

意外語発見手法 の 動画検索への拡張

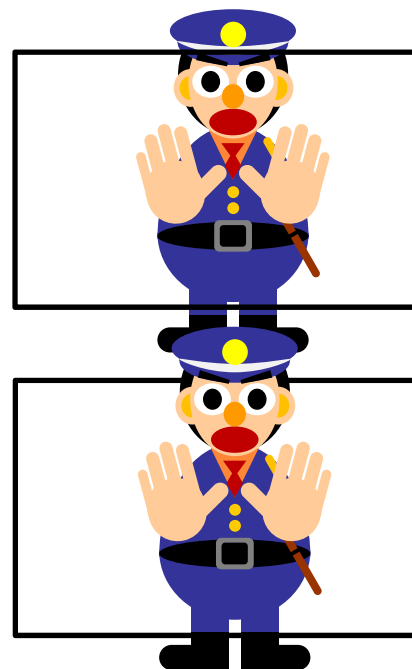
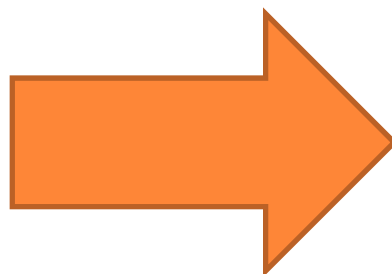
日本大学文理学部 情報システム解析学科

山本広大

研究背景



タグ タグ ... タグ タグ



あたりまえの動画



タグ タグ ... タグ タグ

意外性のある動画

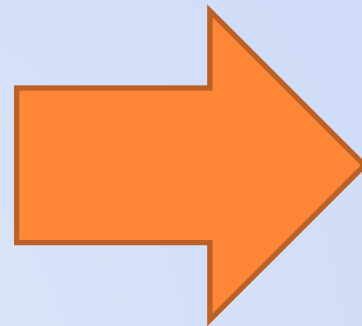
タグ間の意外性を図る

意外語の抽出:[佃ら 情報処理学会トランザクション 2014]

関連研究

語→意外度の高い語

落合博満



ガンダム

意外性とは



意外度 \longleftrightarrow 繋がりの弱さ \times 認知度



意外度の定義

繋がりの弱さ

どれほど認知されているか？

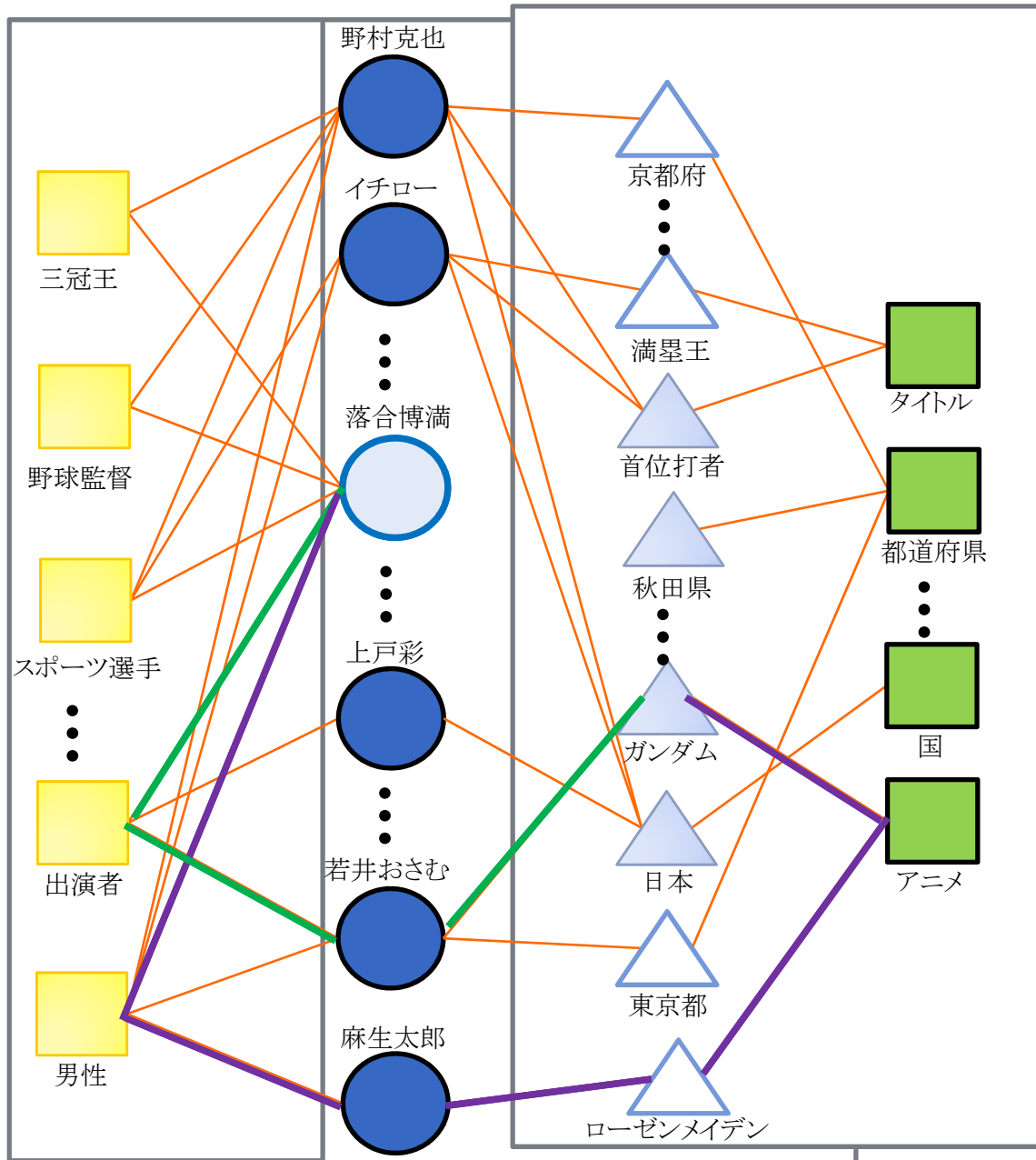
$$\text{意外度} = \frac{1}{\text{典型度}} \times \text{認知度}$$

