F_3 F_3^H = I_3$  になる。

ちなみに、例文を  $T=4$  にした場合のユニタリフーリエ行列を示す。

$$F_4 \approx \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & -0.5i & -0.5 & 0.5i \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0.5 & 0.5i & -0.5 & -0.5i \end{bmatrix}$$

□以下に例文  $T=4$  のユニタリ DFT を適用し、周波数ごとの情報量  $E_n$ 、正規化スペクトル  $p_n$ 、フォン・ノイマンエントロピー VNE を計算すると下図のようになる。



縦棒は、正規化スペクトル  $p_n$

相対再構成誤差：  $6.53 \times 10^{-17}$

(数値的誤差はほぼゼロ)

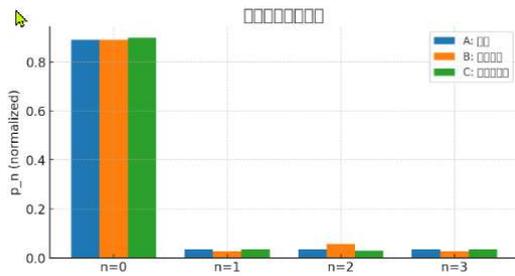
VNE：  $S=0.4623 \text{ nats} (=0.6669 \text{ bits})$

・この例文では、  $p_0$  (低周波) が支配的で、VNE が小さめ → 情報が低周波に集中している状況です。

□そこで、文の語順入れ替えと同義語置換をして比べてみる。

- ・ A (元文)：「猫が素早く走る」  
元の埋め込み行列  $V_A$  を使用
- ・ B (語順入替)：「猫が走る素早く」  
列を入れ替えて語順を変更
- ・ C (同義語置換)：「猫が速やかに走る」  
「素早く」の埋め込みを小さな摂動

下図のようになる。



結果：

- ・ VNE(nats)
  - ・ A(元文)：0.46227
  - ・ B(語順入替)：0.45397
  - ・ C(同義語置換)：0.43599
- ・ VNE(bits)]
  - ・ A: 0.6669 bits
  - ・ B: 0.6549 bits
  - ・ C: 0.6290 bits

□評価：

今回の例文では、語順入替(B)も同義語置換(C)も VNE を減少させた。これは情報がより特定周波数に集中した(分散が小さくなった)ことを示す。

・棒グラフは、各周波数インデックス  $n$  に対する  $p_n$  の比較を示している。元文では低周波( $n=0$ )が支配的でしたが、変更によって多少分布が変わっている。

□まとめ：

- ・ VNE が下がる→周波数(情報)が一部に集中している(高周波または低周波に偏ることが多い)。
- ・ 語順の入れ替えは位相変換を伴い、しばしば高周波側の情報を移すことがある(この例文は、既に低周波寄りだった為に変化は小さい)。
- ・ 同義語の置換は、埋め込みの大きさや方向の微小変更なので、ケースによっては VNE に与える影響が小さいか大きいか

分かれる(語彙距離に依存)。

## 今後の課題

□本研究においては、言語埋め込み系列に対するフーリエ変換の適用と、複素指数ベクトルに基づくフォン・ノイマン・エントロピーの導入により、言語情報の分散度を定量化する理論的枠組みを提示した。しかしながら、本稿で示した成果は理論的考察および小規模な例証に限定されており、今後は以下の諸点について更なる検討を要する。

1. 大規模データに基づく実証的評価  
現段階では、短文を対象とした基礎的な検証にとどまっている。今後は、大規模コーパスや標準的ベンチマークを用い、提案指標が意味解析・推論タスクにおいて一貫した有効性を有するかを精緻に検証する必要がある。
2. 既存大規模言語モデルとの統合的検討  
Attention 機構を周波数選択の観点から再解釈する理論的可能性を提示したものの、現実の Transformer 系モデルへの実装および性能評価は未着手である。今後は、提案手法をモデル内部に組み込み、その学習挙動ならびに推論精度への影響を実証的に明らかにすることが課題となる。
3. 周波数選択原理の高度化  
本研究では、エントロピー最大化を指導原理とした周波数帯域の選択を定式化した。しかし、自然言語処理においては文脈依存性や意味階層構造が不可避的に関与する。従って、動的かつ文脈適応的な周波数選択手法の確立が求められる。
4. 情報理論的指標の比較検討  
フォン・ノイマン・エントロピーの適用可

能性を示したが、今後は Rényi エントロピーや相対エントロピー等の他の情報理論的尺度との比較を行い、多面的な評価枠組みを構築することが望まれる。

#### 5. 応用的展開の可能性

本研究の理論枠組みは、意味的干渉や順序構造の解析にとどまらず、生成過程の安定化、翻訳精度の向上、要約や情報圧縮といった応用領域においても寄与し得る。今後は、具体的応用への展開を通じてその汎用性を検証する必要がある。 □