層の冗長性と層同士の独立性に基づく言語モデルの層交換の成否の特徴づけ

小林春斗 [1] 原知正 [1] 鴨田豪 [1] 横井祥 [2,1,3]
[1] 東北大学 [2] 国立国語研究所 [3] 理化学研究所
{kobayashi.haruto.t8, hara.tomomasa.s8, go.kamoda}@dc.tohoku.ac.jp
yokoi@ninjal.ac.jp

概要

ニューラル言語モデルを一度学習し，後からその層を繋ぎ直すことで，モデルの軽量化や複数のモデルの統合が可能になるという不思議な現象が知られている．本稿では，最も単純な層の繋ぎ変えである隣接層同士の交換の成否が,「層の冗長性」と「層同士の独立性」というふたつの直感的な指標によって特徴づけられることを示す．理論的には，これらの量が十分小さいことが，層同士を交換できることの必要条件になっていることを示した．経験的には，提案指標が学習済みの GPT-2 の層同士の交換しやすさが提案尺度でよく予測できることを確認した．

1 はじめに

自然言語処理分野では，複数の層からなる深層学習モデル [1] を採用した大規模言語モデルが基盤的な役割を果たしている [2–4]．これら大規模言語モデルを構成する層同士を「繋ぎ変え」ても動くという興味深い性質がいくつも報告されている．例えば，同一のモデル内において，隣接する中間層同士の交換 [5–7] や，層の削除 [8,9]，層の並列化 [7] といった操作を行なってもほとんど性能が低下しないことが報告されている．また，複数のモデルを層単位で繋ぎ合わせることでそれらの特徴を受け継ぐ新たなモデルを構築する，モデルマージが可能であること [10] も示されている．

自然な疑問として，モデルがどのような条件を満たすときに「層の繋ぎ変え」が可能になるのだろうか？ Gromov ら [8] や Men ら [9] は，隠れ状態をどの程度変化させないかという層の冗長性によって，層の削除の成否を特徴づけた．しかしながら，これまで報告されている言語モデルの改造方法は，層の削除だけではなく，層の繋ぎ変えやモデルマージといったより複雑な操作を含む．こうした操作の成否が何に特徴づけられているか，いつ成功するのかは重要なオープンクエスチョンとして残されている．

本稿では「層の冗長性」と「層同士の独立性」というふたつの直感的な概念から，層の繋ぎ変えを特徴づけるための新たな指標を構成した (図 1)．とくに繋ぎ変えの最も素朴な設定である隣接する層同士の交換に着目し，理論的にも経験的にも提案指標が良い性質を持つことを確認した．さらに，提案指標を用いることで，交換が失敗する原因を特定できることを示した．

2 提案指標

図 1 に提案指標の概要を示す．まず，図 1 (a) に示すように，交換対象となる 2 つの層のうちいずれか一方の層が何もしない，つまり**層が冗長**になっていれば，交換しても出力は一致する．第 i 層の「非」冗長性の度合い，つまり恒等変換以上のことをする度合いを $\mathrm{Trans}(i)$ で表し，1 つ目の特徴とする．また，図 1 (b) に示すように，仮にそれぞれの層の入力領域がもう一方の層の出力領域と重なっていなければ，すなわち**連続する層が独立**であれば層を交換しても出力は一致する．第 i 層の出力と第 $i+1$ 層の入力の「非」独立性の度合いを $\mathrm{Dep}(i, i+1)$ とし，2 つ目の特徴とする．これら 2 種の量のうち少なくとも一つの値が十分小さければ，交換可能となるだろう．また，最終的な尺度を 2 種の尺度の積で構成することで,「いずれかが十分小さければ小さい」指標を構築する．

2.1 節では，より厳密に提案指標を構成する．3 節では，提案指標が最小値を取ることが 2.2 節で定義する関数としての交換可能性の必要条件になっている（良い尺度になっている）ことを示す．4.1 節では，実際の言語モデルに提案法を利用する場合の計

