

ホログラフ埋め込みに基づく CCG構文解析

持橋大地

統計数理研究所

daichi@ism.ac.jp

NAIST データサイエンス特別講義

2023-10-13 (金)

※ 立命館大学 山木さん(D1)・谷口忠大教授との共同研究
“Holographic CCG Parsing”, ACL 2023 (main, oral presentation)

自己紹介

- 2005年 NAIST博士修了
 - 1998年 東大教養学部
基礎科学科第二卒業
- 2003年～2011年
ATR音声言語コミュニケーション研究所、
NTTコミュニケーション科学基礎研究所
- 2011年～ 統計数理研究所 (東京・立川市)
- 専門：統計的自然言語処理、機械学習



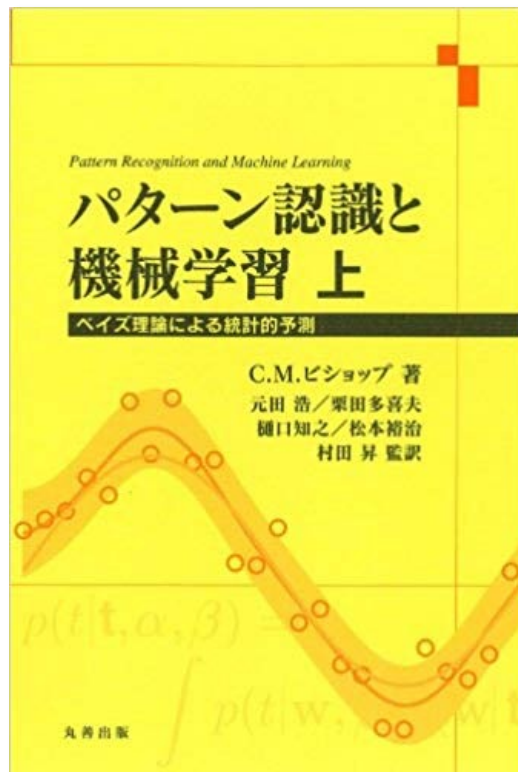
教科書 (1)



- 『ガウス過程と機械学習』
講談社機械学習プロフェッショナルシリーズ(MLP),
2019年発売
 - 持橋大地・大羽成征著
(二人ともNAIST出身)
 - Amazonレビュー122件
- 線形回帰モデルの非常にやさしい導入から入っています

教科書 (2)

- 『パターン認識と機械学習』 翻訳チーム (2006年頃)
 - 10章(下巻)の変分ベイズ法の翻訳を担当
 - 今でも機械学習の最も標準的な教科書

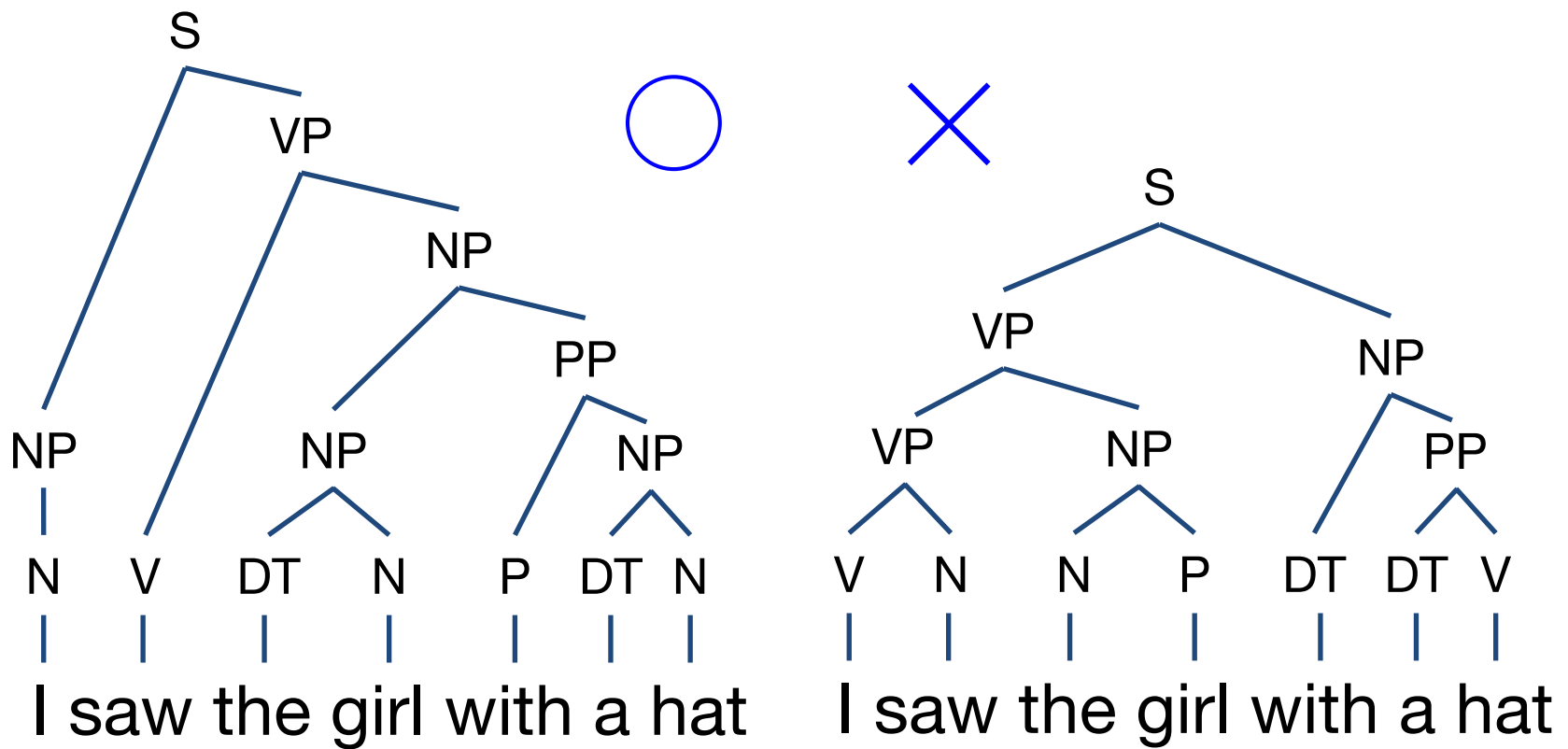


今日の話の概要

- 構文解析とは / その必要性
- CFG (文脈自由文法) と CCG (組み合わせ範疇文法)
- ニューラル構文解析
- ホログラム埋め込み (Nickel+ 2016) とその性質
- ホログラム埋め込みに基づく CCG 構文解析 (Hol-CCG)
- 実験結果
- 複素埋め込み (Trouillon+ 2016) との等価性と解釈
- 画像生成を用いたテキスト理解

構文解析とは

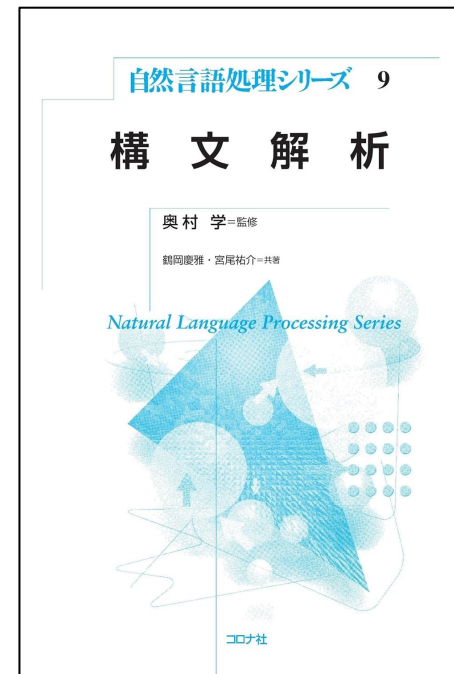
- 言語の文を見て、構文木に変換するタスク
 - 自然言語処理のかつての花形分野



