

everything changes, nothing perishes
Omnia Mutantur, Nihil Interit:
Connecting Past with Present by Finding
Corresponding Terms across Time

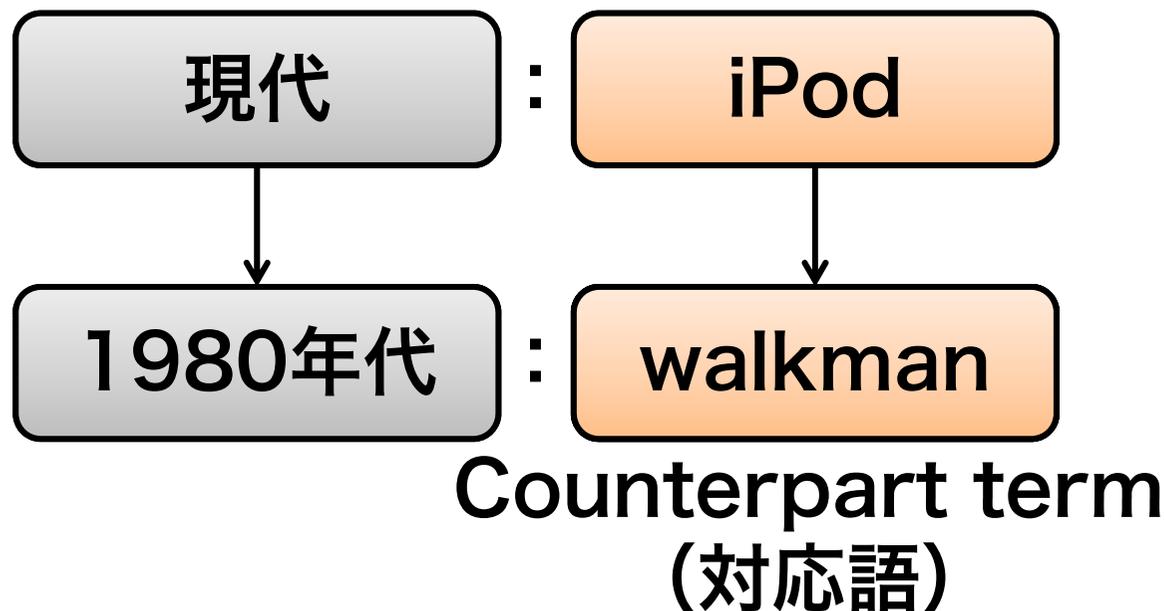
Yating Zhang, Adam Jatowt,
Sourav S Bhowmick, Katsumi Tanaka
ACL 2015

TAKIMOTO Hiroki

Overview

- 時間を跨いだ対応を発見する問題
e.g. 1980年代におけるiPodとは？
→Walkman
- 時代毎にベクトル空間を学習
→空間を変換する行列を学習し利用
- 対応を発見したい語句 (iPod) の文脈語を
参照点として利用し、品質の向上を目指す
– 関連度、一般化語、クラスタリングの三つの
選択方法を比較した

