

Deep Learning x NLP

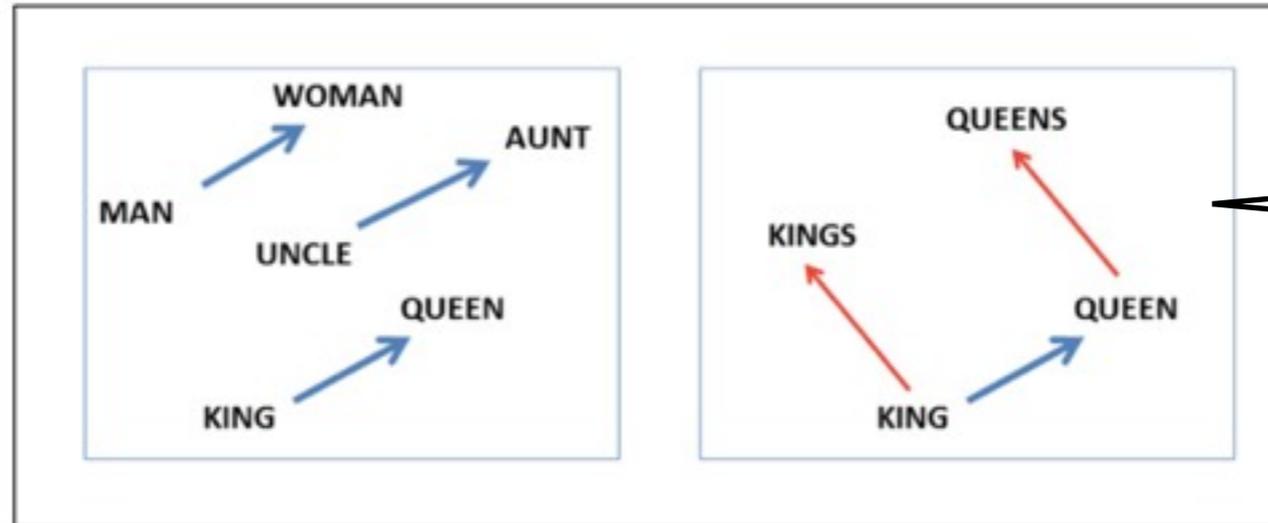
RC-NET: A General Framework for Incorporating Knowledge into Word Representations (CIKM'14)

担当：D2 片岡

Deep Learning x NLP

● この論文の前提知識：“word2vec”

- 以下のAutoencoderを組んでみたら，単語の意味関係を抽出できた！
- イメージだけ伝えると...