次のスライドにて説明

Wikipedia上での
重要度を用いる



典型度



同位語らしさ:
SALSAアルゴリズム

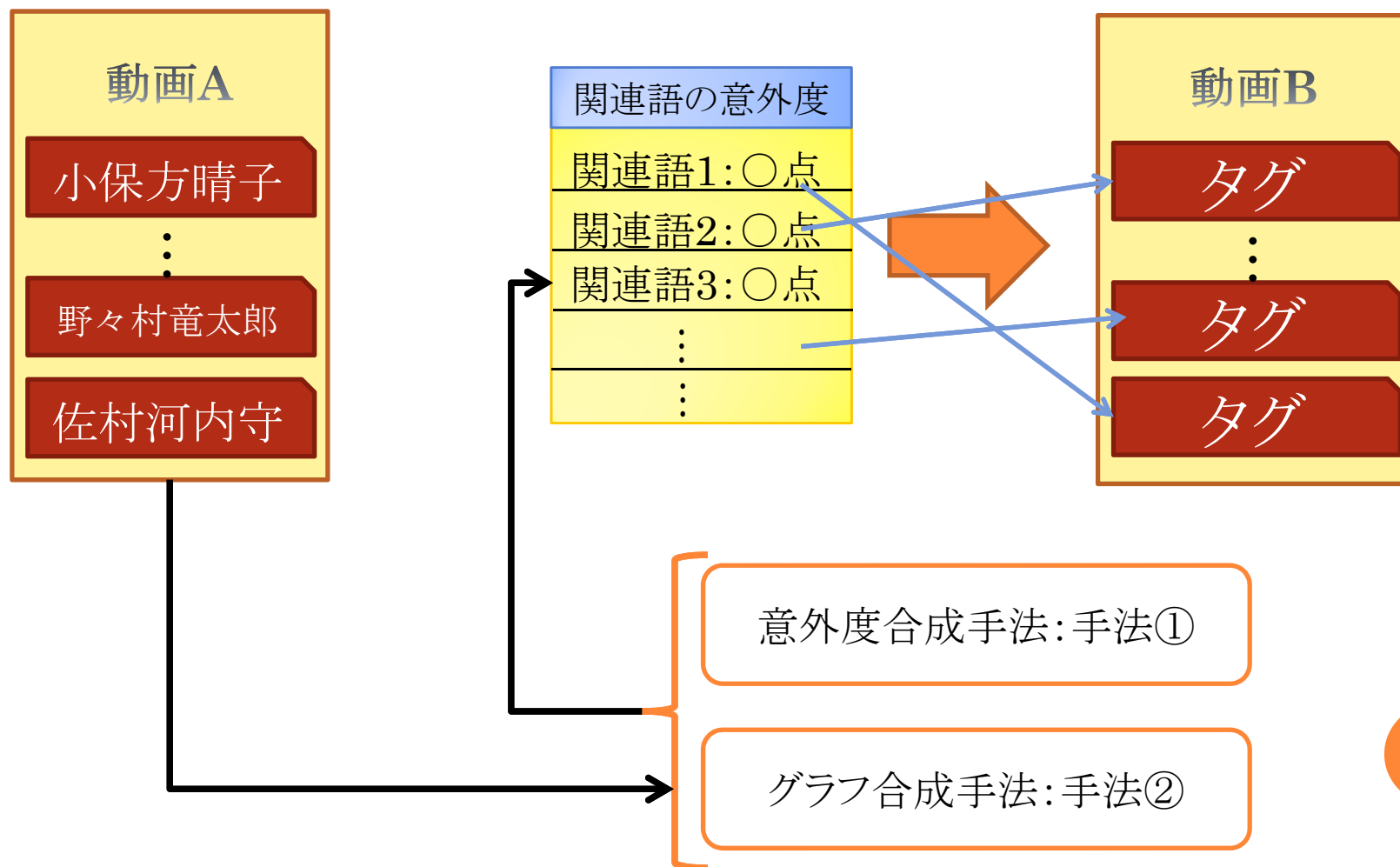
典型度①:
BiasedPageRankアルゴリズム
— のルートでの典型度が測れる