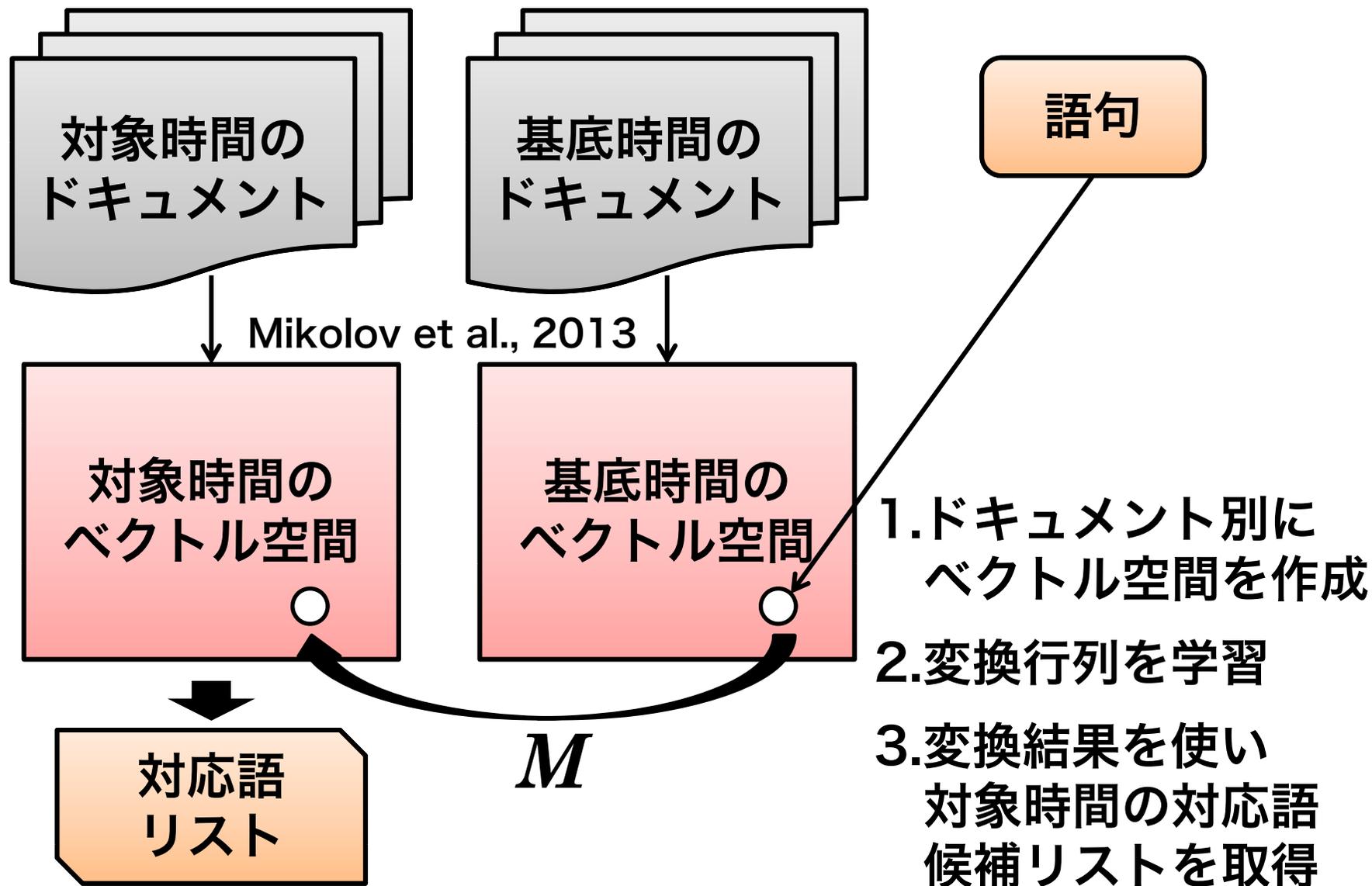
目的と挑戦



挑戦

1. 年代が違くと文脈も異なる
→ 文脈ベクトルを直接比較しても見つけにくい
2. トレーニングデータ (対応語対) が少ない
→ 代表的な機械学習タスクに落としにくい

異なる時間からの対応語の発見



変換行列の学習

対応語対: $\{(\omega_1, w_1), \dots, (\omega_k, w_k)\}$

$$M = \arg \min_M \sum_{i=1}^K \|M \cdot \omega_i - w_i\|_2^2 + \gamma \|M\|_2^2$$

