

Harris, Z. (1954).
Distributional structure.
Word, 10(23): 146-162.

理化学研究所 ロボティクスプロジェクト
知識獲得・対話研究チーム
吉野 幸一郎

自己紹介

- **吉野 幸一郎 Koichiro Yoshino**

- ウェブページ: pomdp.net
- 2020.09- 理化学研究所 ロボティクスプロジェクト
知識獲得・対話研究チーム チームリーダー
- -2020.08 奈良先端大 中村研 助教（現在は客員准教授）

- **研究の興味**

- 言語の理解
- 対話での情報授受
- Embodied Dialogue
- 強化学習

- **ロボティクスプロジェクトで研究員を募集しています！**

- ロボットシステムの構築と言語理解、
知識獲得の枠組み構築
- 特に分散意味論に強い人

みなさん言語処理をする時何をしますか？



初手 word2vec

w

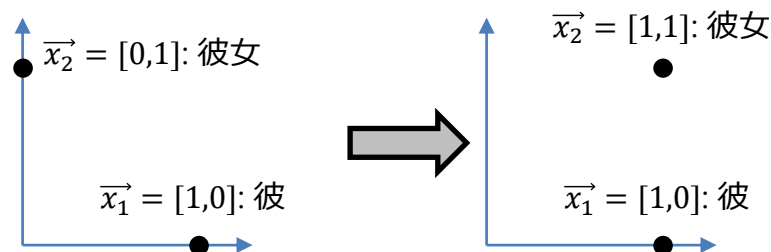
$[0.2, 0.345, 0.12 \dots]$

とりあえず Glove

文脈に応じて BERT (Transformer)

• 単語や文を分散表現へ変換

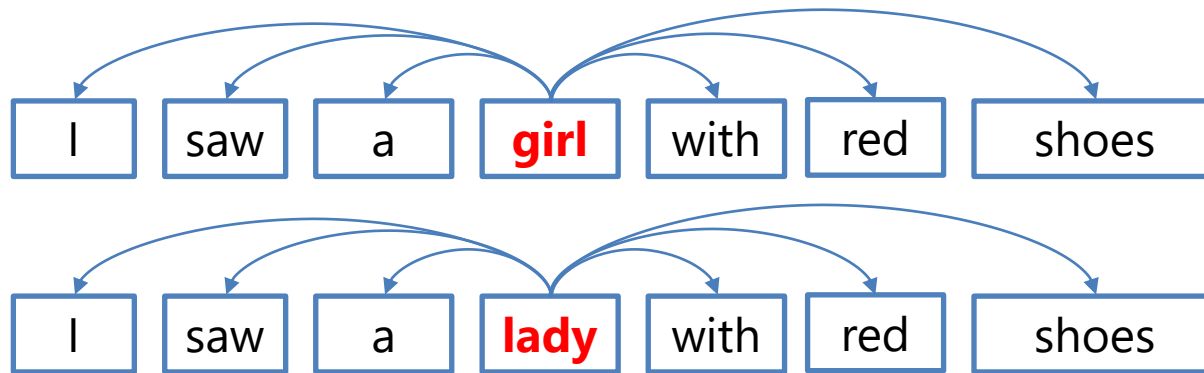
- 離散表現をベクトル表現にできる
- 単語の意味とかを緩く扱いたい
 - それ以前のスタンダードであったベクトル空間モデルは語の類似性などを扱えない



