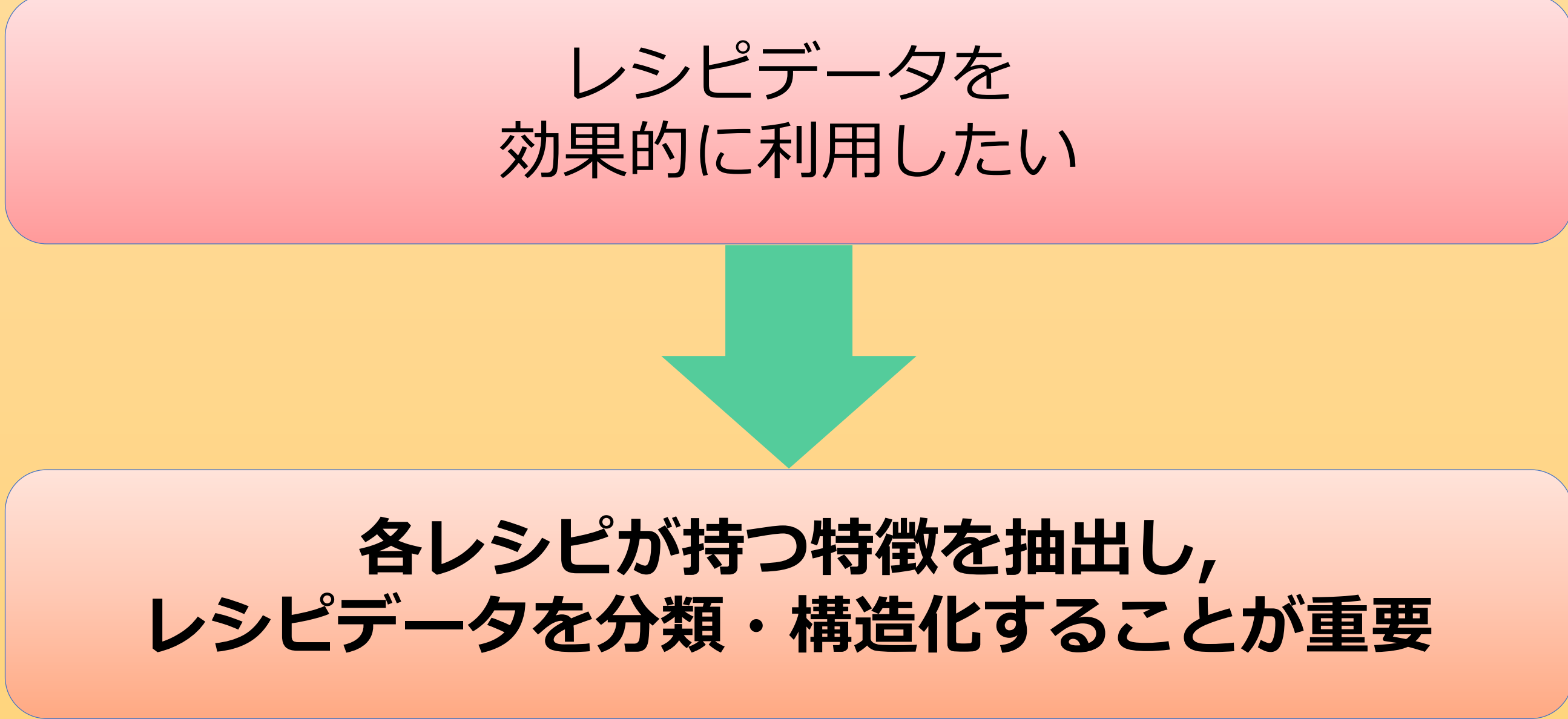


レシピフローグラフによる分散表現の獲得

二宮 あかり, 尾崎 知伸(日本大学)