いや、対応語対が少ないのが問題なんじゃ？

→ Common Frequent Terms (CFTs)を使う

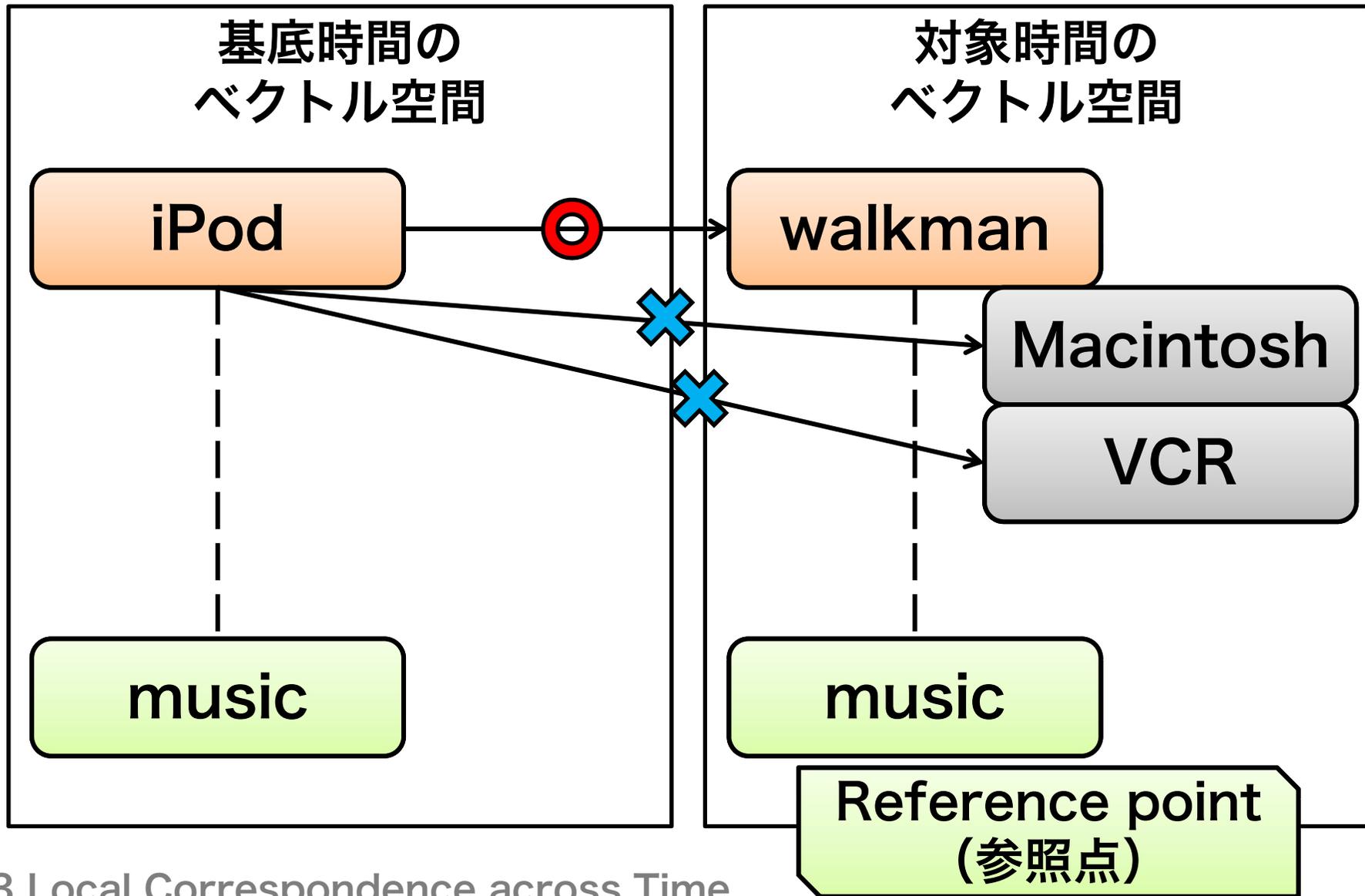
CFTs

- 規定時間でも対象時間でも、統語論的形態が同じ
- どちらの時間でも頻繁に使われている

(e.g., man, women, water, dog, see, three)