いい感じに抽象化されたものを潜在空間上にマッピング

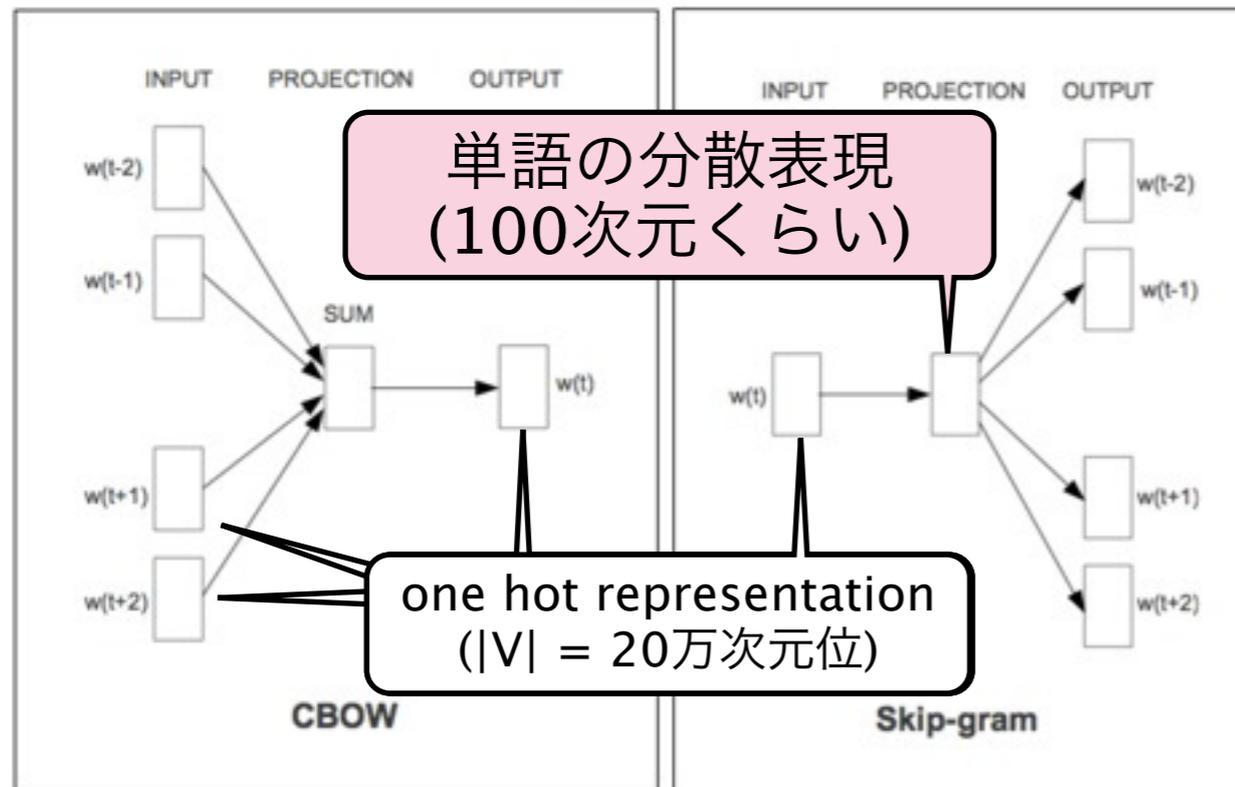


図4-9 Continuous Bag-of-Wordsモデル

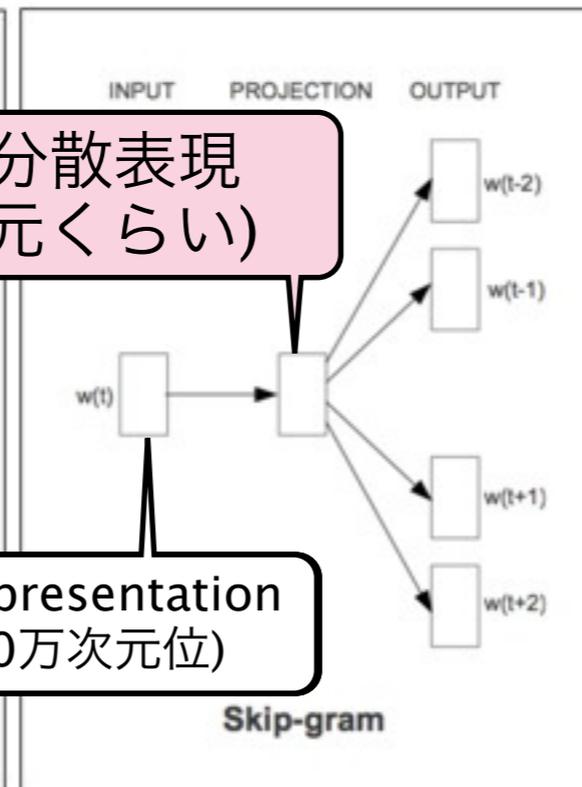


図4-12 Skip-gramモデル¹⁴

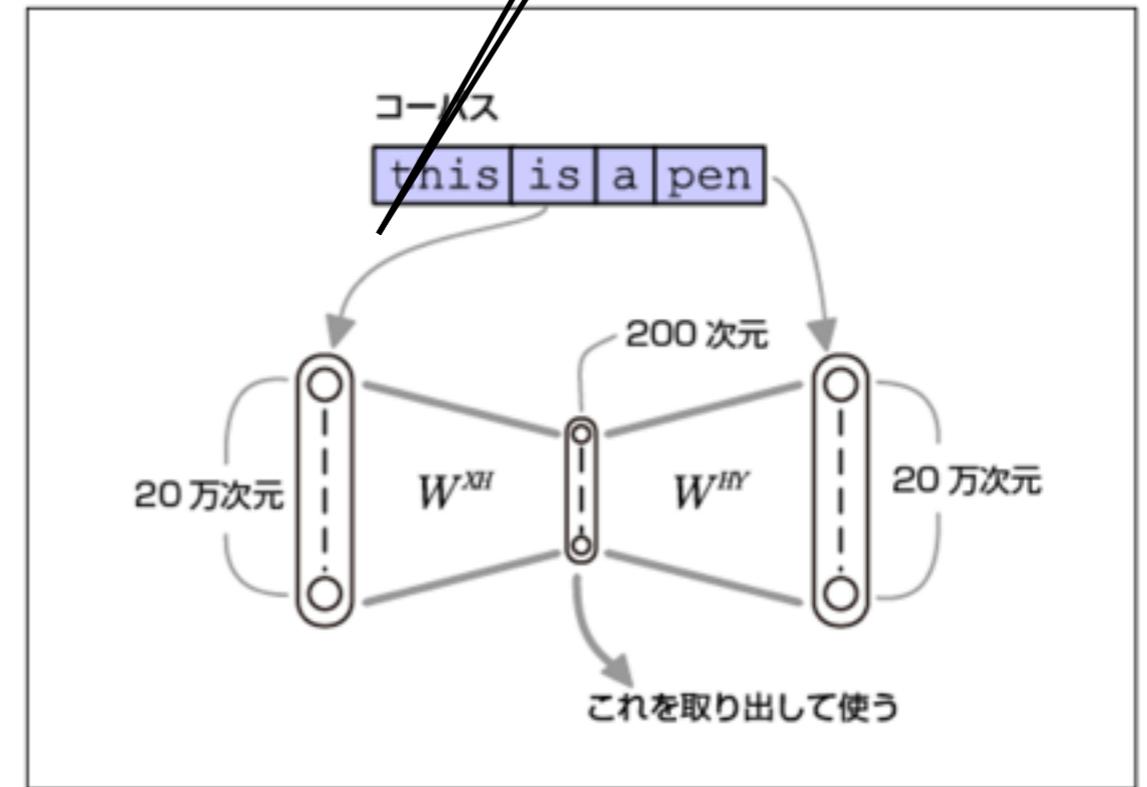


図4-14 シンプルに表現したSkip-gramモデル

“Tomas Mikolov et. al. "Efficient Estimation of Word Representations in Vector Space”

word2vec > Skipgram

● Skipgramモデルの概要

- 入力した単語 w_k を中心として、周囲に出現する単語 $w_{\{k_j\}}$ の予測モデル
- 入力は1-hot representation ($|V|$ 次元のベクトル, 一つ非0の要素を持つ)

● アルゴリズム

- 1, 入力に対して行列 M で次元圧縮($|V|$ 次元 \Rightarrow 100次元くらい)
- 2, 行列 M' で $|V|$ 次元に戻す
- 3, softmaxで予測
- 4, 誤差を行列 M に逆伝播

● 備考

- 全ての単語に対し共通の M 行列を学習
- 単語の分散表現は, $w_k \times M$ で算出
- Mikolov論文はこれを高速化&精度向上

目的関数 L
(最大化する)

$$L = \frac{1}{K} \sum_{k=1}^K \sum_{-N \leq j \leq N, j \neq 0} \log p(w_{k+j} | w_k)$$

$$p(w_{k+j} | w_k) = \frac{\exp(v'_{w_{k+j}} v_{w_k})}{\sum_{w=1}^V \exp(v'_w v_{w_k})}$$

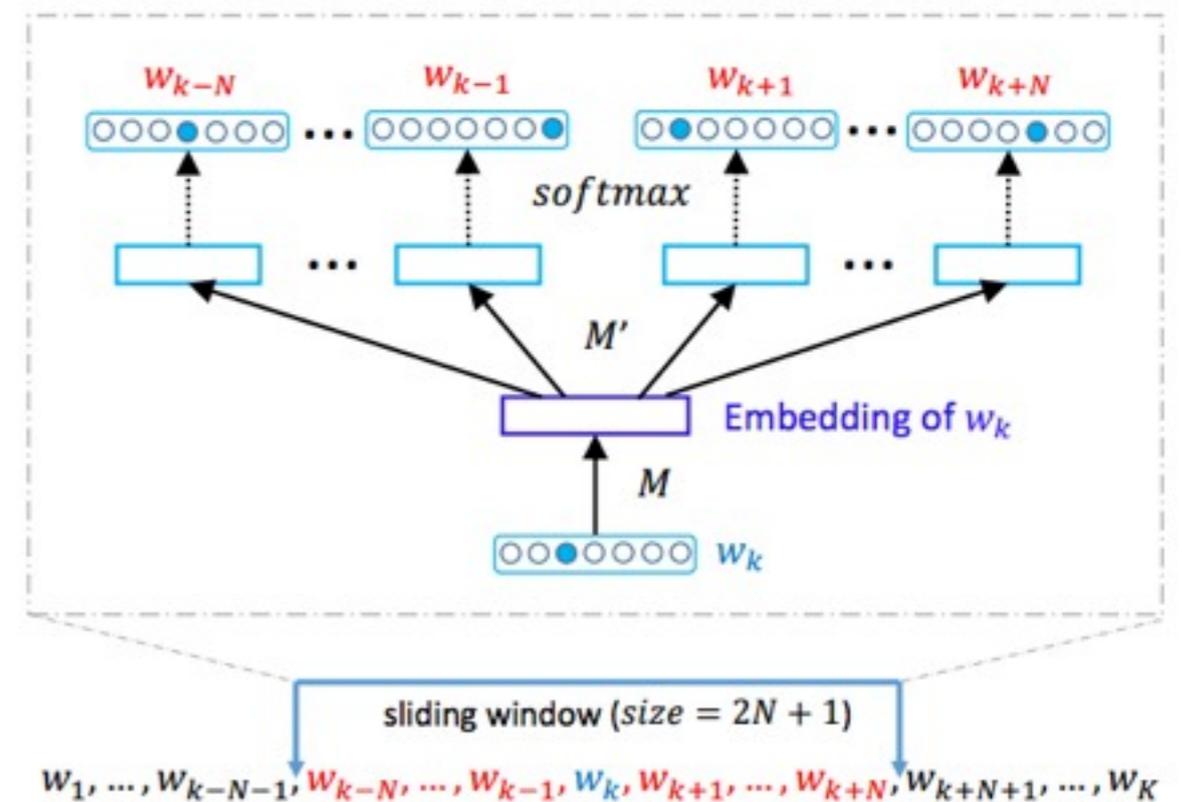


Figure 2: The continuous skip-gram model.

Deep Learning x NLP 疑問

● 単語の意味関係抽出の分野では、既に...

