

# 大規模言語モデルの 原理と可能性

宮尾祐介

東京大学大学院情報理工学系研究科

# 大規模言語モデルとは

- 大量のテキストデータから学習された言語モデル
- 言語モデル：単語列  $w_1, \dots, w_k$  の生成確率  $p(w_1, \dots, w_k)$

$$p(w_1, \dots, w_k) = \prod_{i=1}^k p(w_i | w_1, \dots, w_{i-1})$$

- 自然言語っぽい（人間が書いたっぽい）単語列は高い確率

YU

大規模言語モデルとはなんですか？



大規模言語モデルは、非常に大量のテキストデータを使用してトレーニングされた、自然言語処理（NLP）タスクを解決するための人工知能のモデルです。これらのモデルは、言語の構造や文法を理解し、テキストの意味を推測する能力を持っています。



ChatGPT

# 言語モデルとは

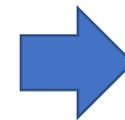
- 単語列  $w_1, \dots, w_k$  の生成確率  $p(w_1, \dots, w_k)$

$$p(w_1, \dots, w_k) = \prod_{i=1}^k p(w_i | w_1, \dots, w_{i-1})$$

- $p(w_k | w_1, \dots, w_{k-1})$  を大量のテキストデータから学習
- 文脈  $w_1, \dots, w_{k-1}$  から次の単語  $w_k$  を予測  
→ 人間が書くような文章が生成できる

我輩 は 猫 で ある 。 名前 は

$w_1$     $w_2$     $w_3$     $w_4$     $w_5$     $w_6$     $w_7$     $w_8$



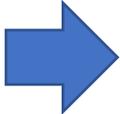
?

$w_9$

# Nグラムモデル

- 直前の  $N - 1$  単語のみに依存して、次の単語の確率が決まる
  - それ以前の単語が異なるものは全て同一と考える

$$p(w_k | w_1, w_2, \dots, w_{k-1}) = p(w_k | w_{k-N+1}, \dots, w_{k-1})$$

我輩 は 猫 で ある 。 名前 は  ?

$w_1$     $w_2$     $w_3$     $w_4$     $w_5$     $w_6$     $w_7$     $w_8$     $w_9$

$p(w_9)$

ユニ(1)グラム

$p(w_9 | w_8)$

バイ(2)グラム

$p(w_9 | w_8, w_7)$

トライ(3)グラム

$p(w_9 | w_8, w_7, w_6)$

4-グラム

$p(w_9 | w_8, w_7, w_6, w_5)$

5-グラム

# $N$ グラムモデルの問題

- 単語数 10,000 としたときのパラメータ数

モデル	パラメータ数
2グラムモデル	$10,000^2 = 1$ 億
3グラムモデル	$10,000^3 = 1$ 兆
4グラムモデル	$10,000^4 = 1$ 京

- $N$  が大きい方が精度が高いと思われるが、単純に  $N$  を増やせない
- 少しでも異なる  $N$  グラムは別の確率事象 → 意味が近い単語列はまとめられるのでは？
  - 「名前 は」と「名まえ は 、」はだいたい同じはず

# 分散表現

- 単語やフレーズの意味をベクトルで表す
  - 単語埋め込み (word embedding) ともいう
- **word2vec**: 周辺の単語が似ている単語 (= 意味が近い単語) は近いベクトルになるよう学習

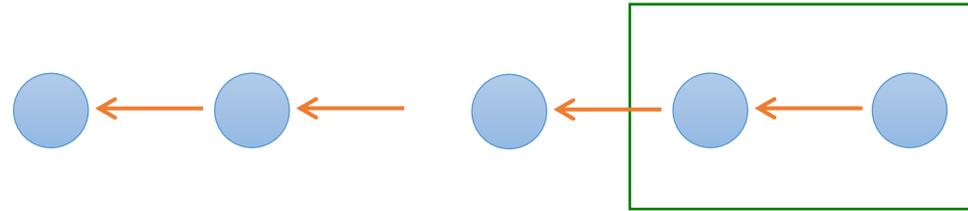
$$p(w_i | w_{i-T}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+T}) = \text{softmax}(\bar{w}_i \cdot \bar{c}_i)$$
$$\bar{c}_i = \bar{w}'_{i-T} + \dots + \bar{w}'_{i-1} + \bar{w}'_{i+1} + \dots + \bar{w}'_{i+T}$$