数はヒューリスティック頼み

関連語を使った精度の向上

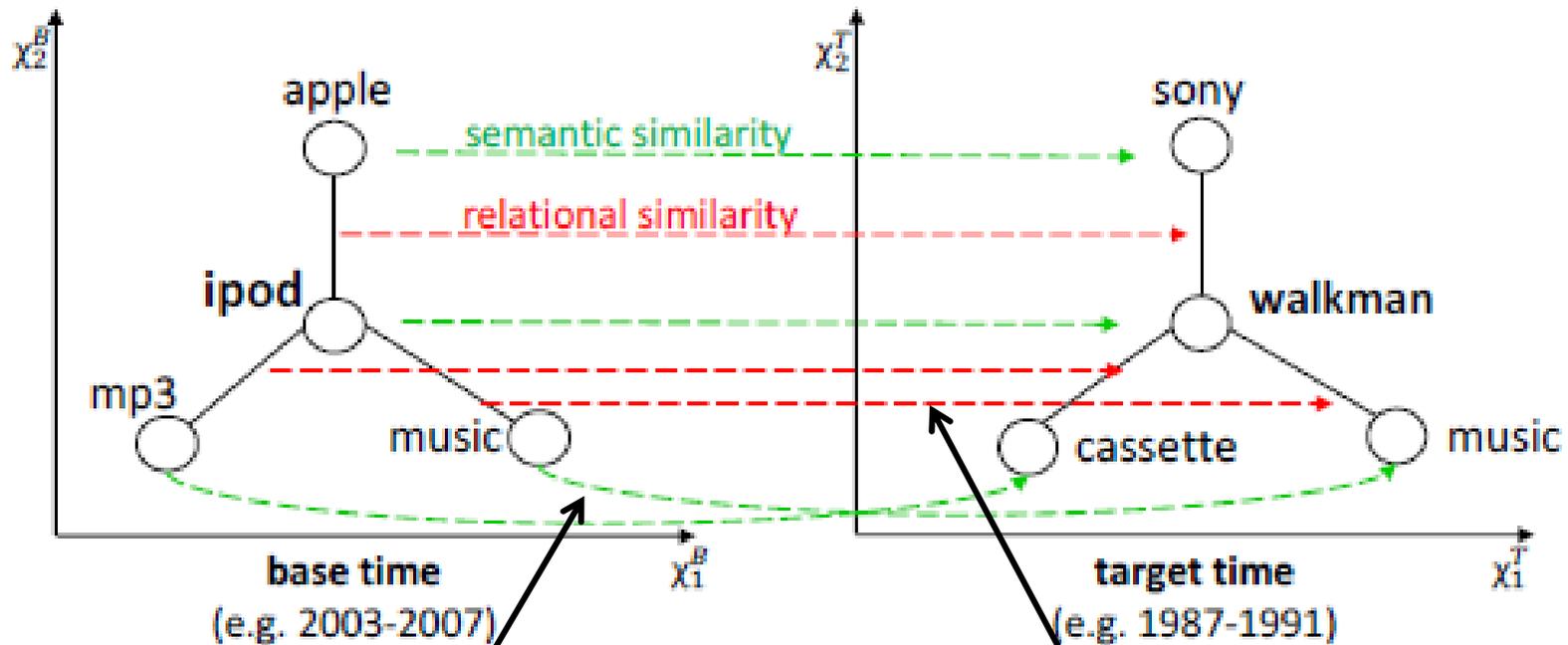


参照点の選択方針三種類

語句の文脈語で以下の何れかを満たす

1. 語句と高い関連度を持つこと
→ *tf*と*relatedness*により計算
2. 十分に一般的であること
→ (Ohshima et al., 2010) の手法により
Hypernymsを決定
3. (参照点が) お互いに独立していること
→ bisecting k-meansにより文脈語を
クラスタリング、代表語を選出し利用

局所グラフ同士の比較



Semantic はノードの位置

relational はノード間の位置関係

ある参照点に対応する点の選び方

→各参照点毎に二つの類似度を最大化する点を選択

訓練データとテストデータ

- 訓練データ
 - New York Times Annotated Corpus
 - 1987-1991, 1992-1996, 1997-2001, 2002-2007 の4つの期間に分けて利用
- テストデータ
 - 人、場所、物の3種類の対応語対を人手作成
 - 全95語対、54対を1987-1991用に、50対を1992-1996用に
 - (この妥当性はちょっと疑問)

テストデータと結果の一例 (表1の一部拡大)

q [2002,2007]	tc [1987,1991]	BOW (baseline)	LSI-Com (baseline)	LSI-Tran (baseline)	GT (proposed)	LT-Cooc (proposed)
Putin	Yeltsin	1000+	252	353	24	1
Chirac	Mitterrand	1000+	8	1	7	19
iPod	Walkman	1000+	20	131	3	13
Merkel	Kohl	1000+	1000+	537	142	76
Facebook	Usenet	1000+	1000+	1000+	1	1
Linux	Unix	1000+	11	1	20	1
email	letter	1000+	1000+	464	1	35
email	mail	1000+	1	9	7	2
email	fax	1000+	1000+	10	3	1
Pixar	Tristar	1000+	549	1	1	1
Pixar	Disney	1000+	4	4	3	2
Serbia	Yugoslavia	1000+	15	1000+	1	1
mp3	compact disk	1000+	56	44	58	17
Rogge	Samaranch	1000+	4	22	12	82

•
•
•

結果 (表 2 : 現在から過去)

Method	MRR	P@1	P@5	P@10	P@20
BOW	4.1E-5	0	0	0	0
LSI-Com	0.206	15.8	27.3	29.5	38.6
LSI-Tran	0.112	7.9	13.6	21.6	22.7
GT	0.298	16.8	44.2	56.8	73.7
LT-Cooc	0.283	18.8	35.3	50.6	62.4
LT-Lex	0.369	24.2	49.5	63.2	71.6
LT-Clust	0.285	14.7	42.1	55.1	65.2