これらのベクトル作成には分布仮説が使われている

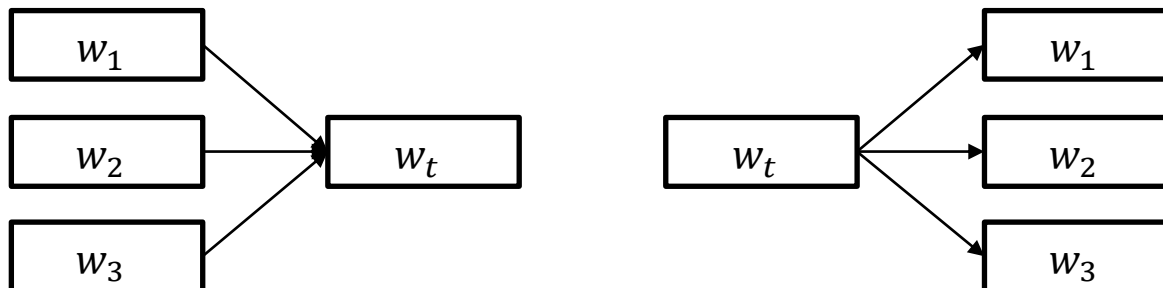
分布仮説(distributional hypothesis)とは？

- 同じ文脈で出現する語が同様の意味を持つ傾向があるという仮説



- 今日の分散意味論の基礎

– CBOW, Skip-gram, next sentence prediction などはこの考えに基づく



分布仮説を議論した論文

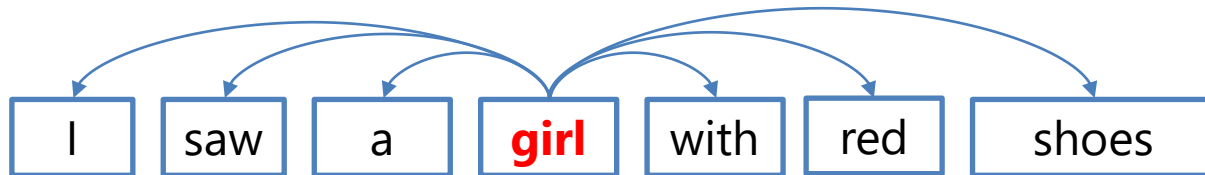
- **Harris, Z. (1954). Distributional structure. *Word*, 10(23): 146-162.**
- **Firth, J.R. (1957). A synopsis of linguistic theory 1930-1955. In *Studies in Linguistic Analysis*, pp. 1-32. Oxford: Philological Society. Reprinted in F.R. Palmer (ed.), *Selected Papers of J.R. Firth 1952-1959*, London: Longman (1968).**
- **McDonald, S., and Ramscar, M. (2001). Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 611-616.**
- **Sahlgren, Magnus (2008) The Distributional Hypothesis. *Rivista di Linguistica (Italian Journal of Linguistics)*, 20 (1). pp. 33-53.**

https://aclweb.org/aclwiki/Distributional_Hypothesis

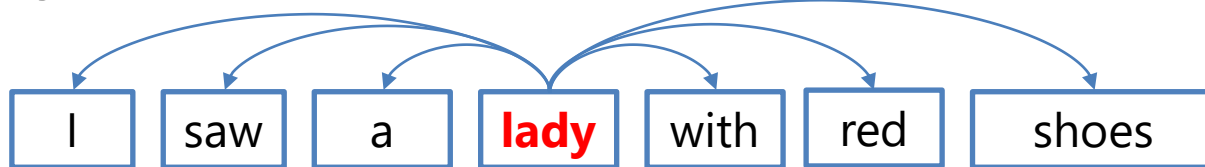
最初の論文には何が書かれているのか？

言語は分布構造を持つのか？

- 単語、音素などの言語を構成する要素が持つ意味をどのように（何でもって）説明するか
- 説明の方法の一つは周囲の環境で説明すること
 - つまり共起語と共起語との相対位置を素性を用いて説明できる



- girl の1語前に a が共起した、1語後に with が共起した



- girl の1語前に a が共起した、1語後に with が共起した

論文の主張

- 1. 言語の各要素は任意の順序・相対位置では発生しない**
 - 単語の生成順序、生成内容には制約がある
- 2. あらゆる要素の生成はその出現位置に制約を持つ**
 - この制約は確率的に議論することができる
- 3. 文中の他の要素との相対状態を見ることで、その要素が文法規則に従うか知ることが出来る**
 - スラングなど従来は文法規範に当てはまらないと考えられてきた現象も他の語との相対位置で説明できる
- 4. 共起する他の要素の分布を見ることで語の説明ができる**
 - もちろん等確率にはならない

※ element を言語の要素、文中の要素と訳しているが、典型的には単語や音素を指す

論文の主張

1. 言語の各要素は任意の順序・相対位置では発生しない
→ 今日の分布意味論 (distributional semantics) の基礎
2. あらゆる要素の生成はその出現位置に制約を持つ
→ 確率的言語モデルそのもの
3. 文中の他の要素との相対状態を見ることで、
その要素が文法規則に従うか知ることが出来る
→ 言語モデル尤度に対する考え方
4. 共起する他の要素の分布を見ることで語の説明ができる
→ 今日の分布意味論 (distributional semantics) の基礎