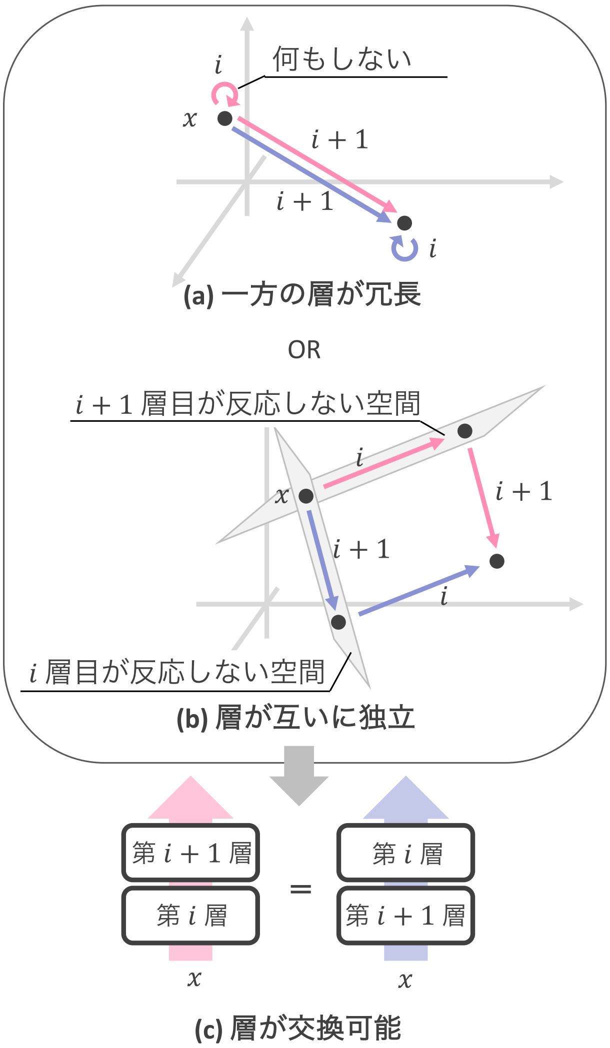

図 1　提案指標の概要．(a) 交換する層のうち，少なくともいずれか一方が十分に冗長であれば，(c) 交換可能であると予想される．また，交換する層が冗長でない場合にも，(b) それぞれの層が互いに独立しているのであれば (c) 交換可能であると予想される．

算方法を述べ，その後実験的な検証をおこなう．

なお，他の特徴付けとして「ふたつの層の機能が同じなので合成関数がどの順番でも一致する」といった観点も考えられる．しかしながら，Lad ら [6] によって隣接する層同士の変換の類似度は低いことが判明しているため，本稿では考慮しないこととした．

2.1　指標の構成

第 i 層の非冗長さを $\mathrm{Trans}(i)$，第 $i+1$ 層の第 i 層への依存度を $\mathrm{Dep}(i,i+1)$ とする．ここで $\mathrm{Trans}(i)$ は，第 i 層が隠れ状態を一切変化させないときに限り 0 をとるように構成する．また $\mathrm{Dep}(i,i+1)$ は，第 $i+1$ 層が第 i 層の影響を一切受けないときに限り 0 となるようにする．

これらを用いると，上記の直観的解釈は以下のような指標にまとめることができる：

$$r(i,i+1) := \mathrm{Trans}(i)\mathrm{Trans}(i+1)\big(\mathrm{Dep}(i,i+1)+\mathrm{Dep}(i+1,i)\big) \tag{1}$$

例えば，第 i 層が冗長であれば，$\mathrm{Trans}(i)=0$ であり $r(i,i+1)=0$ が成り立つ．

2.2　交換しやすさの測定方法

本稿では，隣接する層同士の交換の特徴づけを試みる．そのため，図 1 (c) に示される 第 i 層と第 $i+1$ 層の交換しやすさを，「入力隠れ状態の大きさに対する，交換によって生じた隠れ状態の変化の大きさ」として，次のように実際に測定する[1]：

$$\delta(i,i+1) := \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} \frac{\|L_{i+1}(L_i(\boldsymbol{X}_i))_t - L_i(L_{i+1}(\boldsymbol{X}_i))_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2} \tag{2}$$