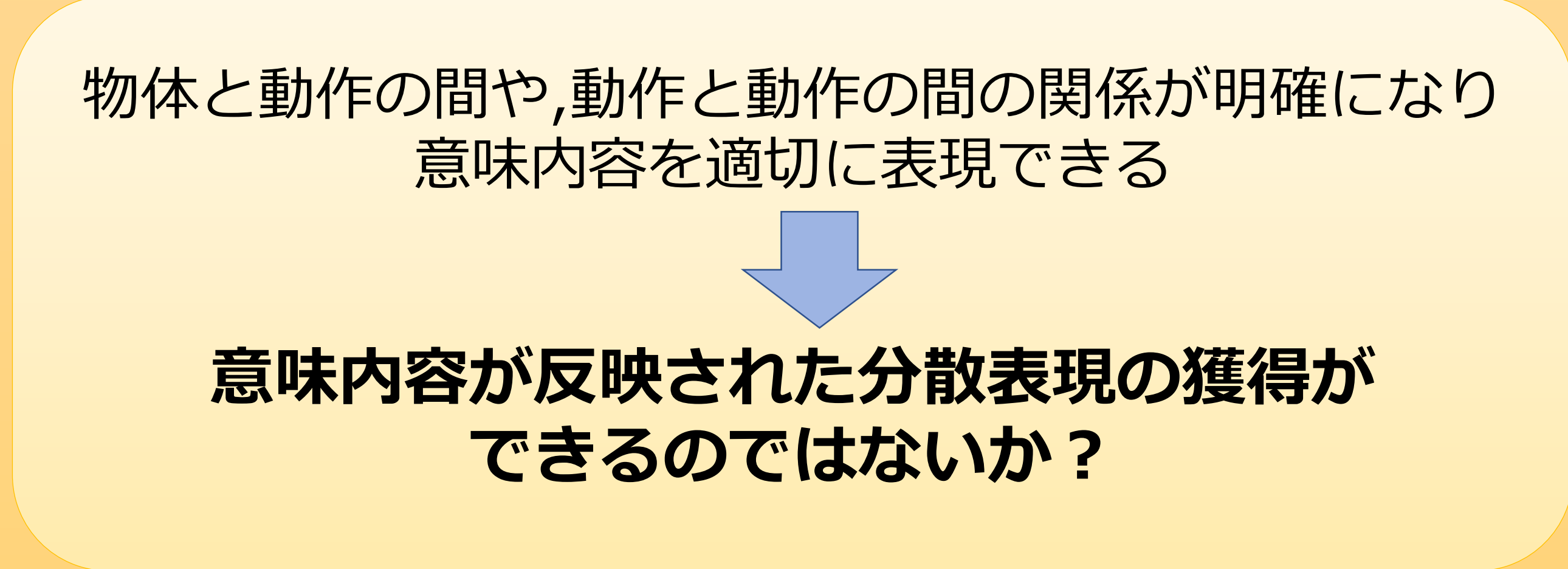
背景



レシピフローグラフ

『手順文書からの意味構造抽出』
前田浩邦, 山肩洋子, 森信介 人工知能学会論文誌, Vol32, 2017

- 調理手順の意味内容を構造化し, 無閉路有向グラフで表現.
- 頂点は食材や動作を表す単語列. 頂点間での関係が認められる場合に枝を張る.

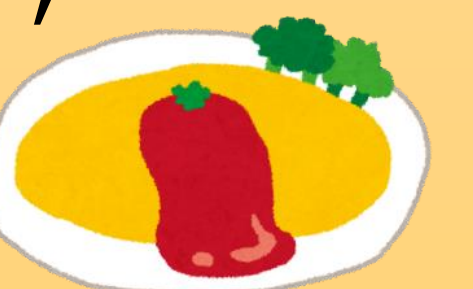


対象データと頻出部分グラフ

クックパッドのデータセットと山肩らのフローグラフのコーパスからレシピのタイトルが「オムライス」, 「コロケ」, 「チャーハン」, 「ハンバーグ」で終わる**レシピ 約33,000件**