- 分散表現を使って確率を与える → 意味が近い単語列をまとめる

$$p(w_k | w_1, w_2, \dots, w_{k-1}) = f(\bar{w}_1, \bar{w}_2, \dots, \bar{w}_{k-1}, \bar{w}_k)$$



# ニューラルネットワーク



$$p(y|x) = \mathbf{a}_n(W_n \mathbf{a}_{n-1}(\dots \mathbf{a}_2(W_2 \mathbf{a}_1(W_1 \mathbf{x}))))$$

活性化関数

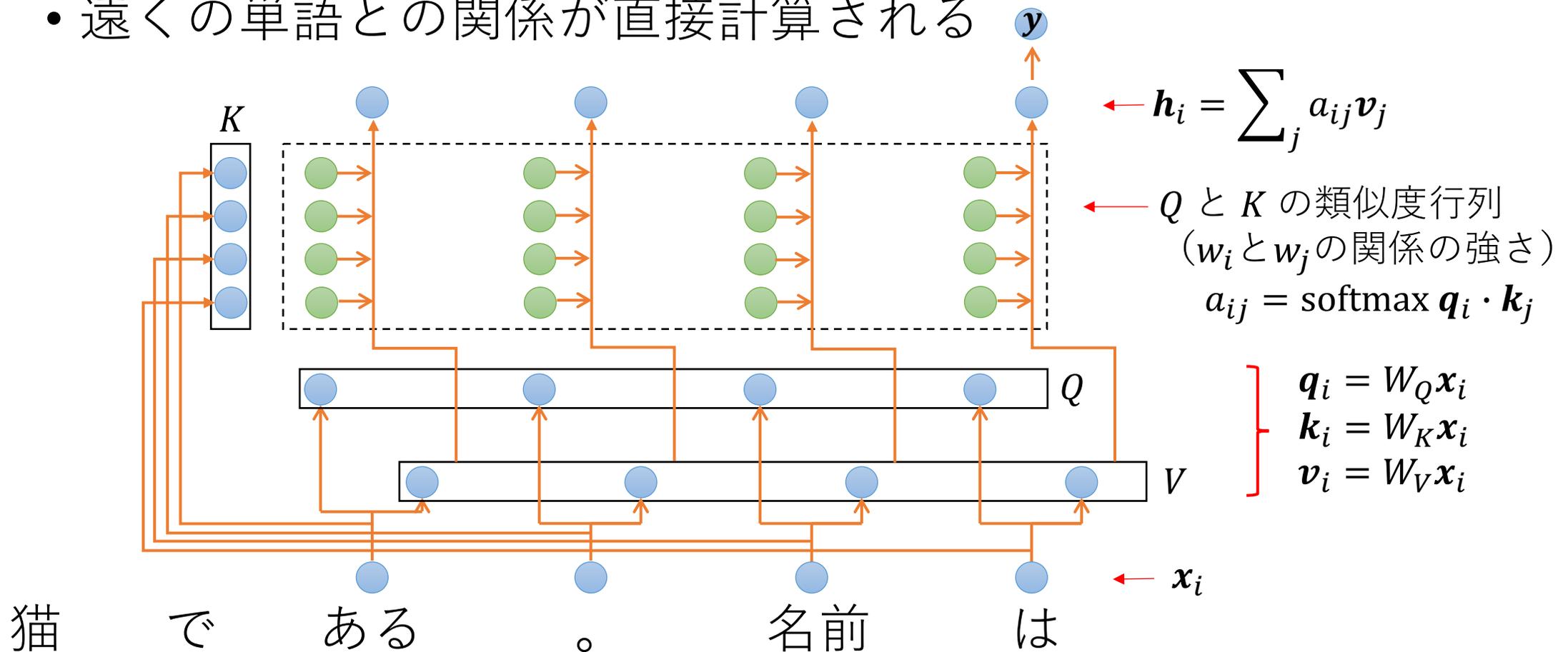
パラメータ行列

- ベクトルからベクトルへの関数（線形写像と非線形関数）を繰り返して適用
- 出力が正しくなる（= 損失関数を最小化する）ようなベクトルになるようにパラメータを学習
  - 言語モデルの場合は、次の単語の予測が正確になるように学習