構文解析の難しさ

- 動的計画法を使っても、文長 N について $O(N^3)$ の計算時間 (CYKアルゴリズム)
- 与えられた文に対して可能な膨大な木構造の中から、正解の一つだけを選ぶタスク

	Book	the	flight	through	Houston
S, VP, Verb, Nominal, Noun			S, VP, X2		S, VP, X2
[0,1]	[0,2]	[0,3]	[0,4]	[0,5]	
	Det	NP		NP	
	[1,2]	[1,3]	[1,4]	[1,5]	
		Nominal, Noun		Nominal	
		[2,3]	[2,4]	[2,5]	
			Prep	PP	
			[3,4]	[3,5]	
				NP, Proper- Noun	
				[4,5]	



詳しくは
こうした
教科書を
参照

構文解析は必要？

- かつては、構文解析をしなければ言語のまともな理解は不可能だと思われていた
 - 形態素解析に次ぐ、NLPの**基礎ステップ**
- 現代のニューラル言語モデルは、**明示的に構文解析をしなくても言語を理解できる**
 - 内部的には、構文構造をかなり正確に捉えている
 - ニューラル言語モデルのとらえている構文構造を取り出すことは、**部分的には可能**

構文解析は必要? (2)

- ロボットへの指示文を構文解析して、論理形式に変換
 - Matuszek&Fox (Experimental Robotics, 2013)
- 解析しないと、とんでもない動作をする危険性がある
(「雰囲気NLP」は怪しい)

go to	the	second	junction	and	go left
S/NP	NP/NP	NP/N	N	$S \setminus S/S$	S
<i>(move-to forward)</i>	<i>[null]</i>	<i>(do-n-times 2 x)</i>	<i>(until (junction current-loc) y)</i>	<i>(do-seq g f)</i>	<i>(turn-left)</i>
		NP		$S \setminus S$	
		<i>(do-n-times 2 (until (junction current-loc) y))</i>		<i>(do-seq g turn-left)</i>	
		NP			
		<i>(do-n-times 2 (until (junction current-loc) y))</i>			
		S			
		<i>(do-n-times 2 (until (junction current-loc) (move-to forward)))</i>			
			S		
			<i>(do-seq (do-n-times 2 (until (junction current-loc) (move-to forward))) (turn-left))</i>		

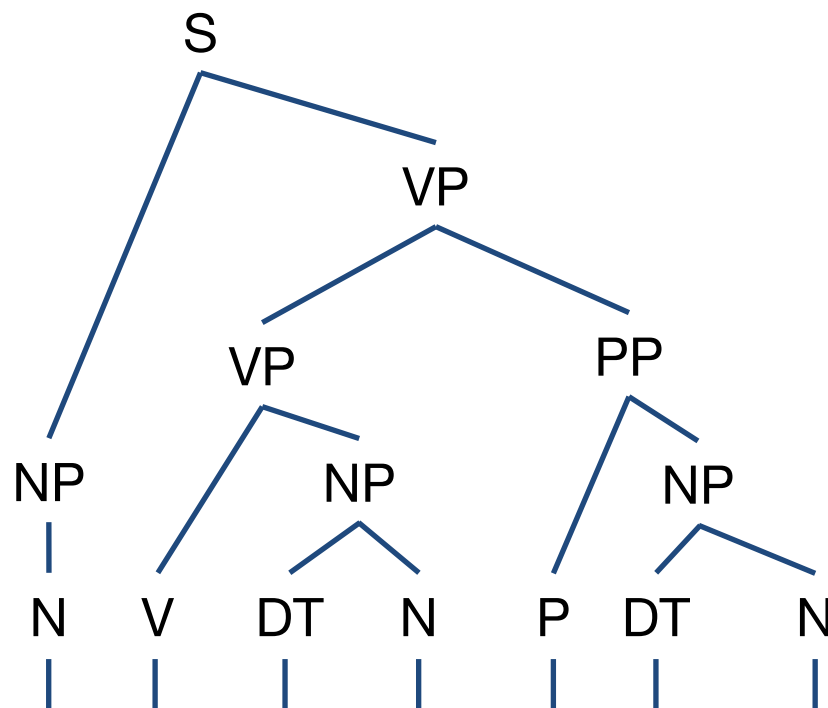
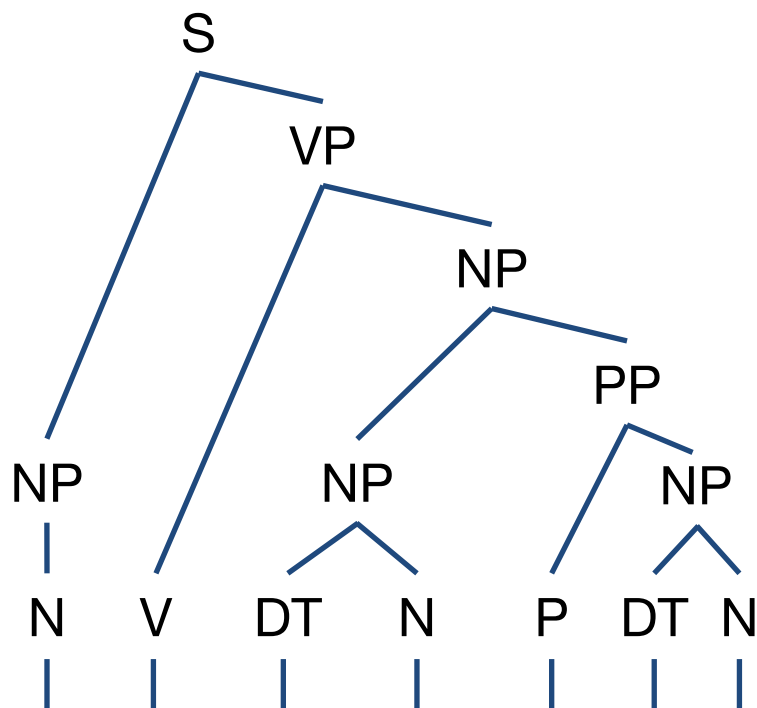
言語の構造をとらえる

- 言語の単語列は本来線形ではなく、明らかに構造がある

きのう見た猫が隣を横切ったのは五時
寝る前には毛布をかけるようにして下さい

- 言語の構造を捉える方法：
 - 構文解析 (PCFG・CCG・依存構造解析)
 - Semantic Parsing
直接、意味的關係を文から抽出する
(ただし、そのために通常は構文解析を利用)

PCFG (確率的句構造文法)