抽出できた**頻出部分グラフ 約21,500個**

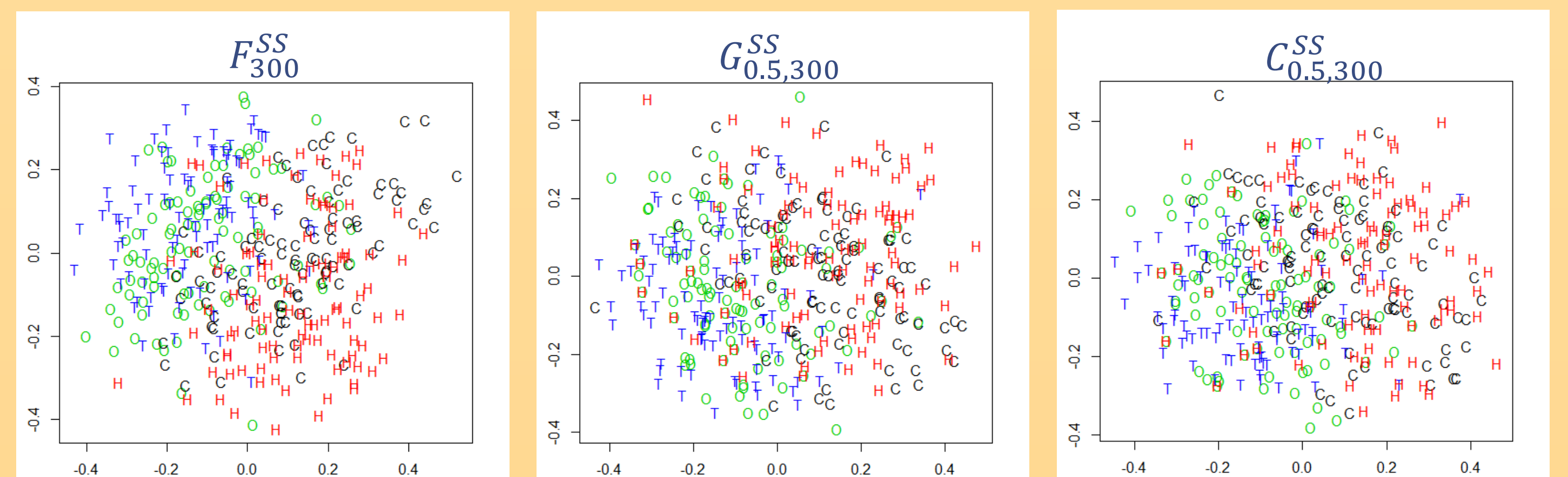
ここから, パラメタ $\delta = \{0.5, 1.0\}$ としたフリーパターン, 飽和パターンを抽出し, 分散表現獲得に利用する.