※ element を言語の要素、文中の要素と訳しているが、典型的には単語や音素を指す

分布仮説を補強するいくつかの事実

- 文の生成途中であっても分布仮説は成立する
- 特定の言語における音素のパターンのようなものも分布仮説は存在する
 - Evaluate → Evaluat**ion**, Investigate → Investigat**ion**
- 分布仮説は既存の文法規則も包含するように説明できる
- 究極には、話者やクラスタによっても分布の傾向は異なるのでこれらのモデル化を行うこともできる
 - あるクラスタに属するテキストから学習した言語モデルに対する対数尤度でクラス所属確率を見ることができるとも
 - ただし発話がそのクラスに属することとイコールではない

分布構造と意味構造の関係

- **言語の分布構造は意味構造を表すのか？**
 - 言語の生成分布に偏りがあることはわかったが...
 - 多くの場合、言語の構造は主観的な意味世界の構造に**対応しない**
 - 「言葉に出来ない感情」「言語化できないアイデア」
 - 時代と共に変わる言葉の使われ方
- **語は単体で多様な意味を持ち、意味（語義）と一対一对応しない**
 - 語義は周囲の単語の分布や談話構造を観測することで一意にできる
 - ただしそれが意味そのものと一対一对応するわけではない

形態素は意味に基づくのか

- **仮説: 単語の形態素（ここで言う形態素は英単語のprefixなど）の意味が周囲の文脈によって決定される**
- **単語を構成する形態素が意味を決定しているわけではない**
 - person と persist は同じ prefix を持つように見えるが関係はない
- **ただし周囲の文字列、形態素などから意味を予測できる**

形態素は意味に基づくのか

• チェロキー語の例

- zinagali'a: I am cleaning
- agiyoseha: I am hungry
- uniyoseha: they are hungry
 - Prefix の z を I と訳すか？ z と ag の関係は？
- ozinagali'a: I and others are cleaning
- ogiyoseha: I and others are hungry

チェロキー語	
ᏍᏏᏉᏍᏏ ᏍᏏᏉᏍᏏ	
話される国	アメリカ合衆国
地域	オクラホマ州、ノースカロライナ州
話者数	12,000人 ~ 22,000人(2007年) ^{[1][2]}
言語系統	イロコイ語族
表記体系	チェロキー文字、ラテン文字
言語コード	
ISO 639-1	なし
ISO 639-2	chr
ISO 639-3	chr



デンプレートを表示

チェロキー語は
Native Americanの少数言語

文の意味とパターンから語義を推測可能
形態素境界は周辺パターンの複雑さによって決定可能

分布で意味を捉える

- 周辺語の分布で形態素の意味（語義）を捉えることは可能
- ただし**スパーシティの問題**は大きい
 - 形態素同士の組み合わせは使われ方に依存
 - neighbor → neighborhood はあらゆる場合に成り立つか？
 - 形容詞は名詞を修飾するが、あらゆる形容詞と名詞の組み合わせが存在するわけではない
 - 音素の組み合わせに関しても同様
 - 日本語における q u e など

**これら組み合わせの偏りを意味のクラスとして捉える
e.g., suffix に “hood” を持つ単語群**

語義の比較

- **周辺語の類似を取ることでおおよその類似度を測ることが可能**
 - eye-doctor と oculist は同義
- **分布の異なる部分でそれらの違いを測ることも可能**
 - knife と knives → 三単現のs や is/are など数詞に関わるものが異なるがそれ以外は似た分布を持つ
- **ある語について全ての共起語を見ることはできない
(とされている)**
 - 現代的言語モデルはそれに近いところまで行っている？