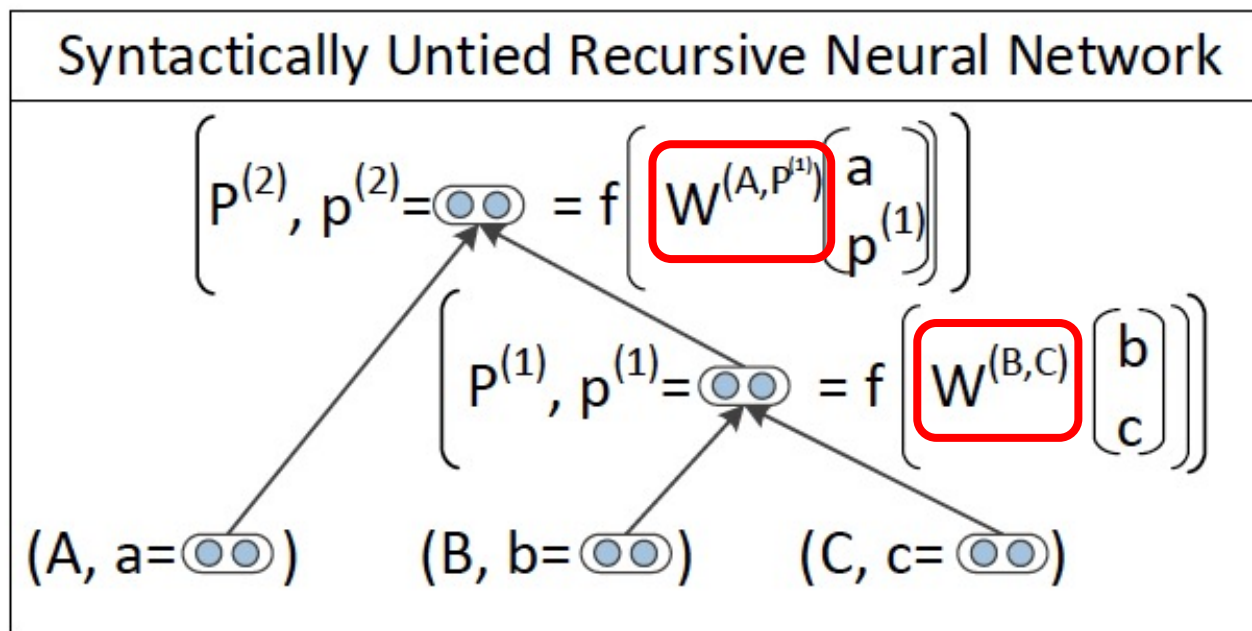
I saw the girl with a hat I saw the girl with a telescope

- 句構造規則：DT N \rightarrow NP，P NP \rightarrow PP，NP PP \rightarrow NP，
V NP \rightarrow VP，NP VP \rightarrow S，VP PP \rightarrow VP
- N: 名詞, DT: 冠詞, P: 前置詞, NP: 名詞句, VP: 動詞句

PCFGの問題

- タグに文法的な意味がない
 - PP → in, with, on, at, ... の確率が高くても、それらが名詞を補語としてとることは言えない (計算機からみると、PPは単なる数字)
 - 同様に、動詞は主語と場合によって目的語をとり文の主辞である、ことはわからない
- 教師なし学習する時、 $Y \rightarrow X_1 X_2$ に対応して infinite PCFGでは $\infty \times \infty$ の確率行列が必要

Compositional Vector Grammar (CVG)



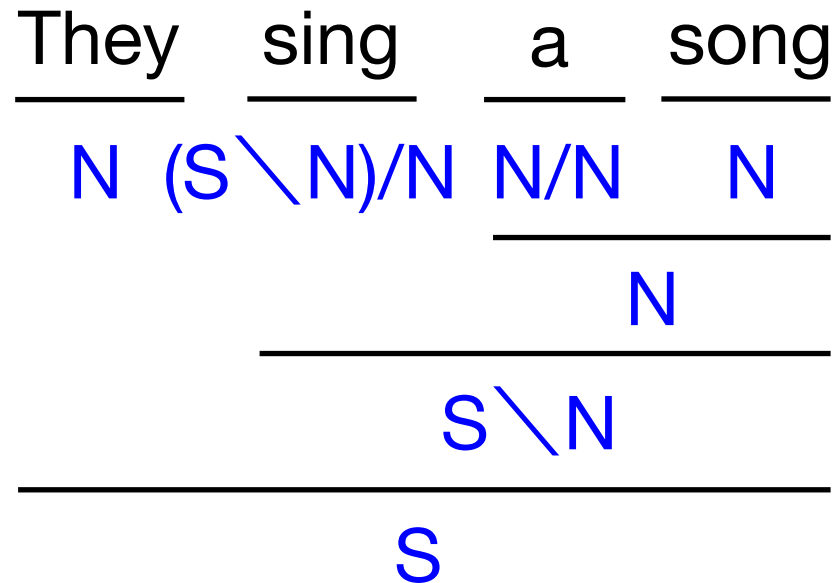
- Socher+ (2013)で提案、構文に沿ったベクトル合成
- ベクトルの合成 $A \rightarrow B C$ に、それぞれ行列 W_{BC} が必要
→ 膨大なパラメータ！ (ex. $500 \times 500 = 250000$ 個の要素/ルール)

組み合わせ範疇文法 (CCG)

- カテゴリとして、文 S と名詞 N しか仮定しない
 - 先験的な“動詞”のような概念を仮定しない
- 例：
 - “the” の意味: the box のように名詞を右にとって、全体で名詞になる $\rightarrow N/N$ (冠詞)
 - “sing” の意味: we sing のように名詞を左にとって、全体で文になる $\rightarrow S \setminus N$ (動詞)
 - “take” の意味: take them のように名詞を右にとって、全体で $S \setminus N$ (=動詞)になる $\rightarrow (S \setminus N)/N$ (他動詞)
 - など

CCGの解析例

- “They sing a song” をCCGで解析する例



- willのような助動詞は動詞句 $(\text{S} \setminus \text{N})$ を右にとって、
また動詞句となるので $(\text{S} \setminus \text{N})/(\text{S} \setminus \text{N})$ になる
など実際はもっと複雑

Holographic CCG

- 基本的なアイデア: CCGをベクトル合成と組み合わせられないか？
- CVGのように合成毎にパラメータがあると、膨大なパラメータが必要で過学習してしまう



- 何か、2つのベクトルを合成する見通しのよい数学的な方法はないか？
→ **Holographic composition** (Plate 1995)

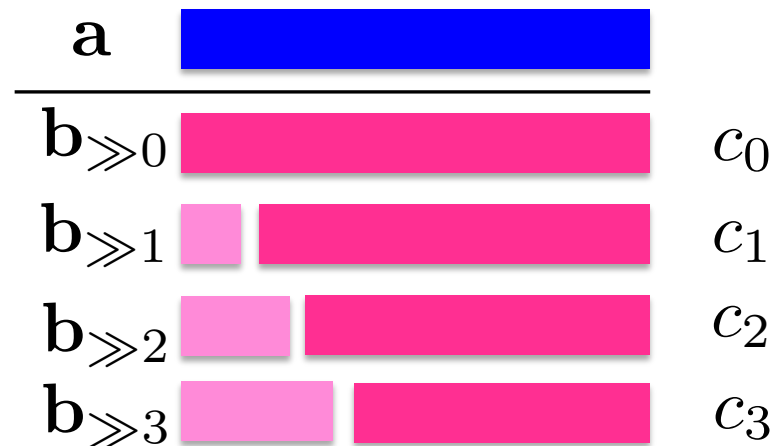
Holographic composition

- 知識グラフの埋め込み手法として、Nickel+(AAAI 2016)で提案 (最初の提案はPlate(1995))
- Circular correlation (巡回相関)

$$\mathbf{c} = \mathbf{a} \star \mathbf{b}$$

- 定義:

$$[c]_k = \sum_{i=0}^{d-1} a_i b_{(k+i) \bmod d}$$



Holographic composition (2)