- wordnetとかfreebaseとかでNamed Entityに関する知識ベースはある

● 疑問 1 :

- 既存の知識ベースを”うまく”AEに考慮できれば、精度上がるのでは？

- **本日紹介する論文はこれ**

● 疑問 2 :

- 単語の分散表現(概念ベクトル化)のような話は昔からあるらしいけど何がnew？

- **特に、SVDとかの行列分解と何が違うの？**

- こちらについては、SkipgramがPMIの行列分解と等価であると証明されたい (NIPS'14)

- <http://www.slideshare.net/nttdata-msi/skip-gram-shirakawa20141121-41833306>

● 疑問 3 :

- 単語の意味関係の抽出タスクは研究的には面白いけど、応用例としては面白くない。別の応用例はないの？

- NLP直球勝負ではなくて、この知見を別分野にどう応用するかが勝負すべきポイントな気がしている

本日紹介する論文

● 出典

- Title : RC-NET: A General Framework for Incorporating Knowledge into Word Representations
- Authors : Chang Xu, Yalong Bai, Jiang Bian, [Bin Gao](#), Gang Wang, Xiaoguang Liu, Tie-Yan Liu
- Conference : CIKM'14

● 概要

- Skipgramによって意味的な情報を分散表現で獲得できるようになった
- しかし、単語の属性や単語間の複雑な意味関係の獲得はまだ難しい
- その理由は、文書コーパスには不完全 或いは 曖昧な情報があるから
- そこで、提案手法は文書コーパスにknowledge graphを組み合わせる
- KGの意味関係とカテゴリ情報を利用⇒精度の高い単語の分散表現を取得
- 3種類の実験で単語の分散表現の精度を評価する

背景

● Deep Learning x NLP

- 単語の分散表現の学習がアツい (~ = word embedding, distributed representations of words)
- CBOWやSkipgramが提案され、その動きは加速
- その原理は、同じ意味を持つ単語の周囲には、同じような他の単語が出現することに基づいている。要するに”共起”。

● 従来手法の問題

- ある単語の前後に出現する単語のデータが少ないと機能しない
- データが多くても、文書コーパスだけではノイズやバイアスという問題がある
 - ノイズやバイアスによって、異なる意味に解釈されてしまう。

● 提案手法

- 基本、Skipgramをベースとする
- 外部情報として、確度の高いknowledge graphを利用する
- knowledge graphとは、FreebaseやWordNetがあり、2種類の情報を持つ
 - 1, Relational Knowledge (is-a, part-of, child-of) : 潜在空間上の意味表現 (方向ベクトルの知識)
 - 2, Categorical Knowledge (gender, location) : 潜在的な類似性の表現 (位置ベクトルの知識)
- Skipgramの正則化項にうまくこの2種類の情報を入れこむ

Knowledge Graphの例

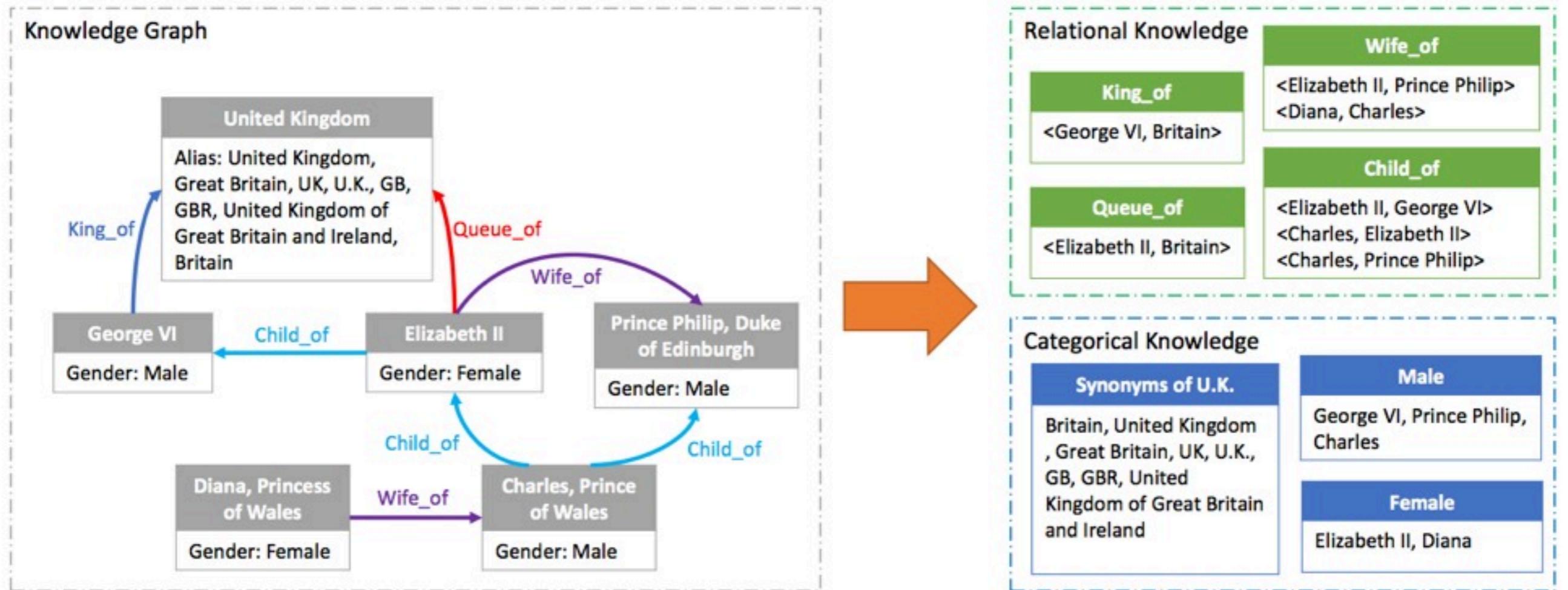
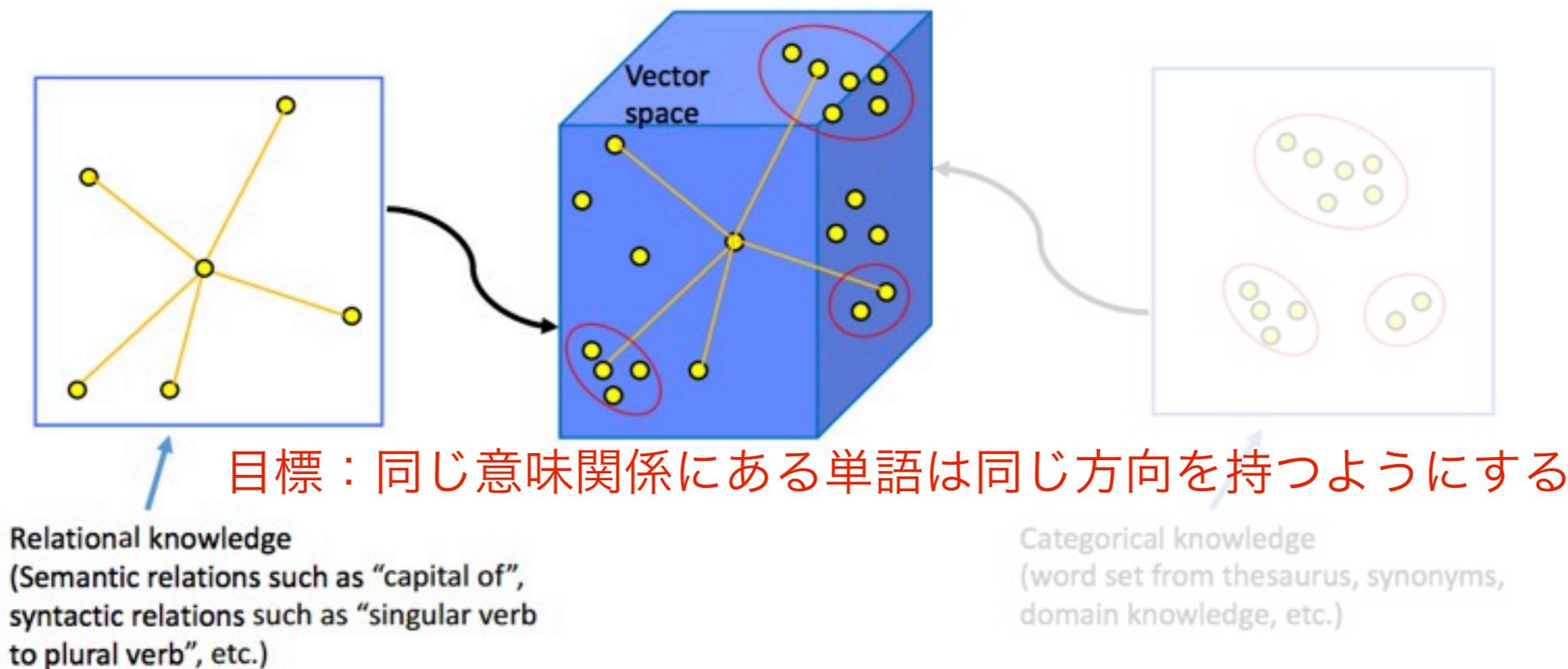