典型度②:
co-HITSアルゴリズム
— のルートでの典型度を考慮した

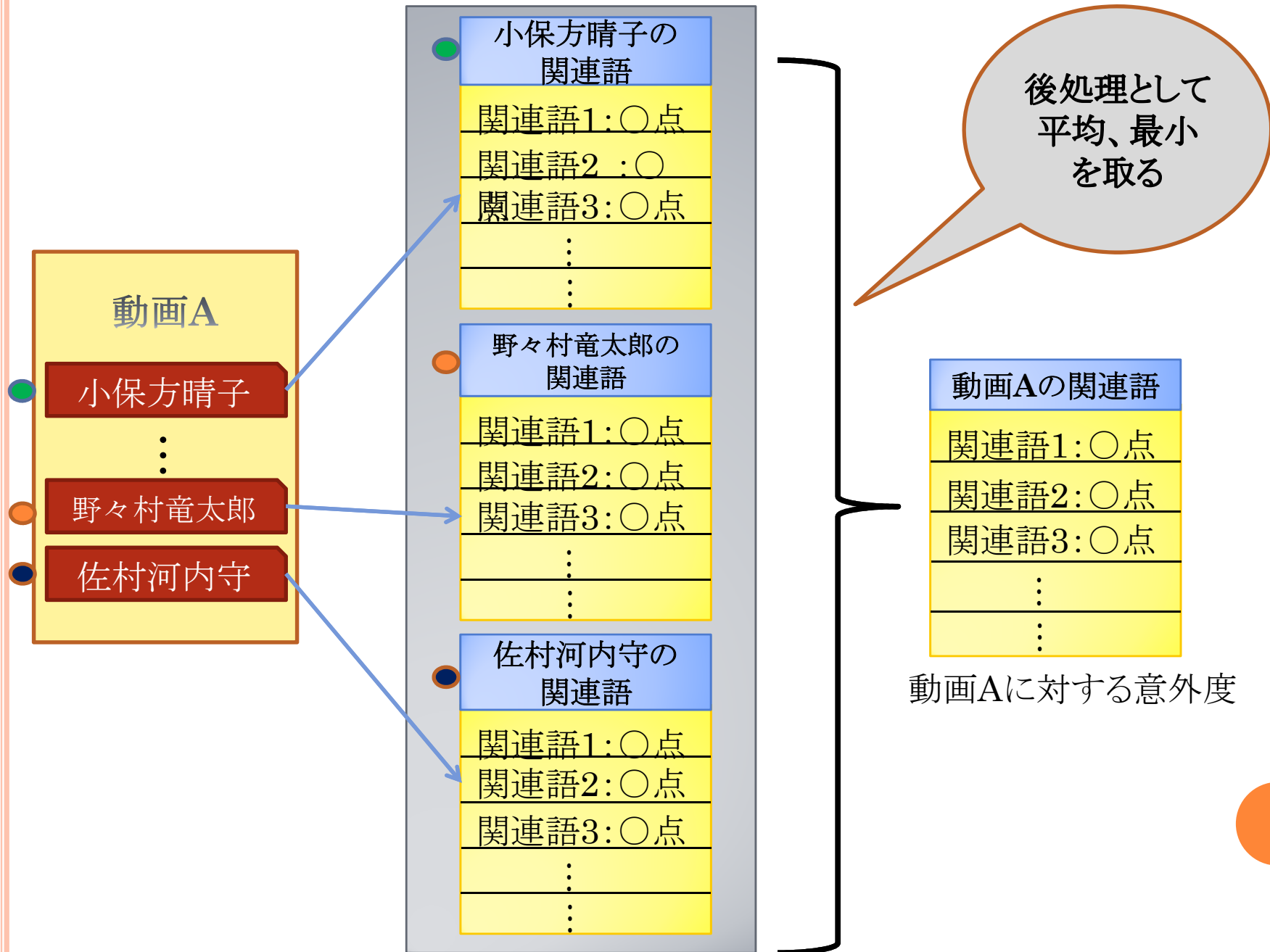


提案手法

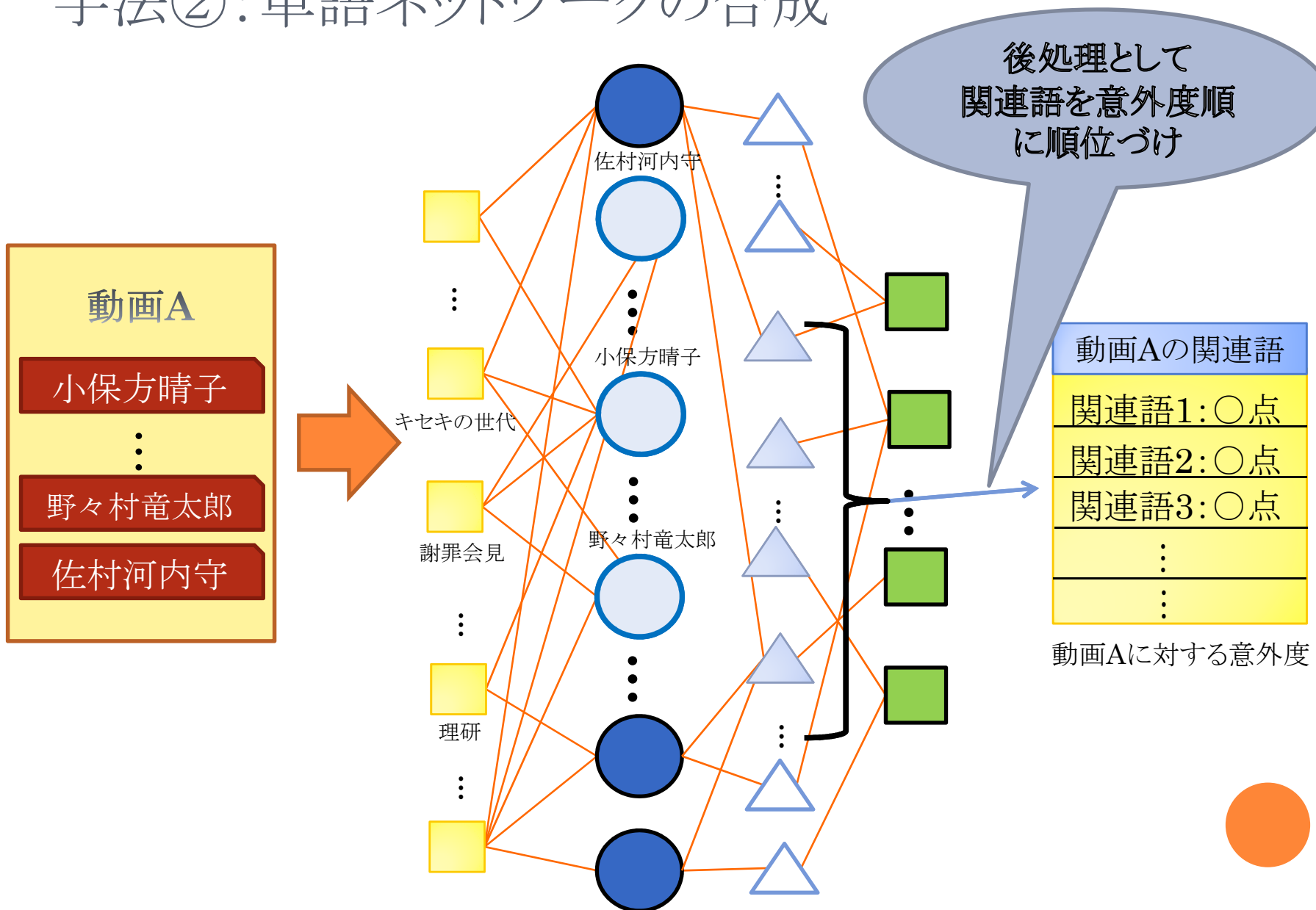
”語→意外度”から”語群→意外度”への拡張



手法①: 意外度の合成



手法②: 単語ネットワークの合成



評価実験

- ① ケンドーコバヤシ, アントニオ猪木
- ② 毛利小五郎, 毛利蘭
- ③ ドラゴンクエスト, ファイナルファンタジー
- ④ スパイダーマン, バットマン
- ⑤ AKB48, TOKIO

Wikipediaオントロジーに含まれている語より。

	①	②	③	④	⑤
総関連語数	453	103	139	188	430

	①	②	③	④	⑤
意外度1の数	209	49	72	111	255
意外度2の数	87	17	44	31	72
意外度3の数	67	18	9	17	59
意外度4の数	90	19	14	29	44

意外度合成手法: 手法①

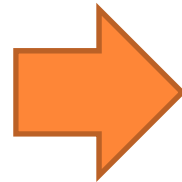
グラフ合成手法: 手法②

評価