ただし，$\boldsymbol{X}_{i,t}$ は，i 層目の入力隠れ状態 $\boldsymbol{X}_i$ の t 行目に当たる行ベクトルを表す．これは，t 単語目の位置の隠れ常態ベクトルに対応する．さらに，$L(\boldsymbol{X}) = \boldsymbol{X} + \ell(\boldsymbol{X})$ であり，$\ell_i(\cdot)$ は第 i 層で行われる非線形変換を指す．また，$\mathbb{E}$ は期待値をとる操作を表す．

3　理論的な分析

本節では，2.1 節で構成した直感的な提案指標が最小値を取ることが，2.2 節の方法で測定する実際の交換可能の必要条件になっていることを示す．いくつかの仮定を置くことで，提案指標が δ の上界と関係していることも示すことができるが，こちらに関しては，参考情報 A.1 に記載した．

指標 $r(i,i+1)=0$ のとき，式 1 より，以下のうち少なくとも 1 つは成立する．

$$\mathrm{Trans}(i)=0 \vee \mathrm{Trans}(i+1)=0 \tag{3}$$

$$\mathrm{Dep}(i,i+1)=0 \wedge \mathrm{Dep}(i+1,i)=0 \tag{4}$$

以下ではこれらで場合分けを行い，必要条件になっていることを述べる．

冗長な場合　まず，式 3 で $\mathrm{Trans}(i)=0$ が成立している場合，2.1 節での定義より，以下が成立する．

$$L_i(\boldsymbol{X}) = \boldsymbol{X} \tag{5}$$

1）　入力隠れ状態の大きさで除して正規化しているのは，層番号によらずに比較するためである．

式 5 を式 2 へ代入すれば，直ちに $\delta(i, i+1) = 0$ が得られる．$\mathrm{Trans}(i+1) = 0$ が成立している場合も同様である．

独立している場合　まず，各層の反応する空間と，各層の出力が含まれる空間の存在を仮定する．このとき，各層に対して，ある直交射影行列 $\boldsymbol{P}_i^{\mathrm{in}}, \boldsymbol{P}_i^{\mathrm{out}}$ と，ある非線形変換 $\ell'_i(\cdot)$ が存在して[2)]，次の等式が成立する．

$$\ell_i(\boldsymbol{X}) = \ell_i(\boldsymbol{X})\boldsymbol{P}_i^{\mathrm{out}} = \ell'_i(\boldsymbol{X}\boldsymbol{P}_i^{\mathrm{in}}) \tag{6}$$

式 4 が成立するとき，2.1 節で課した「$\mathrm{Dep}(i, i+1) = 0$ ならば第 $i+1$ 層が第 i 層の出力の影響を一切受けない」という条件から，以下が成立する．

$$\boldsymbol{P}_i^{\mathrm{out}}\boldsymbol{P}_{i+1}^{\mathrm{in}} = \boldsymbol{O} \land \boldsymbol{P}_{i+1}^{\mathrm{out}}\boldsymbol{P}_i^{\mathrm{in}} = \boldsymbol{O} \tag{7}$$

ここで，式 6，7 より，次式が成立する．

$$\ell_i(\boldsymbol{X} + \ell_{i+1}(\boldsymbol{X})) = \ell'_i(\boldsymbol{X}\boldsymbol{P}_i^{\mathrm{in}} + \ell_{i+1}(\boldsymbol{X})\boldsymbol{P}_{i+1}^{\mathrm{out}}\boldsymbol{P}_i^{\mathrm{in}}) \tag{8}$$

$$= \ell'_i(\boldsymbol{X}\boldsymbol{P}_i^{\mathrm{in}}) \tag{9}$$

$$= \ell_i(\boldsymbol{X}) \tag{10}$$

$$\ell_{i+1}(\boldsymbol{X} + \ell_i(\boldsymbol{X})) = \ell_{i+1}(\boldsymbol{X}) \tag{11}$$