Figure 1: Knowledge graph contains two forms of knowledge: relational knowledge and categorical knowledge.

● 備考

● これらの情報はFreebaseやWordNetでオープンに利用可能

提案手法 R-NET



● Relational knowledge : (head, relation, tail) = (h,r,t) ∈ S

● hとtは単語(位置ベクトル), rは意味ベクトル(方向ベクトル)

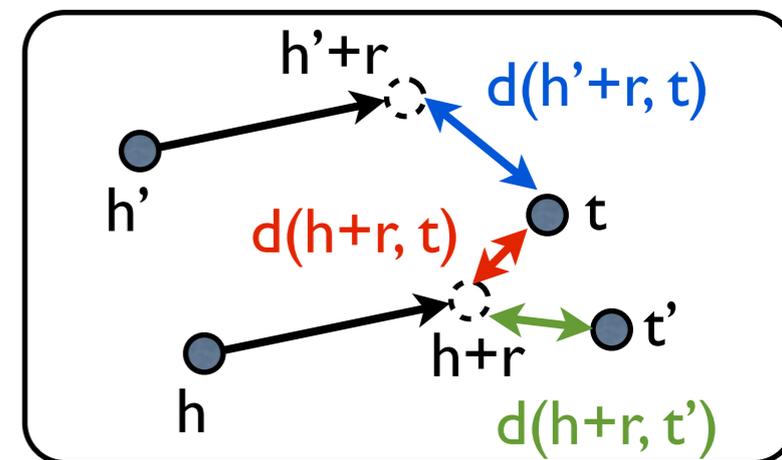
● 最小化する目的関数は次の通り

$$J_r = \alpha E_r - L$$

$$E_r = \sum_{(h,r,t) \in S} \sum_{(h',r,t') \in S'_{(h,r,t)}} \left[\underbrace{\gamma}_{\text{hyper parameter}} + \underbrace{d(h+r,t)}_{\text{知識ベース通りになるよう}} - \underbrace{d(h'+r,t')}_{\text{???}} \right]_+$$

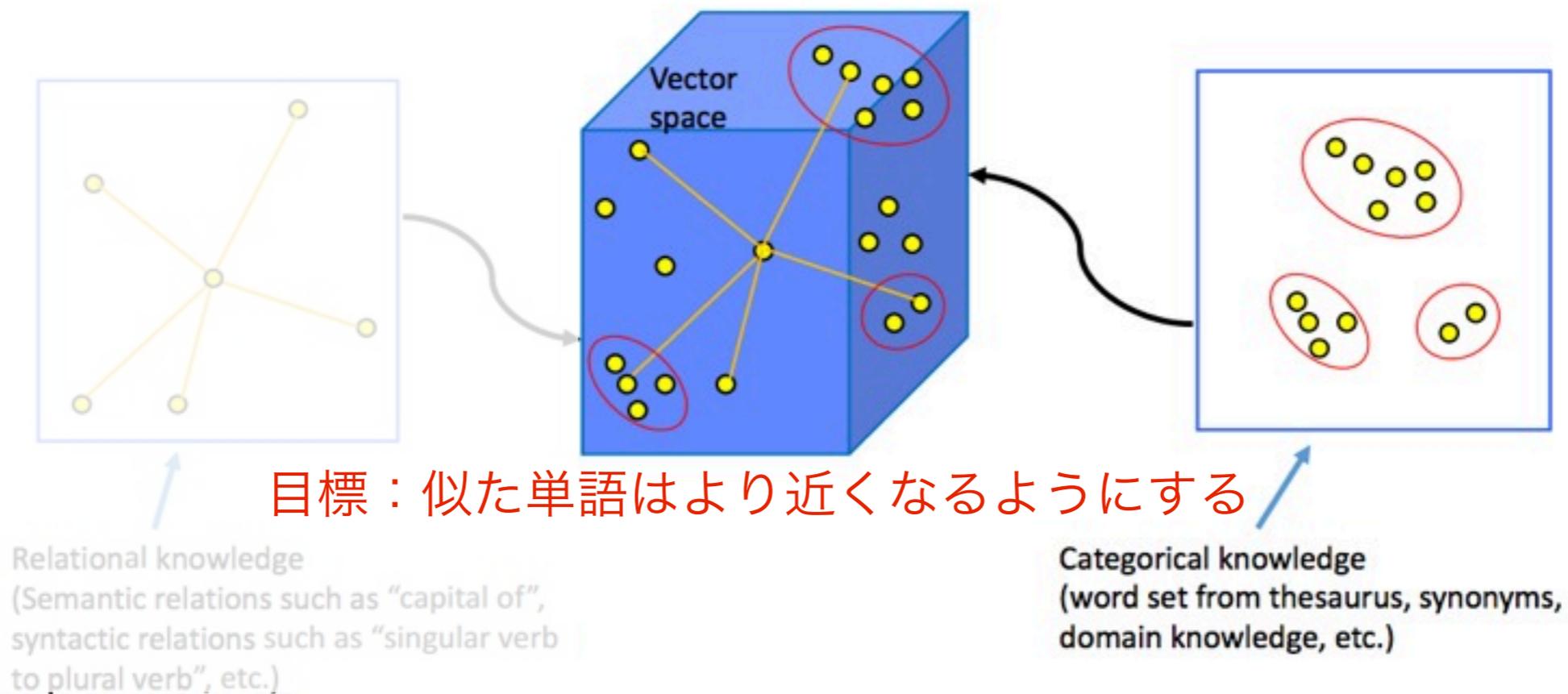
正の値

$$S'_{(h,r,t)} = \left\{ \underbrace{(h',r,t)}_{h': r \text{ と } t \text{ 固定の時の全ての } h} \right\} \cup \left\{ \underbrace{(h,r,t')}_{t': r \text{ と } h \text{ 固定の時の全ての } t} \right\}$$



? rの算出方法は不明.
多分平均...?