# Transformer

- 周辺単語との関係に基づき各単語のベクトルを変換
- 遠くの単語との関係が直接計算される



# 言語モデルがやること・やらないこと

- 大規模言語モデルは言語モデルである
  - 文脈  $w_1, \dots, w_{k-1}$  に応じて人間が書いたかのような文章が生成できる
  - **Transformer** と大規模テキストデータでの学習により、文脈の高度な抽象化が実現されている（ようだ）
- 事実を述べるように設計されてはいない
  - 人間が書いたかのような文章を生成することで、ある程度事実を述べることができる（言語モデルの中に知識が組み込まれている）
  - 学習データに含まれる知識はニューラルネットワークのパラメータとして埋め込まれており、確実に引き出せるわけではない
  - 現状の原理（のみ）では、事実性を保証することは困難
- 感情・意志・情動を持つように設計されてはいない
- （おそらく）人間と同じしくみで言語理解・生成しているわけではない

# 大規模言語モデルの可能性

- インタフェースとしての自然言語理解・生成は人間と同等以上
  - 人間と機械の壁を取り払うことができる
- 実世界のさまざまな問題に応じて言語モデルとその他システムを組み合わせれば、さまざまな応用が可能に
  - 事実を述べてほしい → 検索やデータベースと組み合わせ
  - 感情を持った対話をさせたい → 感情認識・生成と組み合わせ
  - 実世界とのインタラクション → ロボットと組み合わせ
  - 特定タスクを正確に遂行させたい → 個別システムと組み合わせ
  - 楽にいろいろなタスクを実行させたい → プロンプト自動生成・最適化
- 法的・倫理的問題の対応は必要（本講演では割愛）



# 教育に関して

- これまでの講演で既にさまざまな角度から議論されている
- 教育内容・方法 → あまり変わらず、あるいはより効果的に
  - 言語能力、論理的思考、批判的思考、専門知識、**etc.** は大規模言語モデルを使いこなすために必要
  - 教育の個別化など、大規模言語モデルによってより高い教育効果が望める
- 教育における評価 → さまざまな混乱やアップデートが必要なのはしかたがない
- 専門家の育成 → 次世代の言語モデルやさまざまな応用を実現していくにはコンピュータサイエンス等の専門性が必須
  - コンピュータサイエンス：省エネルギー化、品質保証、応用可能性、**etc.**
  - 自然言語処理研究：言語モデルの分析・理解、学習方法、データ整備、**etc.**

# 自然言語処理研究の今

- さまざまな試行錯誤を経て、自然言語処理技術が初めて空を飛んだ
- 新たな疑問が続々と
  - なぜ飛べるのか？
  - 鳥（人間）と何が違うのか？
  - どういう活用ができるか？
  - よりよく飛ぶ方法は？
- 大規模言語モデルや人間の言語理解についてはまだわかっていないことが多い
- 大規模言語モデルによる自然言語処理のインパクトが示されたことで、これらの問いがより重要に



大規模言語モデルの基礎理論、  
分析、応用、新たなモデル



# おわりに

- 今後さまざまな大規模言語モデルや、それを利用した応用システムが実現
  - ChatGPT は一つの製品であり、最終形ではない
  - 大規模言語モデルのインパクトが示されたことで研究開発が加速
- 現在指摘されている問題点は徐々に解決されていく
  - 嘘をつく・正確性
  - 学習・運用のコスト
  - バイアス
  - ブラックボックス性
  - プロンプトへの過敏性
- 人間とは異なる形で言語能力を実現している前提で活用・改善方法やルールを探求すれば、社会のあらゆる場面で活躍が期待される