- 注: 巡回相関は実際には、高速フーリエ変換と逆変換を使うことで、以下で計算できる

$$\mathbf{a} \star \mathbf{b} = \mathcal{F}^{-1}(\overline{\mathcal{F}(\mathbf{a})} \odot \mathcal{F}(\mathbf{b}))$$

- 計算量 $O(d) \rightarrow O(d \log d)$
- 実は、戻さずに最後まで周波数空間で考えてもよい (最後に議論します)
- \odot は要素ごとの積 (アダマール積)

Holographic composition (3)

- 性質:

- Non-commutative: 方向を保存

$$\mathbf{a} \star \mathbf{b} \neq \mathbf{b} \star \mathbf{a}$$

- Non-associative: 構造を保存

$$(\mathbf{a} \star \mathbf{b}) \star \mathbf{c} \neq \mathbf{a} \star (\mathbf{b} \star \mathbf{c})$$

- 構文構造を保存: たとえば “loves to eat” のベクトルは

$$\mathbf{V}_{\text{“loves to eat”}} = \mathbf{V}_{\text{loves}} \star \mathbf{V}_{\text{“to eat”}} = \mathbf{V}_{\text{loves}} \star (\mathbf{V}_{\text{to}} \star \mathbf{V}_{\text{eat}})$$

- これは $(\mathbf{V}_{\text{loves}} \star \mathbf{V}_{\text{to}}) \star \mathbf{V}_{\text{eat}}$ とは異なる

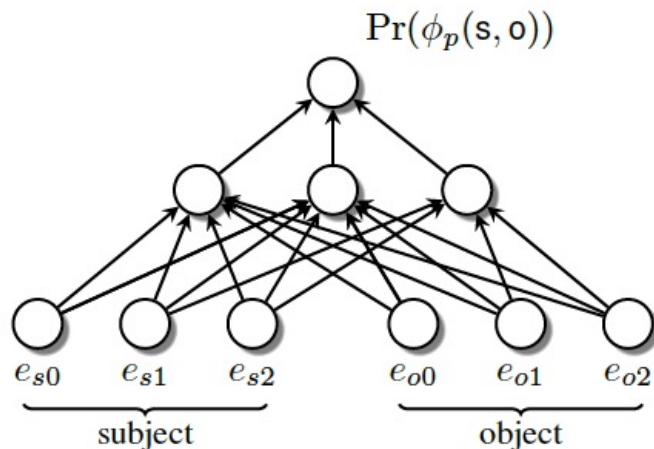
知識グラフでの使用

(Nickel+ 2016)

- 知識グラフ: 主語sと目的語oが関係rにあるかどうか
 - 例: s=“Obama”, o=“Hawaii”, r=“born in”
- そうなる確率:

$$p(\phi_r(s, o) = 1 | \Theta) = \sigma(\mathbf{e}_r^T (\mathbf{e}_s \star \mathbf{e}_o))$$

- sとoの埋め込みベクトルのホログラム合成に対するロジスティック回帰



(b) HOLE



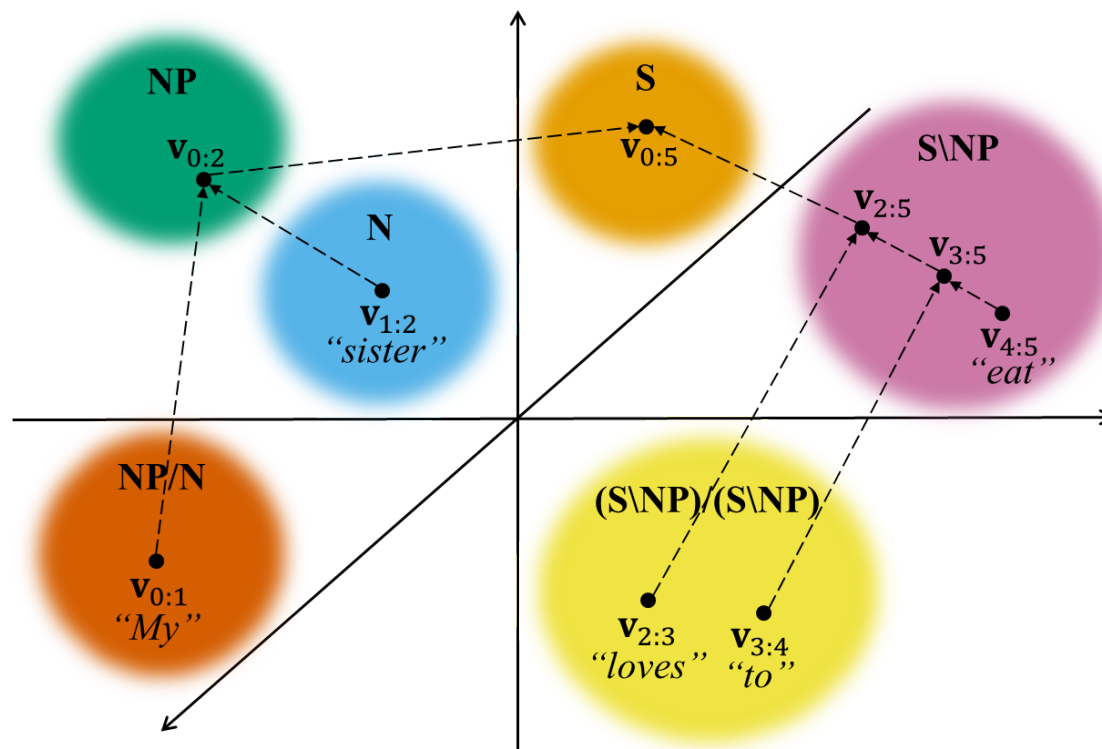
知識グラフでの使用 (2)

- WordnetおよびFreenet知識グラフでの、学習データにない関係の予測精度
 - 他の手法をはるかに超えて、きわめて高い精度で予測可能

Method	WN18					FB15k				
	MRR		Hits at			MRR		Hits at		
	Filter	Raw	1	3	10	Filter	Raw	1	3	10
TRANSE	0.495	0.351	11.3	88.8	94.3	0.463	0.222	29.7	57.8	74.9
TRANSR	0.605	0.427	33.5	87.6	94.0	0.346	0.198	21.8	40.4	58.2
ER-MLP	0.712	0.528	62.6	77.5	86.3	0.288	0.155	17.3	31.7	50.1
RESCAL	0.890	0.603	84.2	90.4	92.8	0.354	0.189	23.5	40.9	58.7
HOLE	0.938	0.616	93.0	94.5	94.9	0.524	0.232	40.2	61.3	73.9

Holographic CCG (2)