提案手法 C-NET



● Categorical knowledge : $s(w_i, w_j) \in$ Similarity Matrix Q

- カテゴリ情報を類似度に変換する
- カテゴリに対してデータが多いと弱い類似度, 少ないと強い類似度
- 最小化する目的関数は次の通り

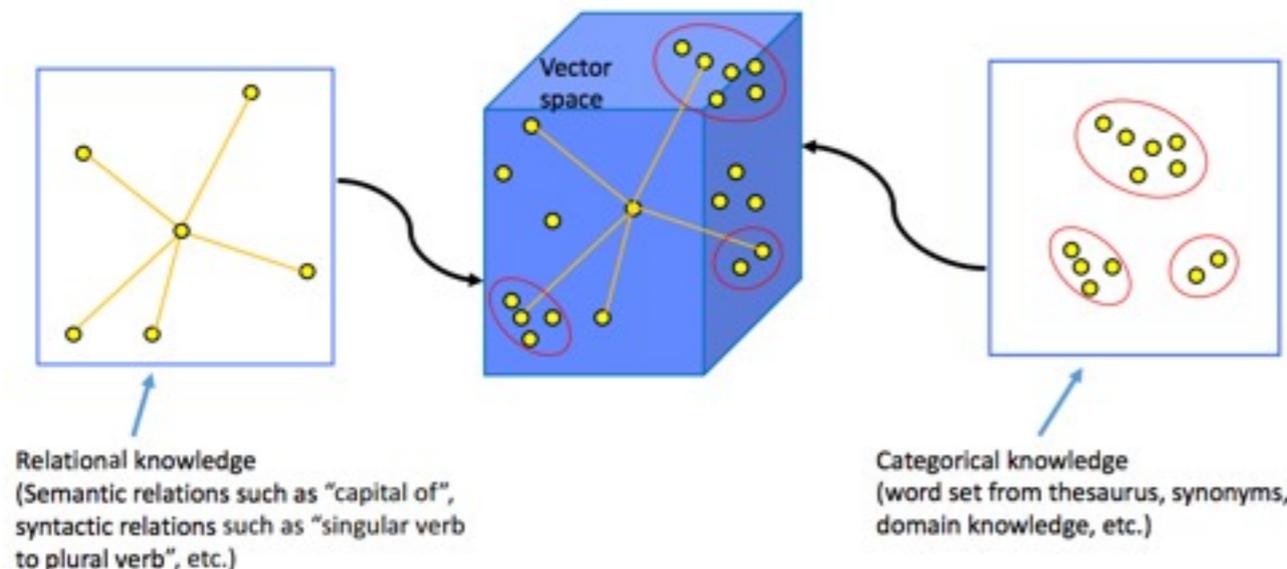
$$J_c = \beta E_c - L$$

$$E_c = \sum_{i=1}^V \sum_{j=1}^V \underbrace{s(w_i, w_j)}_{\text{カテゴリ情報に基づく類似度}} \underbrace{d(w_i, w_j)}_{\text{潜在空間上の距離}}$$

$$\sum_{j=1}^V s(w_i, w_j) = 1,$$

カテゴリ情報的に類似している単語が、潜在空間上で離れているとコストが発生

提案手法 RC-NET



目標：R-NETとC-NETの良いところ取り

$$J = \alpha E_r + \beta E_c - L.$$

● Joint Knowledge Powered Model : RC-NET

- Relational Knowledge : 潜在空間の大局的な構造を作る
- Categorical Knowledge : 潜在空間の局所的な構造を作る

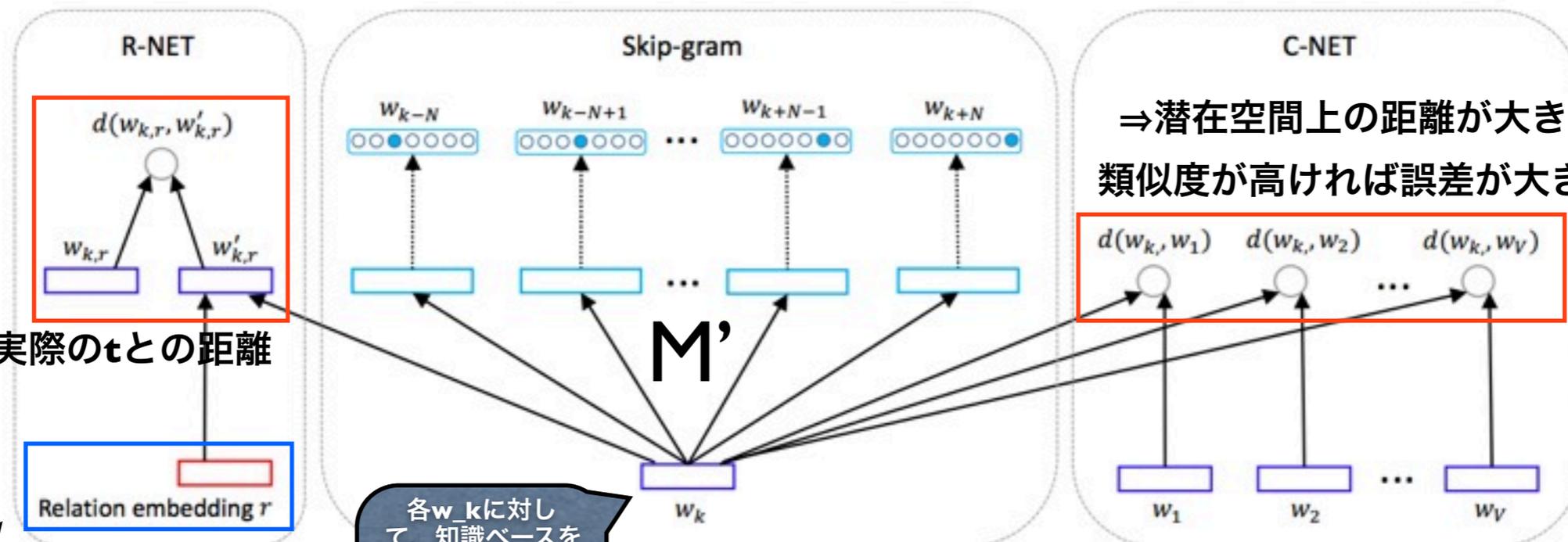


Figure 4: The architecture of RC-NET. The objective is to learn word representations and relation representations based on text stream, relational knowledge, and categorical knowledge.

実験 概要

● 3つの実験で単語の分散表現を評価

- 1, Analogical reasoning
- 2, Word Similarity
- 3, Topic Prediction

● 実験環境など

- 比較手法 : Skipgram, R-NET, C-NET, RC-NET
- パラメータ : α や β は勾配に合わせて動的に変更

実験1 Analogical Reasoning Task

● 同じ関係にある単語の予測タスク

● 以下のようなタプルを計34773個用意. 教師:テストは4:1.

● 例1) Germany : Berlin = France : ?

● 例2) clear : unclear = reasonable : ?