BOW (bag of words) はあまりよい結果にならず
→前提通り、年代が違くと文脈を直接比較してもダメ

結果（表2：現在から過去）

Method	MRR	P@1	P@5	P@10	P@20
BOW	4.1E-5	0	0	0	0
LSI-Com	0.206	15.8	27.3	29.5	38.6
LSI-Tran	0.112	7.9	13.6	21.6	22.7
GT	0.298	16.8	44.2	56.8	73.7
LT-Cooc	0.283	18.8	35.3	50.6	62.4
LT-Lex	0.369	24.2	49.5	63.2	71.6
LT-Clust	0.285	14.7	42.1	55.1	65.2

LSIには今回の変換方法はうまく働かず
→semanticの変換には向いていても
Topicの変換には向いていなかったか

結果 (表2 : 現在から過去)

Method	MRR	P@1	P@5	P@10	P@20
BOW	4.1E-5	0	0	0	0
LSI-Com	0.206	15.8	27.3	29.5	38.6
LSI-Tran	0.112	7.9	13.6	21.6	22.7
GT	0.298	16.8	44.2	56.8	73.7
LT-Cooc	0.283	18.8	35.3	50.6	62.4
LT-Lex	0.369	24.2	49.5	63.2	71.6
LT-Clust	0.285	14.7	42.1	55.1	65.2

全体的に提案手法 (GTやLT) は他手法を凌ぐ結果中でもLT-Lex (文脈語を一般化した語を参照点に選択) が最も良い結果を出した
他の二つの参照点はGTの改善に失敗した

Conclusions

- 入力語句に対応する、別の時代の語を発見するタスク
- 異なる時代のベクトル空間を変換する行列を学習し利用、更に入力語句の文脈語を参照点として利用し、品質の向上を目論んだ
- 提案手法がベースラインを凌いだ
- 変換行列はLSIには上手く働かなかった
- 文脈語を一般化した語を参照点として利用すると品質が向上した

付録

結果（表2：現在から過去）

Method	MRR	P@1	P@5	P@10	P@20
BOW	4.1E-5	0	0	0	0
LSI-Com	0.206	15.8	27.3	29.5	38.6
LSI-Tran	0.112	7.9	13.6	21.6	22.7
GT	0.298	16.8	44.2	56.8	73.7
LT-Cooc	0.283	18.8	35.3	50.6	62.4
LT-Lex	0.369	24.2	49.5	63.2	71.6
LT-Clust	0.285	14.7	42.1	55.1	65.2

iPod→apple は walkman→Sony

→hypernymsにすると両方共 →Company

→他の二つではこうは行かないのが悪化の原因？

変換行列はCFTsを基準に学習される

→LT-Lexは変換行列の基準とよく合いそう（所感）

結果（過去から現在）

Method	MRR	P@1	P@5	P@10	P@20
BOW	3.4E-5	0	0	0	0
LSI-Com	0.181	13.2	19.7	28.9	35.5
LSI-Tran	0.109	5.3	17.1	21.1	23.7
GT	0.226	15.2	27.3	33.3	45.5
LT-Cooc	0.231	14.7	30.7	36	46.7
LT-Lex	0.235	16.7	28.8	31.8	48.5
LT-Clust	0.228	13.6	28.8	31.8	47

逆方向でも何となく使えてるっぽい
現在から過去の結果とは非対称的
→過去から現在方向は物事の多様性が増すから
難しいのでは

参照点の選び方

Algorithm 3 Local Graph Matching

Input: local graph of q , S_q^{FB}
 W = top k corresponding terms of q (by Eq. 2)
 FF = {top k corresponding terms of each f in reference points $F_B = \{f_0, f_1, \dots, f_u\}$ } (by Eq. 2)
for $w = W[1:k]$ do:
 sum_cos = 0 # total graph similarity score
 for $F = FF[1:u]$ do:
 max_cos = 0 # current maximum similarity
 for $c = F[1:k]$ do:
 find c which maximizes current graph similarity
 end for
 sum_cos += max_cos
 end for
end for
sort W by sum_cos of each w in W .
Output: sorted W as ranked list of temporal counterparts

ある対応候補の…

ある参照点に対して…

参照点对応候補の中から…

一番良いものを選ぶ

✕ 参照点候補の全組み合わせを試す

○ 参照点毎に最適な物を選ぶ

異なる年代間での実験結果

Method	MRR	P@1	P@5	P@10	P@20
LSI-Com	0.115	10.6	14.9	21.3	23.4
GT	0.132	8.5	27.7	40.4	53.2
LT-Lex	0.169	10.6	34.1	48.9	55.3

Table 4: Results of searching from present to past (present: 2002-2007; past: 1992-1996).