これを式 2 へ代入すれば，直ちに $\delta(i, i+1) = 0$ が得られる．

上記の議論から，指標が最小値を取ることは，層同士が交換可能な必要条件になっている．以降では，どのようにして層の冗長性と層同士の独立性を測定するかについて述べる．

4　経験的な分析

4.1　測定方法

理論的には 3 節に示す関係が成立するが，複雑な構造を持つ実際のモデルにおいて「層の冗長性」，「層同士の独立性」を直接測るのは困難である．そこで本稿では，実際のモデルの層を線形近似してから「層の冗長性」，「層同士の独立性」を推定し，実際のモデルにおける層同士の交換しやすさとの関係を分析する．

近似の詳細　層の近似は以下の式に基づいて行う：

$$\boldsymbol{A}_i = \underset{\boldsymbol{A}_i}{\mathrm{argmin}}\left(\underset{\boldsymbol{X},t}{\mathbb{E}} \|\boldsymbol{X}_{i,t} + \hat{\boldsymbol{X}}_{i,t}\boldsymbol{A}_i - \boldsymbol{X}_{i+1,t}\|_2\right) \tag{12}$$

ただし，$\hat{\boldsymbol{X}}$ は，t 番目の語と直前 3 単語に当たる隠れ状態を平均したものを表し，注意機構による文脈の利用を簡易的に考慮している．なお，本近似の妥当性については，4.4 節で議論する．

冗長性の測定方法　層の非冗長性 $\mathrm{Trans}(i)$ を，以下のように測定する．これは，「i 層で生じる隠れ状態の変化の大きさが，入力隠れ状態に対して平均してどの程度大きいか」を意味し，図 1 (a) に示した概念に対応する．

$$\mathrm{Trans}(i) := \frac{1}{\sqrt{d}}\|\boldsymbol{A}_i\|_{\mathrm{F}} \tag{13}$$

層同士の依存度の測定方法　層同士の独立性 $\mathrm{Dep}(i, i+1)$ を，以下のように測定する．これは，第 i 層の出力する部分空間 $\mathrm{Im}(\boldsymbol{A}_i)$ と第 $i+1$ 層が反応しない部分空間 $\mathrm{Ker}(\boldsymbol{A}_{i+1})$ との間に，$\mathrm{Im}(\boldsymbol{A}_i) \subseteq \mathrm{Ker}(\boldsymbol{A}_{i+1})$ が成り立っているときに限り 0 をとるような尺度であり，図 1 (b) に示した概念に対応する．

$$\mathrm{Dep}(i, i+1) := \frac{\|\boldsymbol{A}_i\boldsymbol{A}_{i+1}\|_2}{\|\boldsymbol{A}_i\|_2\|\boldsymbol{A}_{i+1}\|_2} \tag{14}$$

以降では，上記の枠組みでモデルの「層同士の 交換しやすさ」を評価できるのかを検証する．

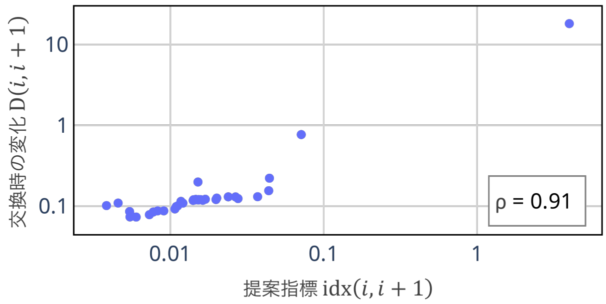

図 2　提案指標と交換時の影響 δ との関係．各点が，隣接する層同士の交換に対応する．また，図中の ρ は spearman の順位相関係数を表している．

4.2　提案指標と交換しやすさ

本節では，2 節で述べた提案指標について，その有効性の検証を行う．

実験設定　提案指標の値と交換しやすさとの関係を実験的に確認する．モデルは，36 層からなる GPT-2 Large [3)] を使用した．各指標を計測するためのデータは，gpt-2-output dataset[4)] 内の webtext データを使用した．汎化性能を測るために，提案指標の測定に webtext.train.jsonl の一部を，交換しやすさ δ の測定には webtext.valid.jsonl の一部を使用した．