● Mikolovが公開しているデータセット <https://code.google.com/p/word2vec/source/browse/trunk/questions-words.txt>

● + WordRep <http://arxiv.org/pdf/1407.1640v1.pdf>

● 推定

● $\text{vec}(\text{France}) + (\text{vec}(\text{Berlin}) - \text{vec}(\text{Germany}))$ とコサイン類似度が一番高い単語

● 比較手法

● Skipgram, Skipgram+R-NET, Skipgram+C-NET, Skipgram+RC-NET

● 分散表現の次元 = 100, 300

● 備考

● R-NET : 教師データの意味関係から訓練

● C-NET : 各タスクのheadに該当する単語群を一つのカテゴリにまとめる (後述)

●

Example of Dataset

● Mikolov

1	:	capital-common-countries	8368	:	family				
2	Athens	Greece	Baghdad	Iraq	8369	boy	girl	brother	sister
3	Athens	Greece	Bangkok	Thailand	8370	boy	girl	brothers	sisters
4	Athens	Greece	Beijing	China	8371	boy	girl	dad	mom
5	Athens	Greece	Berlin	Germany	8372	boy	girl	father	mother
6	Athens	Greece	Bern	Switzerland	8373	boy	girl	grandfather	grandmother
7	Athens	Greece	Cairo	Egypt	8374	boy	girl	grandpa	grandma
8	Athens	Greece	Canberra	Australia	8375	boy	girl	grandson	granddaughter
9	Athens	Greece	Hanoi	Vietnam	8376	boy	girl	groom	bride
10	Athens	Greece	Havana	Cuba	8377	boy	girl	he	she
11	Athens	Greece	Helsinki	Finland					

● WordRep

Table 1. Summary of Mikolov et al.'s evaluation set on analogical reasoning task.

Subtask	Word pair 1 (<i>a, b</i>)		Word pair 2 (<i>c, d</i>)		# word pairs	# tuples
Common capital city	Athens	Greece	Oslo	Norway	23	506
All capital cities	Astana	Kazakhstan	Harare	Zimbabwe	116	4,524
Currency	Angola	kwanza	Iran	rial	30	866
City-in-state	Chicago	Illinois	Stockton	California	68	2,467
Man-Woman	brother	sister	grandson	granddaughter	23	506
Adjective to adverb	apparent	apparently	rapid	rapidly	32	992
Opposite	possibly	impossibly	ethical	unethical	29	812
Comparative	great	greater	tough	tougher	37	1,332
Superlative	easy	easiest	lucky	luckiest	34	1,122
Present Participle	think	thinking	read	reading	33	1,056
Nationality adjective	Switzerland	Swiss	Cambodia	Cambodian	41	1,599
Past tense	walking	walked	swimming	swam	40	1,560
Plural nouns	mouse	mice	dollar	dollars	37	1,332
Plural verbs	work	works	speak	speaks	30	870
# Total	/	/	/	/	573	19,544

実験1 結果

● 各比較手法における正解率

Model	Vector Dimensionality	Accuracy[%]		
		Semantic	Syntactic	Total
Skip-gram	100	25.06	36.49	31.30
	300	28.76	40.31	35.07
R-NET	100	26.91	39.37	33.56
	300	32.64	43.46	38.55
C-NET	100	29.67	36.12	33.19
	300	37.07	40.06	39.00
RC-NET	100	32.02	43.92	38.52
	300	34.36	44.42	39.85

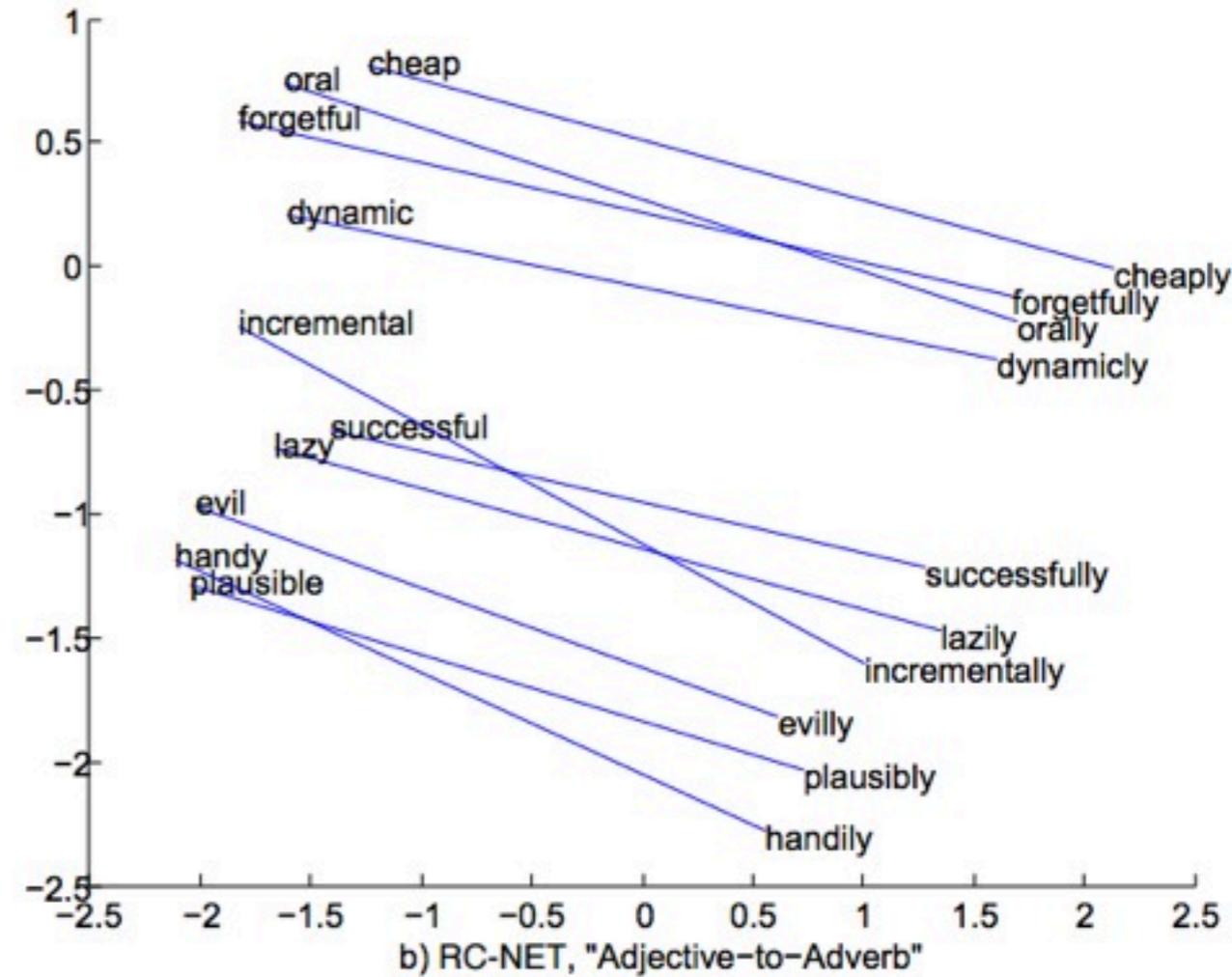
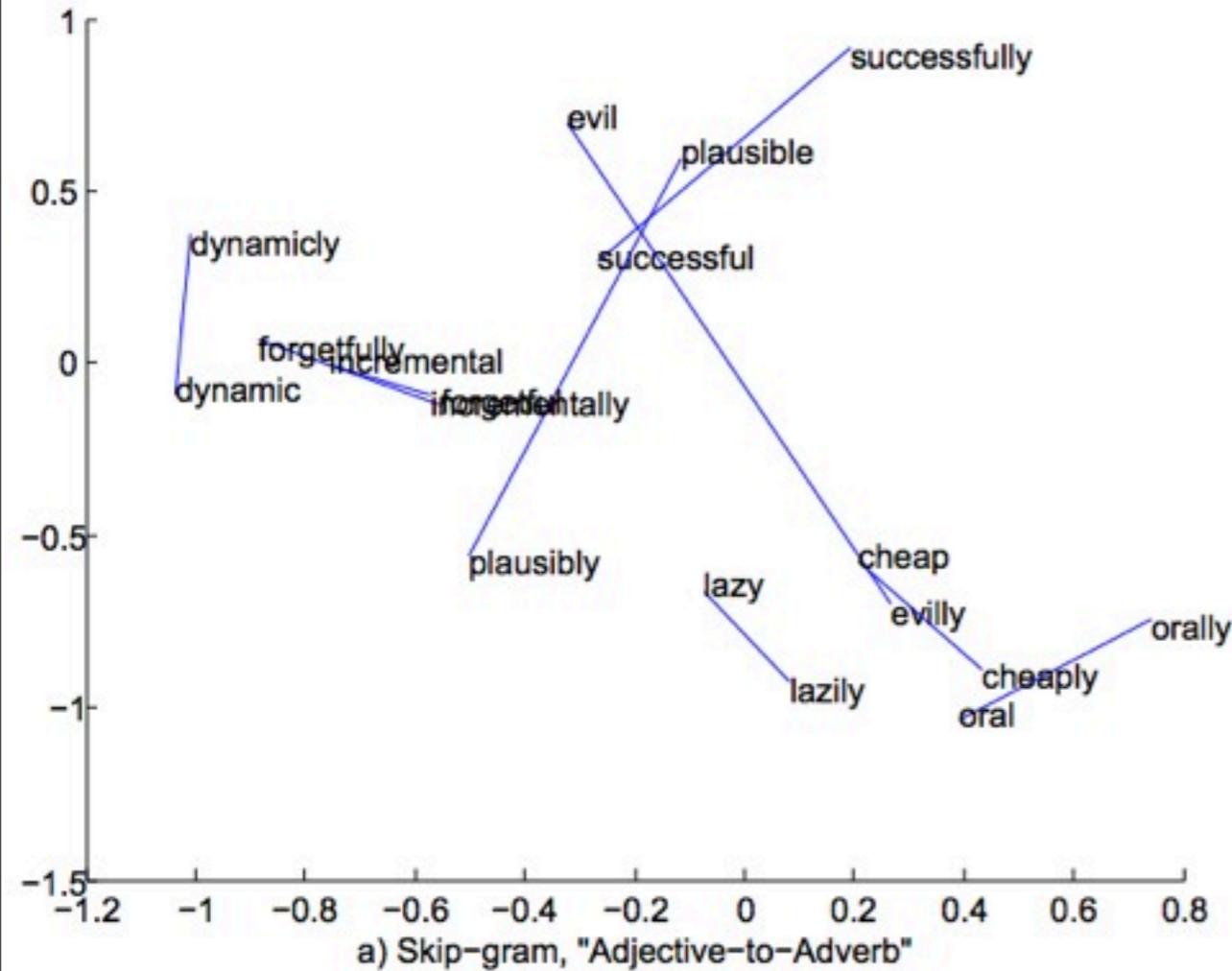
“knowledge powered word embedding is of higher quality than the baseline model with no knowledge regularizations.”