Method	MRR	P@1	P@5	P@10	P@20
LSI-Com	0.148	11.6	18.6	22.2	20.2
GT	0.184	11.6	23.3	33.3	40.2
LT-Lex	0.212	14	28	38	45

Table 5: Results of searching from past to present (present: 2002-2007; past: 1992-1996).

あれ、過去から
現在のほうがいい結果？
→テストセットが
違うから？

CFTsが結果に与える影響

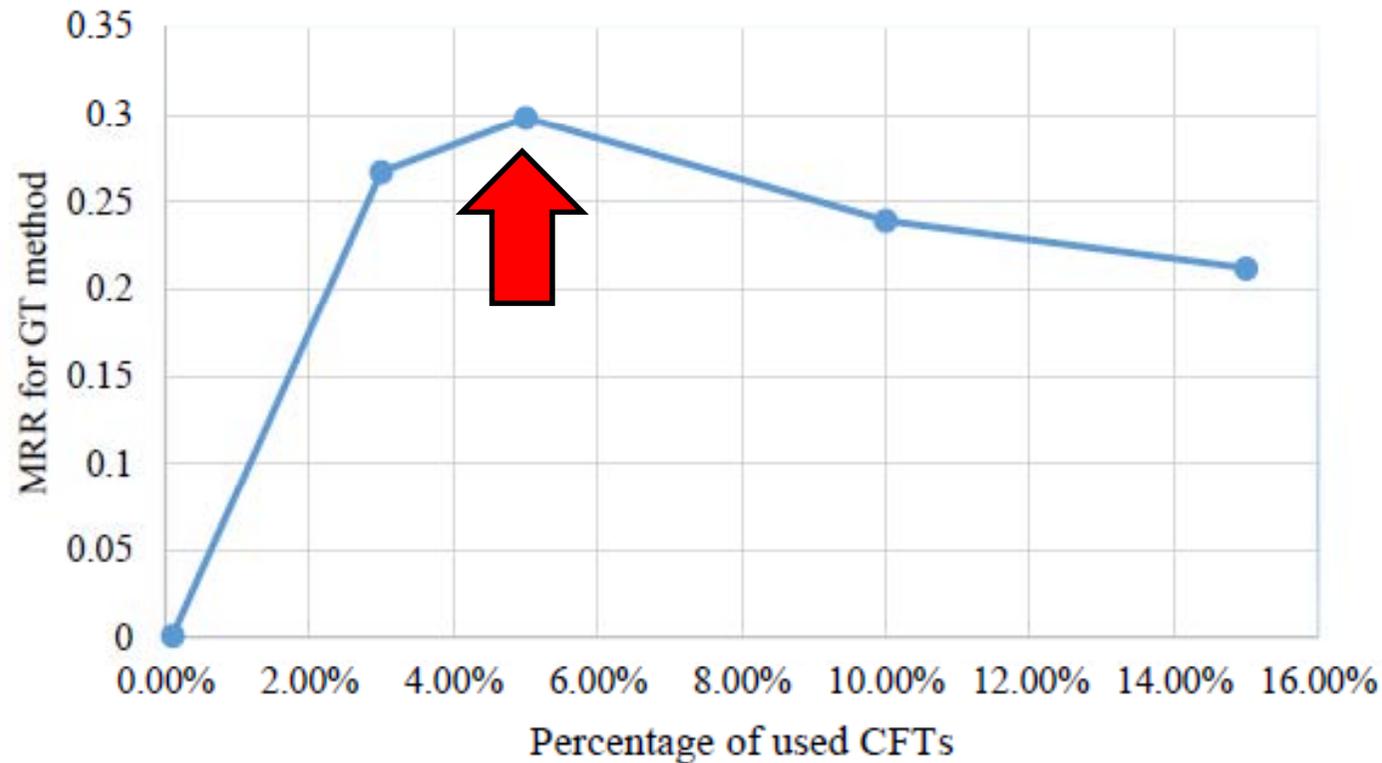


Figure 4: Results of MRR for **GT** method depending on number of used *CFTs*.

Parameters

- Skip-gramやLSIの次元数
200
- CFTsの数
トップ5% (18k words)
- 参照点の数
5