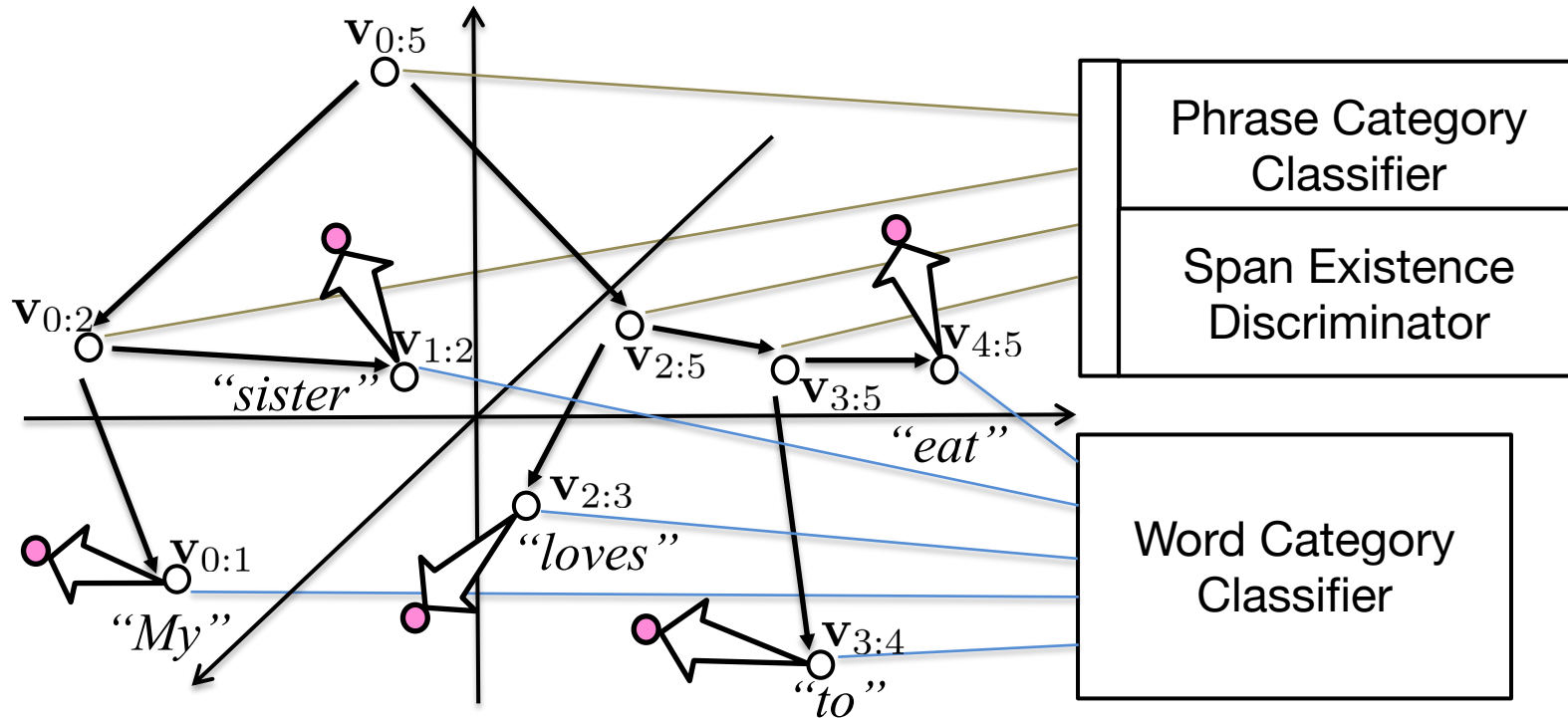
- Hol-CCGのイメージ



- 単語ベクトル自体を最適化して、合成して最終的に文Sのベクトルが得られるようにする

実際の学習の様子

“My sister loves to eat”

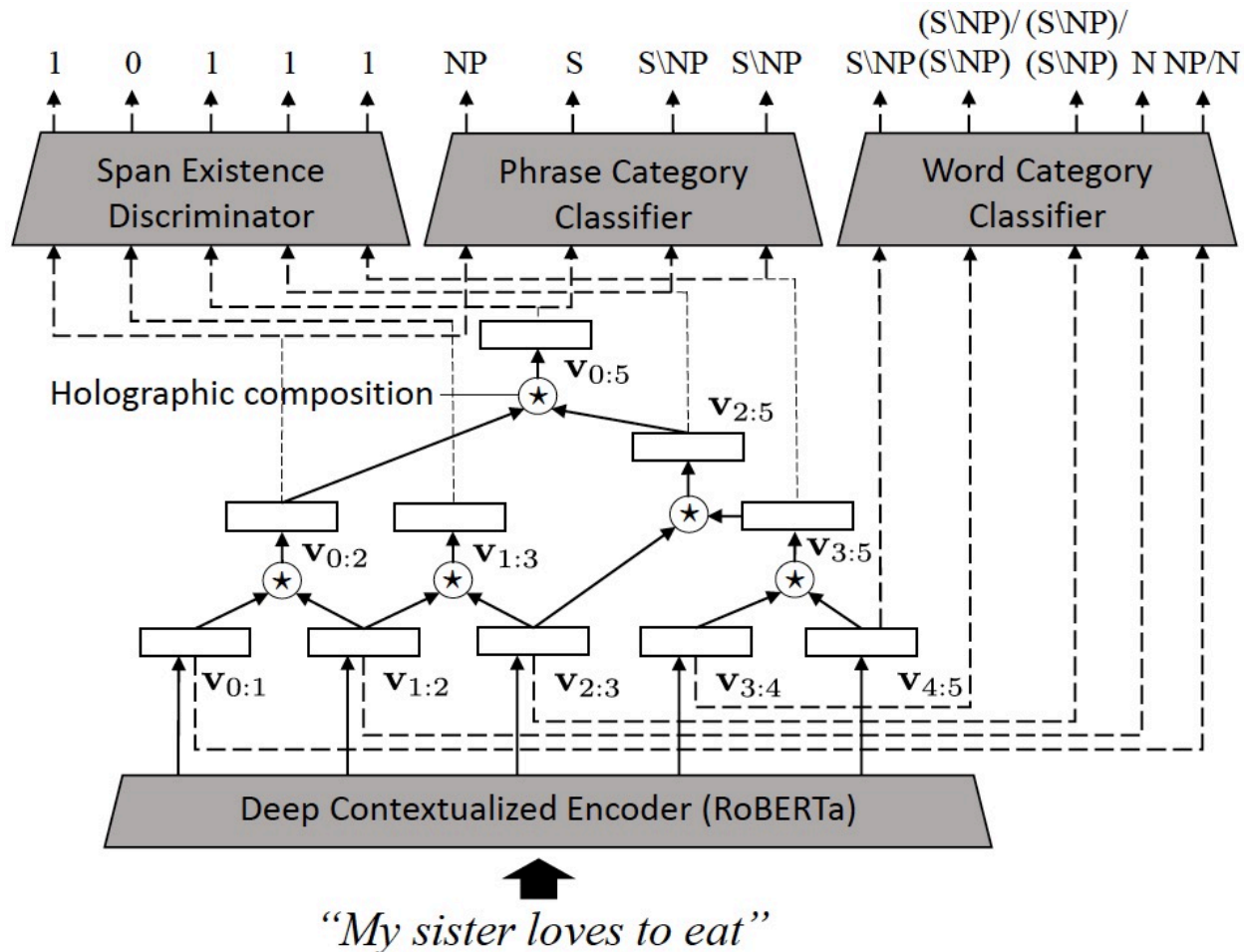


- 合成した結果ベクトルがそれぞれの句に正しく分類されるよう、逆伝播で最初の単語ベクトルを最適化
- 最終的には、文Sとなればよい



Transformerから始める場合

- 各単語の入力埋め込みとしてRoBERTaを使用



学習の目的関数

- 単語・句のベクトルをFFNNに通して3種類の確率を計算, 誤差を逆伝搬
 - $P_w(i, i+1)$: 単語に対するカテゴリ割当の確率分布 (supertagging) $\rightarrow \mathcal{L}_w$
 - $P_p(i, j)$: スパン(句)に対するカテゴリ割当の確率分布 $\rightarrow \mathcal{L}_p$
 - $P_s(i, j)$: スパン自体の存在確率 $\rightarrow \mathcal{L}_s$
- 単語ベクトルを使うより、BERTの埋め込みを使う方が高性能：
単語埋め込みの最適化 \rightarrow 「BERT埋め込みの変換」の最適化
(計算にはRoBERTaを使用)