結果　結果を図 2 に示す．提案指標と実際のモデルの 交換しやすさとの間に高い相関が確認された，

2）　厳密には斜交射影行列も考えられるが，ここでは直交射影行列に限定する．

3）　https://huggingface.co/openai-community/gpt2-large
4）　https://github.com/openai/gpt-2-output-dataset

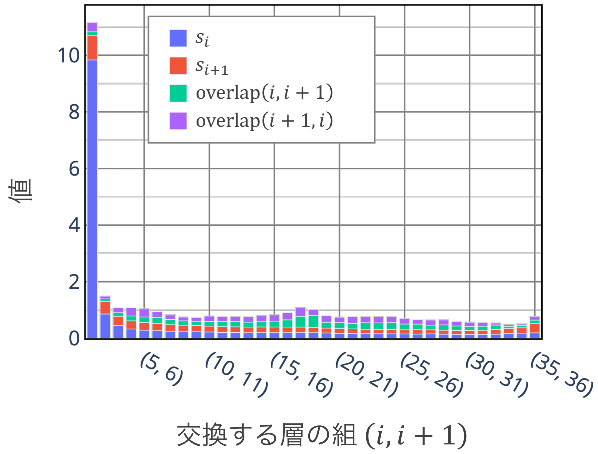

図 3　提案指標の各構成要素の値.

spearman の順位相関係数は 0.91 であり，層の冗長性と層同士の独立性を考慮した提案指標が有効であることが示唆された.

4.3　特に影響を与える要素の検討

提案指標 (式 1) は，層の冗長性に対応する項 2 つと，層同士の独立性に対応する項 2 つの計 4 種類で構成されている．本節では，特にどの項が交換の難しさに影響していたかの特定を行い，層の特性を考察する.

実験設定　式 1 で表される指標の各構成要素の値を確認する．モデル・データは 4.2 節と同一のものを使用した.

結果　結果は図 3 のようになった．第 1 層と第 2 層の組に対する指標の値は，Trans(1) によって大きな値をとっていたことがわかる．また，第 35 層と第 36 層の組に対する指標の値は，Trans(36) によって大きな値をとっていた．これらの原因として，モデルは最初の層と最後の層で大きく隠れ状態を変化させ，単語埋め込みをモデルが扱いやすいような形に変換している可能性が考察される．第 17 層と第 18 層の組に対する指標の値は，Dep(17, 18) によって大きな値をとっていた．これらの層ではふたつの層をかけて他の層よりも深い推論を行なっている可能性が考察される.

4.4　近似の妥当性

提案指標がどの程度理論に沿っているかは，近似行列がどの程度実際のモデルを反映しているかに依存する．本節では，近似行列がどの程度実際のモデルを反映しているかを検討し，提案指標の妥当性を考える.

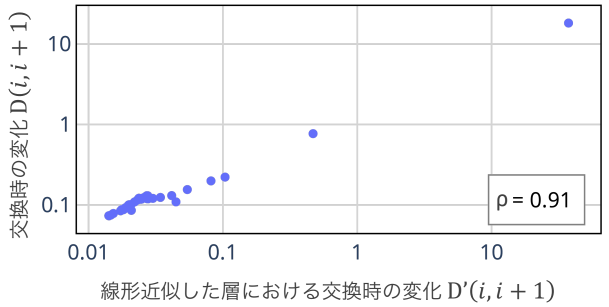

図 4　線形なモデルにおける交換の影響 δ' と，実際のモデルにおける交換の影響 δ の関係.

実験設定　実際のモデルで測定した 交換しやすさ δ と，線形近似したモデルで測定した 交換しやすさ δ' との対応の確認を行う．ただし，線形近似したモデルにおける 交換しやすさ δ' は以下のように定義した：

$$L'_i(X) := X + XA_i \tag{15}$$

$$\delta'(i, i+1) := \mathop{\mathbb{E}}_{X_{i,t}} \frac{\|L_{i+1}(L_i(\hat{X}_i))_t - L_i(L_{i+1}(\hat{X}_i))_t\|_2}{\|\hat{X}_{i\,t}\|_2} \tag{16}$$

結果　結果を図 4 に示す．spearman の順位相関係数は 0.91 であり，強い相関が認められた．これにより．線形近似は，層の交換を分析する上ではある程度妥当であることが支持された.