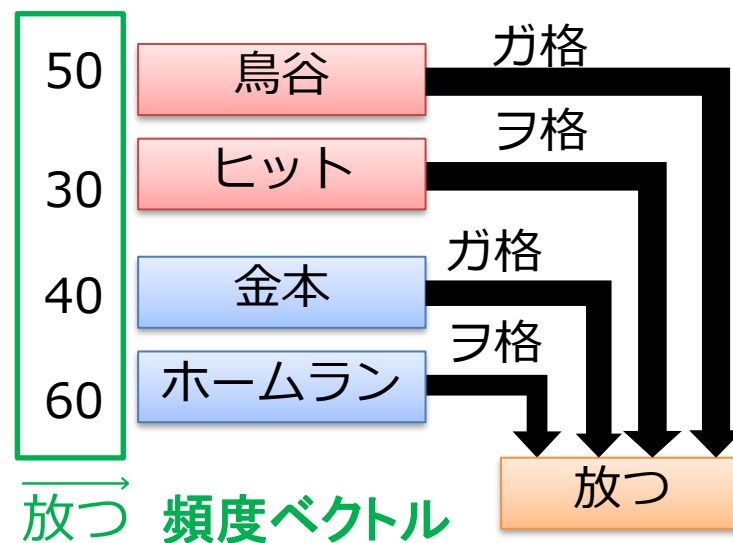
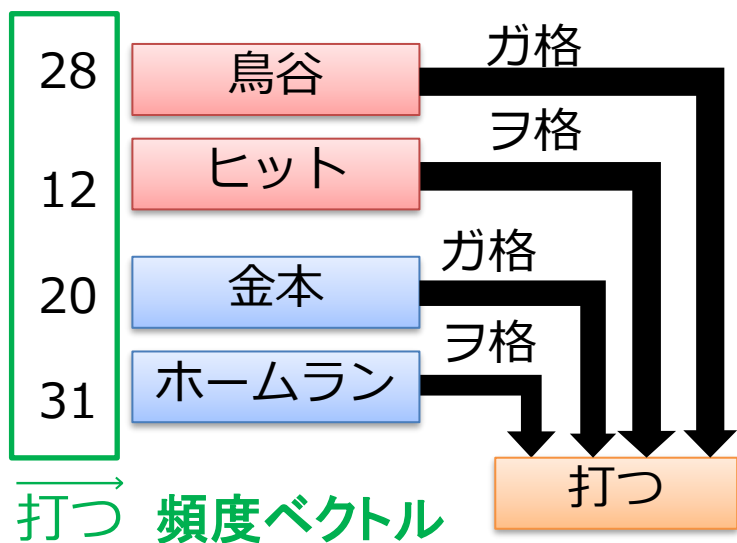
語義から文意へ

- **文中の任意の部分に対して分布仮説が成り立つ**
 - 文の順序それ自体も意味を持つ
 - **Next sentence prediction の意味！**
 - 代名詞などの意味も取ることができるはず
 - 照応解析にも使える？

- **分布仮説は談話を跨ぐことはない**
 - 談話単位の順序に意味があるケースではそういう場合もあるのでは？

NLPにおける分布仮説

- **Statistical semantics (=Distributional Semantics; 分散意味論)**
 - (多くの場合分布仮説に基づいて文脈情報を利用し) 語の意味を捉える、語の類似度を捉えることが試みられてきた
 - 分布仮説は構文構造上でも成り立つことが知られる



Yoshino et al., Proc. SIGDIAL, 2011

W2Vの偉いところ

- **分布仮説に基づく Skip-gram のタスクを設定し
単語予測の問題に落とし込んだ**
 - 良い学習目標を考えた
 - データ量で殴るのに識別タスクはよい
- **単語ベクトルを低次に落としとして固定長にした**
 - 扱いやすい、スパースネスの解消
- **容易に使えるツールの整備**
 - みんなが使う

より詳細なところは PFN 海野さんの
「Statistical Semantic入門 ~分布仮説からword2vecまで~」
がわかりやすいです

まとめ

• 分布仮説の基礎

- 言語の意味は共起語と共起語との相対位置を素性を用いて説明可能
- 単語以外の要素にも分布仮説は成り立つ
- 厳密なルールではなく確率的に捉えることが有効
- 周辺語の分布の偏りにより、類似点だけでなく、
どういう点で異なるかについても見る事が可能

• 分布仮説の実用上の問題点と解決策

- スパーシティの問題 → NNに与えるデータ数で殴る
- 分布仮説をどう実装するかは自明ではなかった
→ 学習データが十分な場合識別的アプローチが有効なので
周辺語を予測する言語モデリングタスクとして定式化