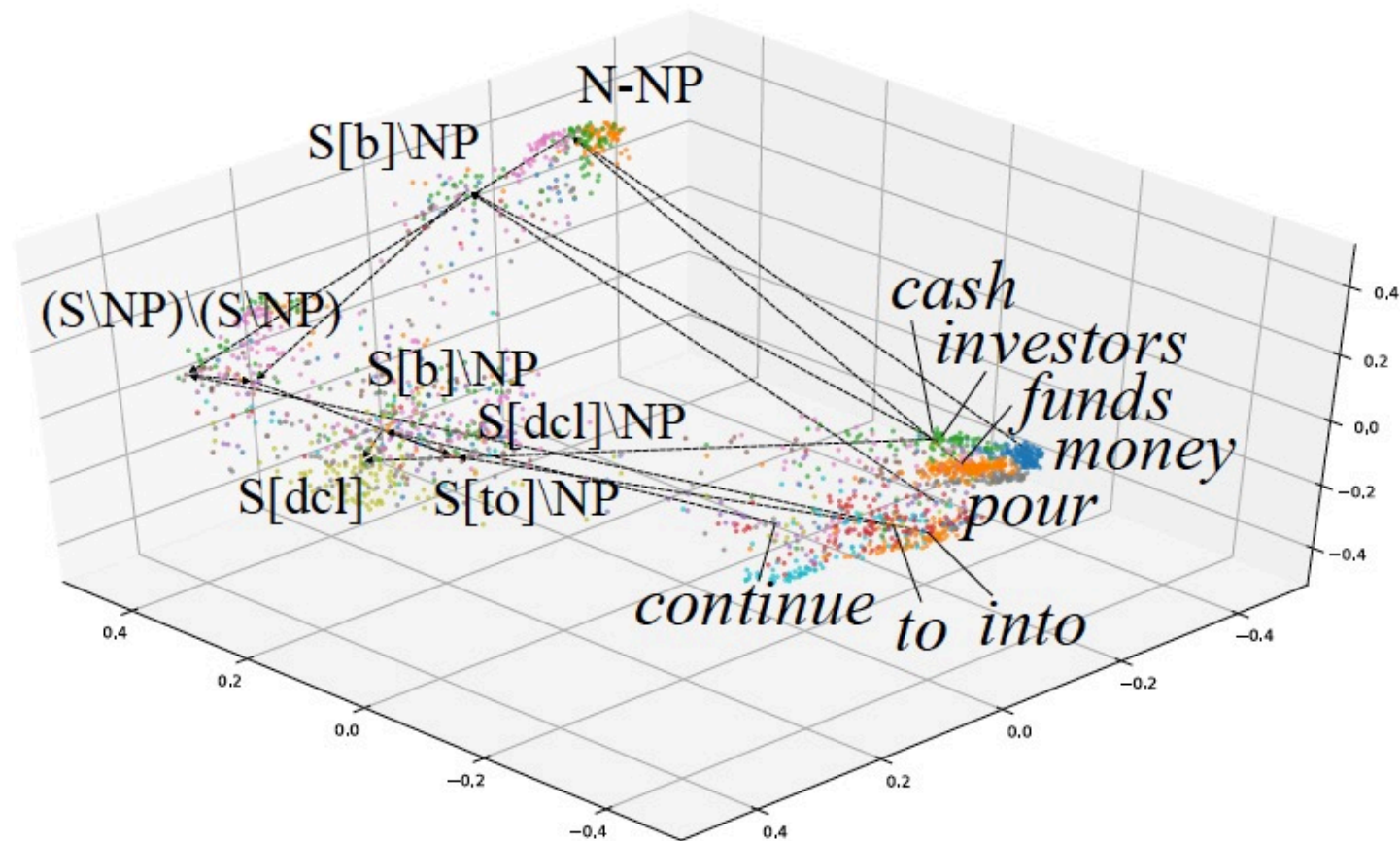
実験結果

- CCGBankというWall Street Journalのデータで評価
- 現在、ヒューリスティックな方法を抜いて世界最高性能

Model	Super-Tagger	Parser	Acc	LF
Lewis et al. (2016)	LSTM	A*	94.7	88.1
Vaswani et al. (2016)	LSTM	C&C	94.5	88.32
Yoshikawa et al. (2017)	LSTM	A* (LSTM)	–	88.8
Stanojević and Steedman (2020)	LSTM	Shift-Reduce (LSTM)	–	90.6
Tian et al. (2020)	Attentive-GCNN	EasyCCG	96.25	90.58
Bhargava and Penn (2020)	LSTM decoder	C&C	96.00	90.9
Liu et al. (2021)	Category Generator	C&C	96.05	90.87
Prange et al. (2021)	Tree-Structured decoder	C&C	96.22	90.91
Kogkalidis and Moortgat (2022)	Heterogeneous Dynamic Convolutions	–	96.29	–
Clark (2021)	Tian et al. (2020)	C&C	–	91.9
		Span-based	–	92.9
Ours ($\mathcal{L}_w + \mathcal{L}_p + \mathcal{L}_s$, Real)	Holographic	C&C	96.60	92.12
		Span-based	–	92.67

Table 2: Comparison of the proposed model and existing methods; best results are shown in bold.

実際のベクトルの合成の様子



- 768次元のベクトル空間をPCAで3次元に可視化

文法に即した穴埋め

- 「句」のベクトルから、ランダムに単語に戻すことが可能 → 文法に準拠した穴埋めが可能に

ID	Sentence	Replacement Candidate	Sim.	NPMI
1	<i>Mr. Vinken is chairman of Elsevier N.V. , the Dutch publishing group .</i>	<i>Mr. Baris</i>	1.00	0.19
		<i>Dr. Novello</i>	1.00	0.10
		<i>Ms. Ensrud</i>	1.00	0.11
2	<i>When Scoring High first <u>came out</u> in 1979, it was a publication of Random House .</i>	<i>turned up</i>	0.94	0.27
		<i>sold out</i>	0.91	0.29
		<i>sells out</i>	0.90	0.24
3	<i>In early trading in Hong Kong Thursday , gold was quoted <u>at \$ 374.19 an ounce</u> .</i>	<i>for \$ 25.50 a share</i>	0.94	0.33
		<i>for \$ 60 a bottle</i>	0.94	0.29
		<i>at \$ 51.25 a share</i>	0.93	0.34
4	<i>Judges are not getting <u>what they deserve</u> .</i>	<i>what she did</i>	0.96	0.28
		<i>what they do</i>	0.96	0.36
		<i>what we do</i>	0.89	0.35
5	<i><u>Despite recent declines in yields</u> , investors continue to pour cash into money funds .</i>	<i>Despite the flap over transplants</i>	0.89	0.22
		<i>In a victory for environmentalists</i>	0.86	0.22
		<i>On the issue of abortion</i>	0.82	0.27
6	<i>Despite recent declines in yields , investors continue <u>to pour cash into money funds</u> .</i>	<i>to provide maintenance for other manufacturers</i>	0.83	0.27
		<i>to share data via the telephone</i>	0.79	0.21
		<i>to cut costs throughout the organization</i>	0.77	0.26