5　おわりに

本稿では,「層の冗長性」と「層同士の独立性」に着目し，隣接する層同士の交換しやすさの特徴づけを試みた．まず，直感的な概念から提案指標を構成した．さらに，理論的・経験的に機能することを確認したほか，指標を用いたモデルの簡単な解釈も行なった．今後，手法を発展させ，より複雑な繋ぎ変えの可否の分析が可能になることが期待される．また，指標を用いて繋ぎ変えを阻害する要因の特定し，その修正を試みるといった方向への発展も期待する.

謝辞

本研究は JST FORET JPMJFR2331，JSPS 科研費 JP22H05106，JP22H00524 の助成を受けたものです．

参考文献

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In **Advances in Neural Information Processing Systems (NIPS)**, 2017.

[2] Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper, and Bryan Catanzaro. Megatron-lm: Training multi-billion parameter language models using model parallelism. **arXiv preprint**, 2020.

[3] Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, and Laurent Sifre. Training compute-optimal large language models. **arXiv preprint**, 2022.

[4] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models. **arXiv preprint**, 2023.

[5] Srinadh Bhojanapalli, Ayan Chakrabarti, Daniel Glasner, Daliang Li, Thomas Unterthiner, and Andreas Veit. Understanding robustness of transformers for image classification. In **IEEE/CVF International Conference on Computer Vision (ICCV)**, pp. 10211–10221, 2021.

[6] Vedang Lad, Wes Gurnee, and Max Tegmark. The remarkable robustness of llms: Stages of inference? **arXiv preprint**, 2024.

[7] Qi Sun, Marc Pickett, Aakash Kumar Nain, and Llion Jones. Transformer layers as painters. **arXiv preprint**, 2024.

[8] Andrey Gromov, Kushal Tirumala, Hassan Shapourian, Paolo Glorioso, and Daniel A. Roberts. The unreasonable ineffectiveness of the deeper layers. **arXiv preprint**, 2024.

[9] Xin Men, Mingyu Xu, Qingyu Zhang, Bingning Wang, Hongyu Lin, Yaojie Lu, Xianpei Han, and Weipeng Chen. Shortgpt: Layers in large language models are more redundant than you expect. **arXiv preprint**, 2024.

[10] Takuya Akiba, Makoto Shing, Yujin Tang, Qi Sun, and David Ha. Evolutionary optimization of model merging recipes. **arXiv preprint**, 2024.

A　参考情報

A.1　交換の難しさの上界と提案指標

第 i 層に対して反応する線形部分空間 U_i と，各層の出力が含まれる線形部分空間 W_i の存在を仮定する．また，$\boldsymbol{X}_{i,t}$ は常に U_i に含まれるものとする．このとき，

$$\delta(i,i+1) = \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} \frac{\|L_{i+1}(L_i(\boldsymbol{X}_i))_t - L_i(L_{i+1}(\boldsymbol{X}_i))_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2} \tag{17}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} \frac{\|(\boldsymbol{X}_{i,t} + \ell_i(\boldsymbol{X}_i)_t + \ell_{i+1}(\boldsymbol{X}_i + \ell_i(\boldsymbol{X}_i))_t) - (\boldsymbol{X}_{i,t} + \ell_{i+1}(\boldsymbol{X}_i)_t + \ell_i(\boldsymbol{X}_i + \ell_{i+1}(\boldsymbol{X}_i))_t)\|_2}{\|\boldsymbol{X}_{i,t}\|_2} \tag{18}$$

$$\leq \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} \frac{\|\ell_i(\boldsymbol{X}_i)_t - \ell_i(\boldsymbol{X}_i + \ell_{i+1}(\boldsymbol{X}_i))_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2} + \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} \frac{\|\ell_{i+1}(\boldsymbol{X}_i)_t - \ell_{i+1}(\boldsymbol{X}_i + \ell_i(\boldsymbol{X}_i))_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2} \tag{19}$$