実験

※次元数は100, 200, 300の3種

• 多次元尺度法による視覚化



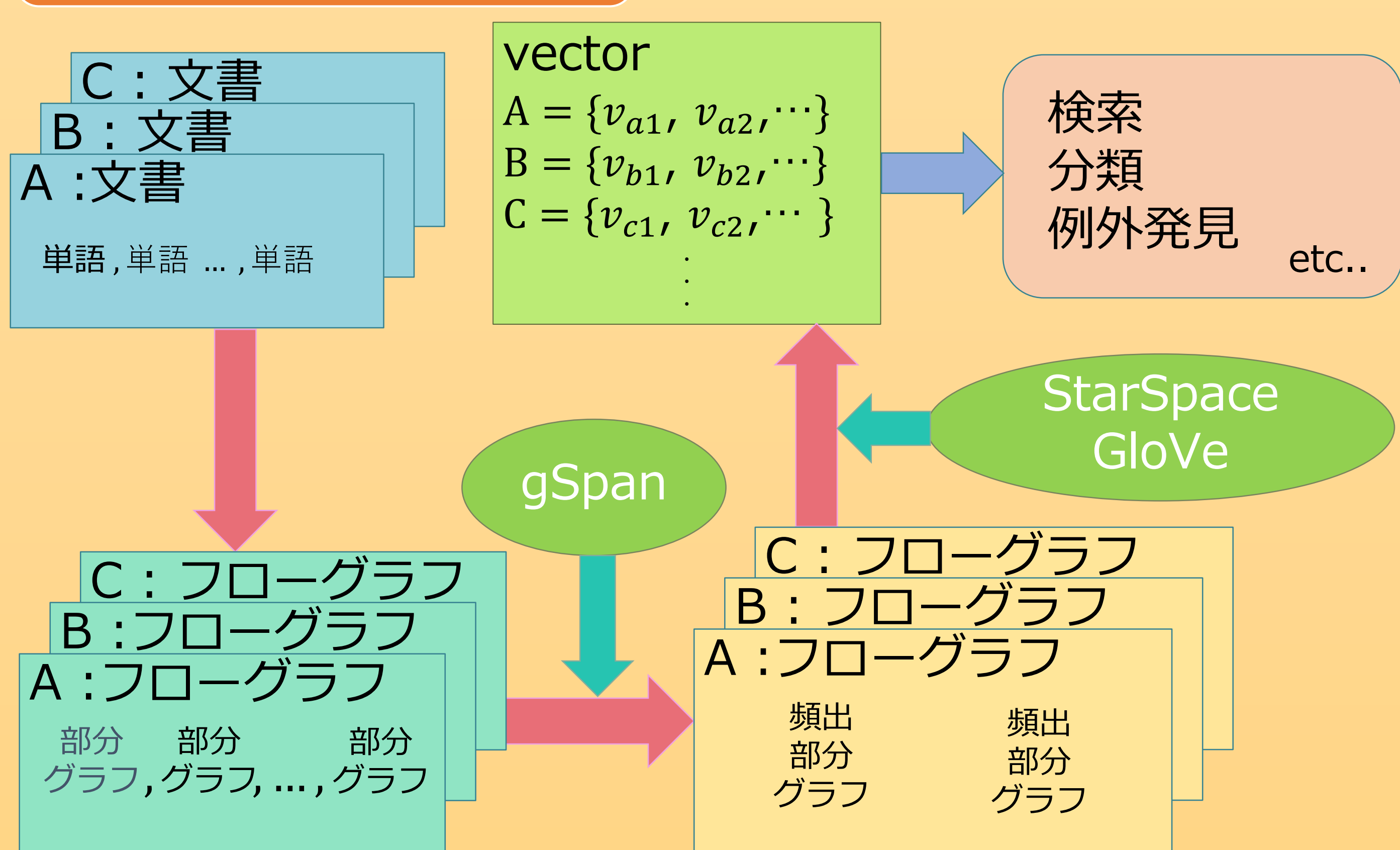
• Jaccard係数の算出

$F_{300}^{SS} \cdot G_{0.5,300}^{SS}$	0.271	$F_{300}^{GV} \cdot G_{0.5,300}^{GV}$	0.175
$F_{300}^{SS} \cdot G_{1.0,300}^{SS}$	0.265	$F_{300}^{GV} \cdot G_{1.0,300}^{GV}$	0.151
$F_{300}^{SS} \cdot C_{0.5,300}^{SS}$	0.222	$F_{300}^{GV} \cdot C_{0.5,300}^{GV}$	0.119
$F_{300}^{SS} \cdot C_{1.0,300}^{SS}$	0.151	$F_{300}^{GV} \cdot C_{1.0,300}^{GV}$	0.010

• SVMによる料理種の種類

	F_{300}^{SS}			F_{300}^{GV}		
	P	R	F	P	R	F
オムライス	0.61	0.41	0.49	0.23	0.01	0.02
コロケ	0.86	0.80	0.83	0.81	0.43	0.56
チャーハン	0.78	0.88	0.83	0.62	0.71	0.66
ハンバーグ	0.85	0.90	0.87	0.59	0.88	0.70

分散表現の獲得



文書を単語の集合としてとらえるように、**フローグラフを部分グラフの集合**としてとらえる

頻出部分グラフのみをレシピフローグラフの構成要素とする
⇒ **他のレシピとの共通点を取り入れた分散表現の獲得**

まとめ・今後の課題

- 視覚化の分布では差異が見られなかったが, Jaccard係数によってそれぞれの手法によって異なる分散表現を得ていることがわかる.
- GloVeでの結果は外れ値が大きく出やすく, StarSpaceに比べると分類の精度が劣る結果となった.

- ✓ 代替食材などの異なるタスクを用いた評価
- ✓ グラフ分散表現獲得手法といった他の手法との比較
- ✓ 対象データの拡大
- ✓ 時系列, 表記ゆれ, ノードの中身の考慮
- ✓ 画像, 手順文書等の他データの併用