CCGとλ計算

- CCG構文解析により、λ式による意味計算が可能になる
- ccg2lambda: Pascual+ (ACL2016), Tanaka+ (NLP2016) など

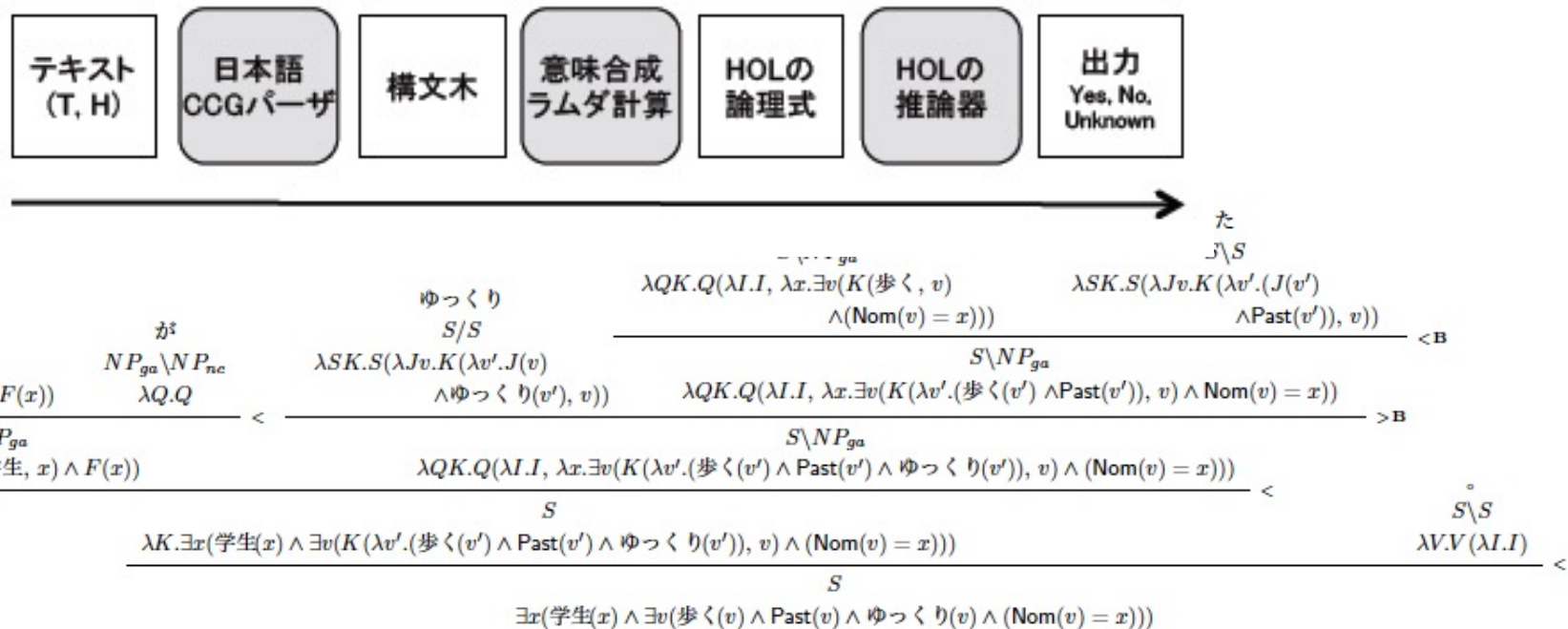


図 4: 「学生がゆっくり歩いた。」の導出木と意味表示.

複素埋め込みとの関係

- 巡回相関は、複素埋め込み (Trouillon+, ICML 2016) と数学的に等価 (Hayashi+, ACL 2017)
 - フーリエ変換して、すべて周波数空間で計算してもよい

$$\mathbf{a} \star \mathbf{b} = \mathcal{F}^{-1}(\overline{\mathcal{F}(\mathbf{a})} \odot \mathcal{F}(\mathbf{b}))$$

- 入力を先にFFTで変換しておけば、複素共役 + 単なる要素ごとの積でOK
 - $\alpha = \mathcal{F}(\mathbf{a})$, $\beta = \mathcal{F}(\mathbf{b})$ とおけば、

$$\mathcal{F}(\mathbf{a} \star \mathbf{b}) = \overline{\alpha} \odot \beta$$

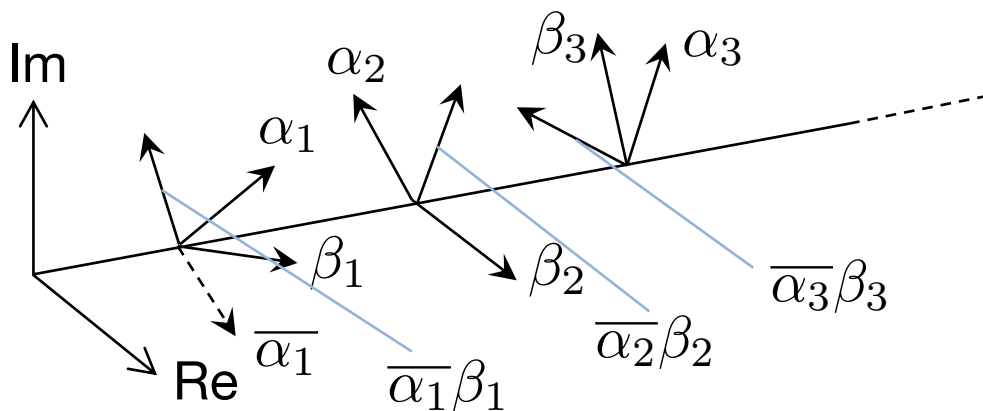
複素埋め込みとの関係 (2)

$$\mathcal{F}(\mathbf{a} \star \mathbf{b}) = \bar{\alpha} \odot \beta$$

- $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_D), \beta = (\beta_1, \beta_2, \dots, \beta_D)$ とおくと、

$$\bar{\alpha} \odot \beta = (\bar{\alpha}_1\beta_1, \bar{\alpha}_2\beta_2, \dots, \bar{\alpha}_D\beta_D)$$

- 各複素数 α_k, β_k のノルムを1にすれば、ベクトル合成は周波数空間で、高次元角度ベクトルを次元毎に回転していることに相当する！(検討中)



画像解析/生成への応用

(NLP2023 優秀研究賞)