● 概して、知識を入れた方が精度は高い

● # $|V| = 20$ 万語の中で、正解率40%@1 という精度はすごい...

● 定性的な評価は次のスライドで

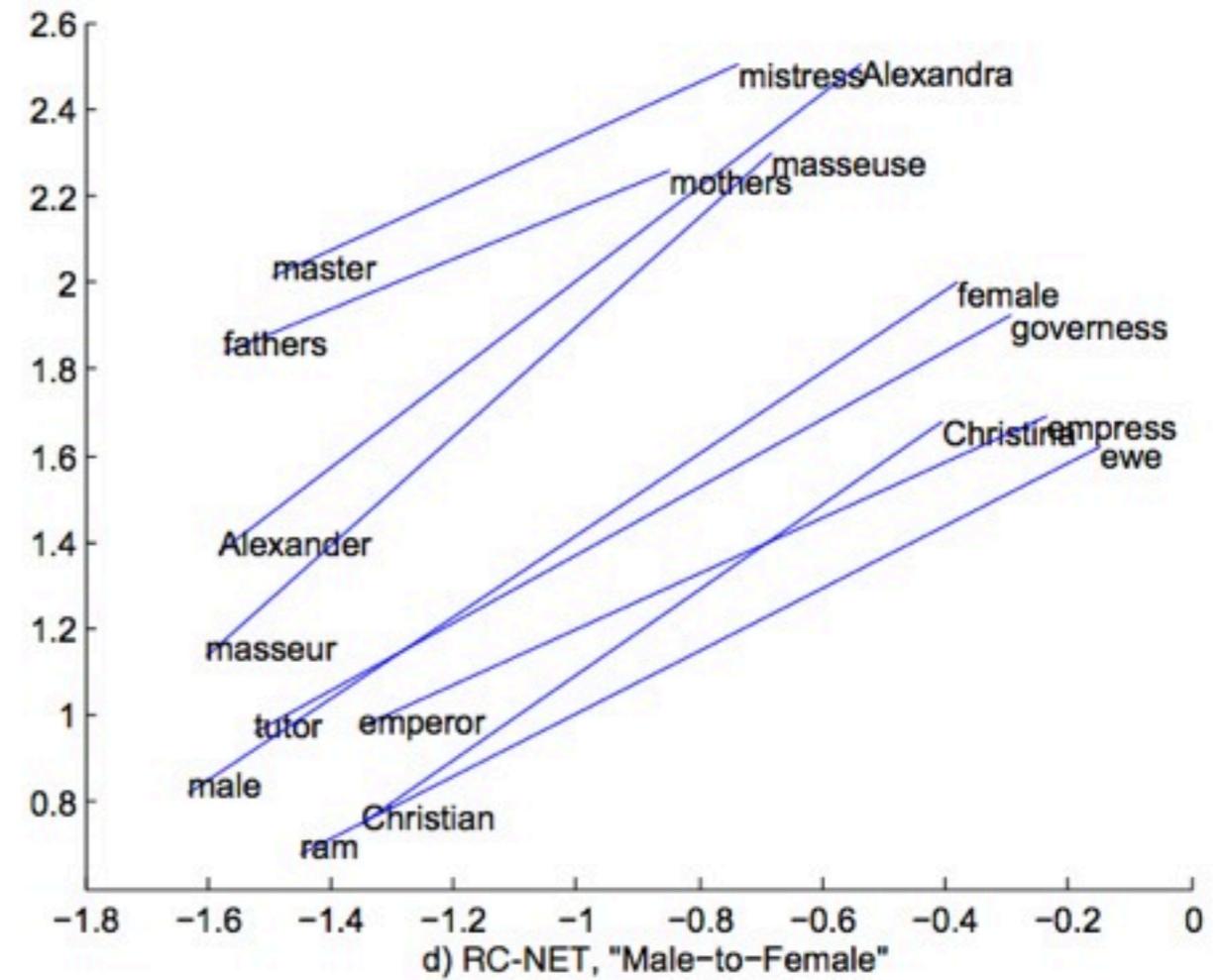
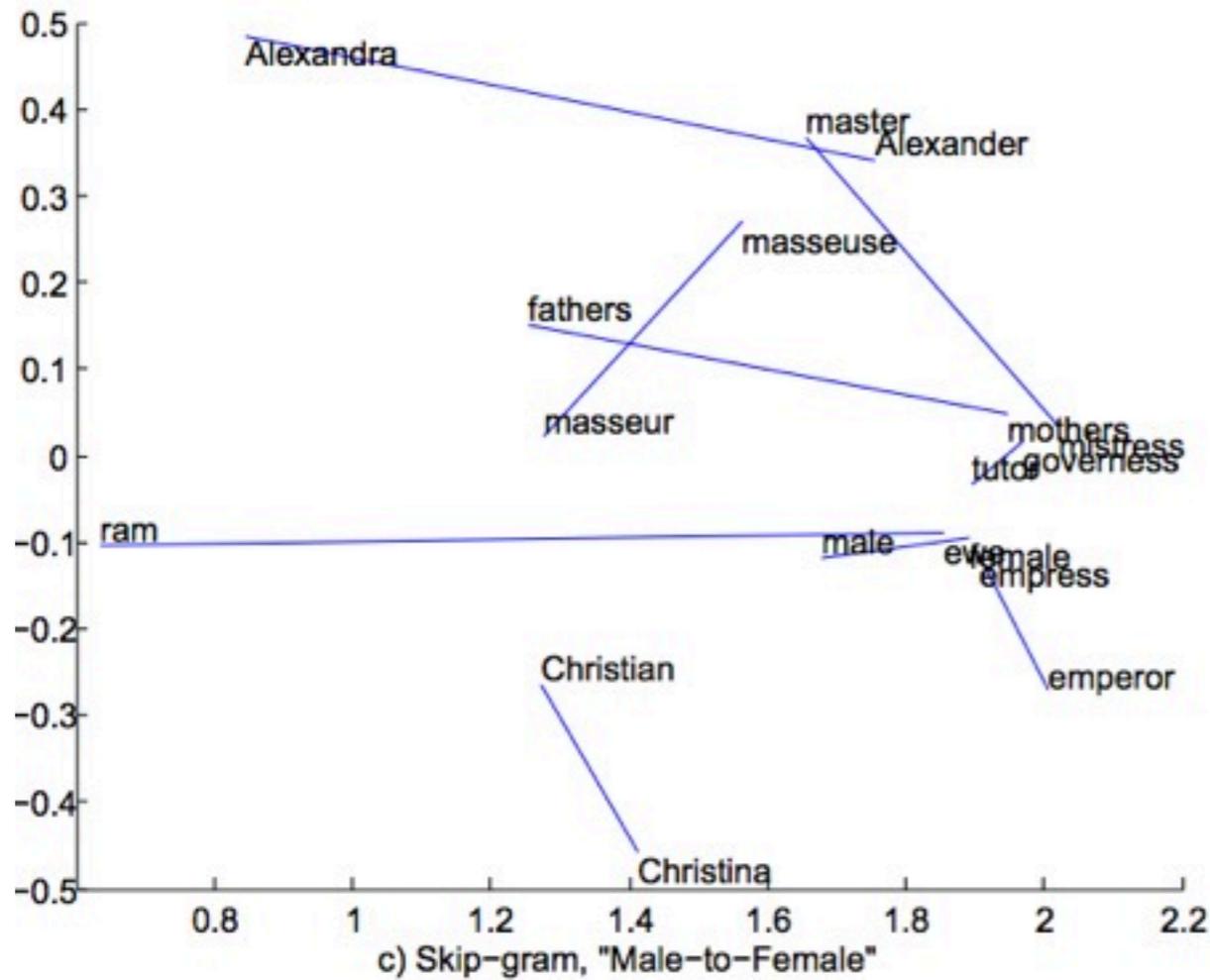
実験1 結果