$$= \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} \frac{\|\ell'_i(\boldsymbol{X}_i)_t - \ell'_i(\boldsymbol{X}_i + \ell_{i+1}(\boldsymbol{X}_i)\boldsymbol{P}_{i+1}^{\mathrm{out}}\boldsymbol{P}_i^{\mathrm{in}})_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2} + \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} \frac{\|\ell'_{i+1}(\boldsymbol{X}_i)_t - \ell'_{i+1}(\boldsymbol{X}_i + \ell_i(\boldsymbol{X}_i)\boldsymbol{P}_i^{\mathrm{out}}\boldsymbol{P}_{i+1}^{\mathrm{in}})_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2} \tag{20}$$

ここで，以下のように表せると仮定する．

$$\frac{\|\ell'_i(\boldsymbol{X}_i)_t - \ell'_i(\boldsymbol{X}_i + \ell_{i+1}(\boldsymbol{X}_i)\boldsymbol{P}_{i+1}^{\mathrm{out}}\boldsymbol{P}_i^{\mathrm{in}})_t\|_2}{\|\ell'_i(\boldsymbol{X}_i)_t\|_2} = f\left(\frac{\|\ell_{i+1}(\boldsymbol{X}_i)_t\boldsymbol{P}_{i+1}^{\mathrm{out}}\boldsymbol{P}_i^{\mathrm{in}}\|_2}{\|\boldsymbol{X}_{i,t}\|_2}\right) \tag{21}$$

ただし f は単調増加関数であり，$f(0) = 0$ とする．これは，左辺で表される ℓ_i で生じる相対的な変化は，ℓ_i への入力が $\boldsymbol{X}_i$ に近いほど小さくなると仮定することを意味する．さらに $f(x) \leq cx$（c は正定数）と仮定すれば，

$$\delta(i,i+1) \leq \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} c\frac{\|\ell'_i(\boldsymbol{X}_i)_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2}\frac{\|\ell_{i+1}(\boldsymbol{X}_i)_t\boldsymbol{P}_{i+1}^{\mathrm{out}}\boldsymbol{P}_i^{\mathrm{in}}\|_2}{\|\boldsymbol{X}_{i,t}\|_2} + \mathop{\mathbb{E}}_{\boldsymbol{X},t} c\frac{\|\ell'_{i+1}(\boldsymbol{X}_i)_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2}\frac{\|\ell_i(\boldsymbol{X}_i)_t\boldsymbol{P}_i^{\mathrm{out}}\boldsymbol{P}_{i+1}^{\mathrm{in}}\|_2}{\|\boldsymbol{X}_{i,t}\|_2} \tag{22}$$

$$\leq \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} c\frac{\|\ell_i(\boldsymbol{X}_i)_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2}\frac{\|\ell_{i+1}(\boldsymbol{X}_i)_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2}\|\boldsymbol{P}_{i+1}^{\mathrm{out}}\boldsymbol{P}_i^{\mathrm{in}}\|_2 + \mathop{\mathbb{E}}_{\boldsymbol{X}_{i,t}} c\frac{\|\ell_{i+1}(\boldsymbol{X}_i)_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2}\frac{\|\ell_i(\boldsymbol{X}_i)_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2}\|\boldsymbol{P}_i^{\mathrm{out}}\boldsymbol{P}_{i+1}^{\mathrm{in}}\|_2 \tag{23}$$

さらに $\frac{\|\ell_i(\boldsymbol{X}_i)_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2}$ と $\frac{\|\ell_{i+1}(\boldsymbol{X}_i)_t\|_2}{\|\boldsymbol{X}_{i,t}\|_2}$ とが独立であると仮定すれば，

$$\delta(i,i+1) \leq c\mathrm{Trans}(i)\mathrm{Trans}(i+1)(\mathrm{Dep}(i,i+1) + \mathrm{Dep}(i+1,i)) = cr(i,i+1) \tag{24}$$

このように，様々な仮定のもとで，指標は影響の上界と関係していることが示せる．