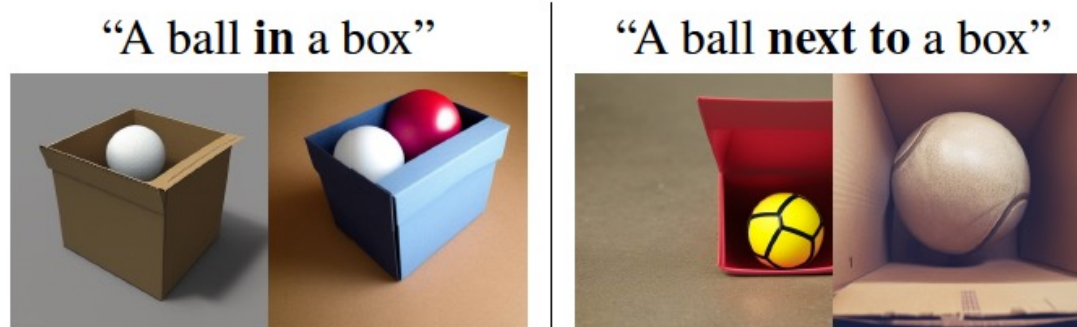
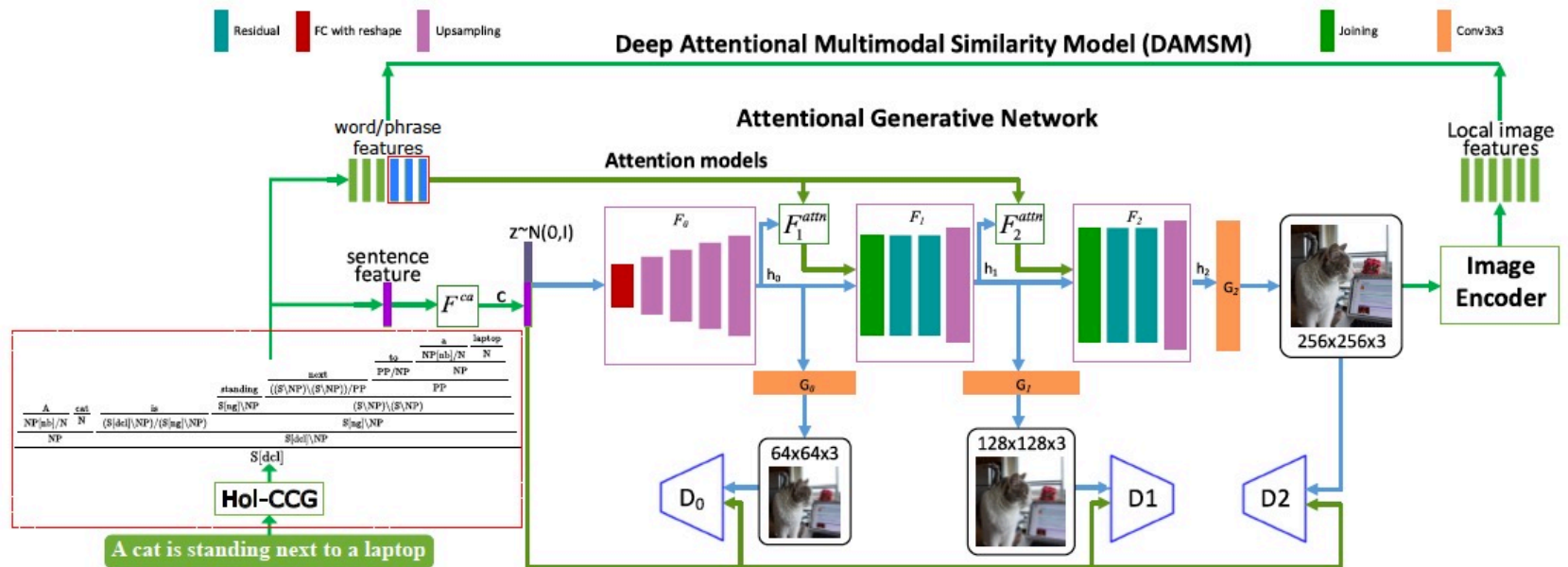


図 1: Stable Diffusion が生成した画像の例.

- 問題: 今の画像処理は、画像のキャプション文の文法的な構造を明示的にとらえていない
 - 「雰囲気NLP」
 - 工学的な適用を考えると危険性が大きい
- 認知的にみても、人間は画像のような外界情報と言語の構造を同時に学習している

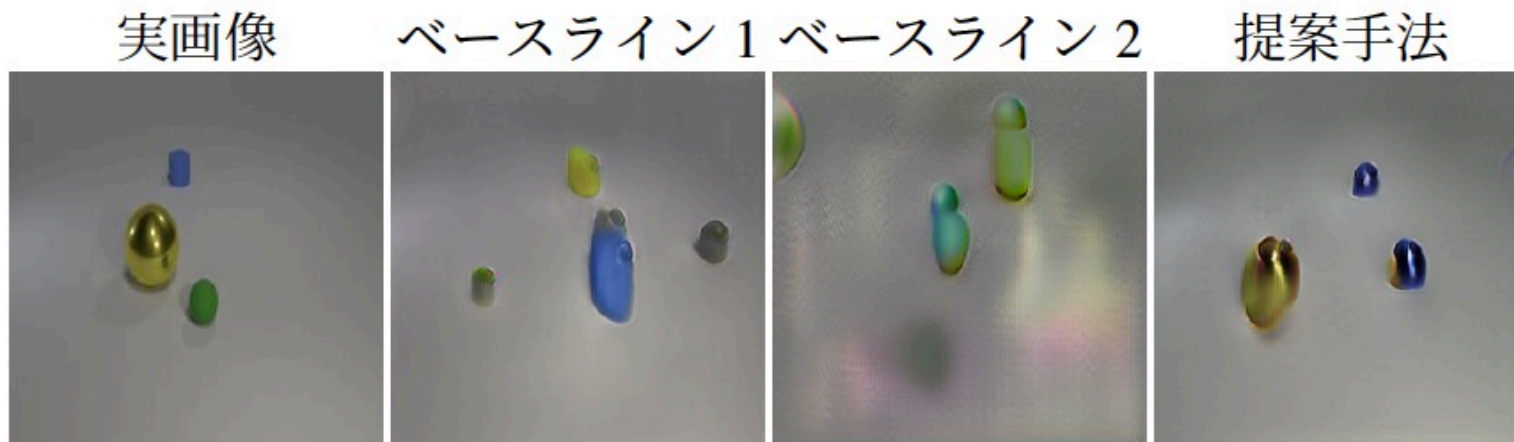
画像解析/生成への応用

- Attention GANによる画像/キャプションの対応モデルを拡張
 - Hol-CCGによる句構造ベクトルを、アテンションで画像と対応させる



実験結果

- 入力テキスト: “large yellow sphere is to the left front of small blue cylinder, and large yellow sphere is to the left behind of small green sphere”



- 提案手法は、構文構造を利用して物体の位置関係 (“to the left front of”) をより正確に再現

実験結果 (2)

表 1: COCO データセットにおける客観評価の結果.

手法	FID↓	R-prec↑
AttnGAN [11]	35.49	85.47
AttnGAN + VICTR [14]	29.26	86.39
ベースライン 1	60.76	87.07
ベースライン 2	39.55	93.59
提案手法	31.29	93.12

- 数値評価でも、ベースラインより性能を改善
- ただし、GANベースなので、Stable Diffusionベースにししないと実際の画像生成の品質は微妙 (現在検討中)

まとめ

- 潜在空間で二つのベクトルを合成する巡回相関 (ホログラフ埋め込み) 演算を再帰的に用いた、Holographic CCG 構文解析を提案
 - C&Cパーザを用いる中で世界最高精度
 - 完全ブラックボックスのTransformerと同程度の性能
- ルール毎に膨大なパラメータが必要だったCVG (Socher+ 2013) のパラメータを不要に
- 材料科学など、ベクトルを再帰的に合成する必要のある分野への応用も期待できる
- 複素埋め込みと等価 = 周波数空間における disentangling (因子分解)