● Syntactic Relationshipの定性評価

- 例) Adjective to Adverbの評価 (2次元PCA)
- 同じ距離&同じ方向に配置されるようになった
- 同じような意味の単語がまとまるようになった。(ホント...?)

実験1 結果



● Semantic Relationshipの定性評価

- 例) Male to Femaleの評価
- 同様に方向と距離がいい感じにまとまった
- 意味は依然としてまとまったのか不明

実験2 Word Similarity Task

● 単語の類似性の予測

● 人手でラベル付けされた単語の類似性の評価値当てる

● WordSim-353というデータセット 例) (cup, drink) \Rightarrow 7.25, (cup, substance) \Rightarrow 1.92

● #13 ~ 16人の人が, 0~10の評価値で類似性を判定したらしい

● 比較手法

● SkipgramとC-Netの2者間比較

● 理由は, 類似度タスクにはR-NETは関係ないからとのこと...

● カテゴリの知識は, Freebaseからtype-ofの知識から取得

● astronomy, biology, boats, chemistry, computer, fashion, food, geology, interests, language

● 評価値: スピアマンの順位相関係数

● A(正解データ): 353個のラベルデータを順位付け

● B(提案手法): 353個の類似度をコサイン類似度(?)で順位付け

実験2 結果

Table 2: Results obtained by the different methods on the word similarity task.

Methods	Vector Dimensionality	Spearman's ρ correlation
Skip-gram	100	0.652
	300	0.678
C-NET	100	0.661
	300	0.683

- 若干良い結果
- Freebaseを明示的に学習プロセスに考慮できた

実験3 Topic Prediction Task

● ある単語に対するトピック(単語)の予測

- 例) “star”, “earth” \Rightarrow topic:astronomy
- 例) “cell”, “neuron” \Rightarrow topic:biology

● データセット

- 全実験同様, Freebaseのトピック10個

● 推定方法

- h : ある単語(入力), r : topic-relation, t : topic word
- 潜在空間上で, $h+r$ に一番近い t' を予測したトピックとする
 - t は事前に決めた10個のどれか. だからランダムにやっても10%は当たる

● その他

- 訓練データ, テストデータは1:1
- SkipgramやC-NETでは, r は学習しないので $t-h$ の平均を r とする

実験3 結果

Table 3: Results obtained by the comparing methods on the topic prediction task.

Relation	Error Rate[%]			
	Skip-gram	C-NET	R-NET	RC-NET
astronomy	2.00	2.00	8.00	2.00
biology	6.29	4.91	4.91	4.32
boats	11.76	5.88	5.88	7.84
chemistry	6.67	5.71	9.52	15.24
computer	19.54	8.62	6.32	4.02
fashion	22.08	24.68	22.08	22.08
food	17.98	13.60	11.40	7.89
geology	0.00	0.00	11.54	7.69
interests	6.80	8.16	6.12	6.12
language	33.33	33.33	18.52	7.41
Total	12.21	9.37	8.57	7.15

- 多くのトピックで、提案手法が上回っている
- 特に、RC-NETが最も良い結果と言える
- 一部悪くなっている理由
 - 1, 訓練データが少なかった
 - 例) astronomy, geologyはデータが少なかった
 - 2, 稀な単語が多い場合はうまくいかない
 - 例) chemistryでは、稀な単語が多かった

まとめ

● 論文の結論

- 精度の高い単語の分散表現の学習はNLPでは重要
- この論文ではknowledge graphを学習プロセスに組み込む方法を提案
- 関係性に関する知識はエンティティ間の関係の、
カテゴリに関する知識はエンティティの属性の推定に利用できる
- 3種類の実験により、単語の分散表現の精度が高まった事を示した
- 今後は、形態学(morphological)の知識の組み込み方などを検討予定