評価方法

手動で関連語に予想値を付与

関連語	予想値
A	3
B	4
C	1
D	2
E	4
F	4
G	4
H	3



意外度	関連語	予想値
22.07845	A	2
8.675918	B	4
22.95586	C	3
14.94621	D	2
18.04504	E	1
22.62594	F	4
9.963881	G	1
8.437681	H	3

ソート



意外度	関連語	予想値
22.95586	A	3
22.62594	B	4
22.07845	C	2
18.04504	D	1
14.94621	E	2
9.963881	F	1
8.675918	G	4
8.437681	H	3

precision@k
k位以内での正解率
average precision
で評価

提案手法で意外度を付与



実験結果

- グラフ合成手法 > 意外度合成手法
- 平均, 最小の値は、ほぼ変化せず
- カテゴリが狭いと精度が高い

precision @10

	ケンドーコバヤシ アントニオ猪木	毛利小五郎 毛利蘭	ドラゴンクエスト ファイナルファンタジー	スパイダーマン バットマン	AKB48 TOKIO
グラフ合成手法	0.2	0.5	0.3	0.3	0.0
意外度合成手法 avg	0.2	0.4	0.1	0.2	0.1
意外度合成手法 min	0.2	0.4	0.1	0.2	0.1

average precision

	ケンドーコバヤシ アントニオ猪木	毛利小五郎 毛利蘭	ドラゴンクエスト ファイナルファンタジー	スパイダーマン バットマン	AKB48 TOKIO
グラフ合成手法	0.12	0.33	0.35	0.23	0.23
意外度合成手法 avg	0.12	0.35	0.16	0.21	0.21
意外度合成手法 min	0.12	0.31	0.16	0.21	0.21

動画検索

- 提案手法で求められた関連語の意外度が上位である語を持つ動画を意外な動画とみなす。
- ニコニコ動画のタグがwikipedia上に存在しない語の場合、意外度を評価することができない。



まとめ・今後の課題

---まとめ---

- 意外度評価の単語集合への拡張

---今後の課題---

- 関連語の意外度を評価する際、クエリの上位語の関連語も意外度評価の対象にする。
→精度が上がったのでは？
- ニコニコのオントロジーを用いる
→動画検索の精度が上がるのでは？
- 具体的なツール開発を行う

