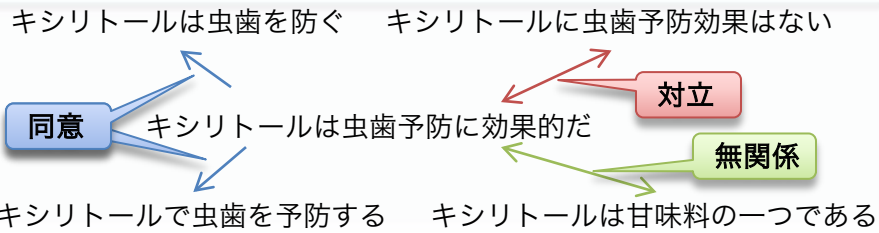


文間関係認識における局所構造アライメントの有効性

水野淳太^{†1,†2} 増田祥子^{†3,†1} 村上浩司^{†1} 渡邊陽太郎^{†2,†1} 乾健太郎^{†2,†1}
^{†1} 奈良先端科学技術大学院大学 ^{†2} 東北大学 ^{†3} 大阪府立大学

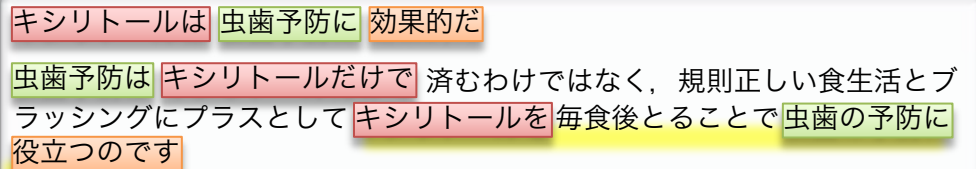
文間関係認識

- 1組の文対が与えられたとき、一方が他方に対してどのような関係にあるかを同定・分類する課題

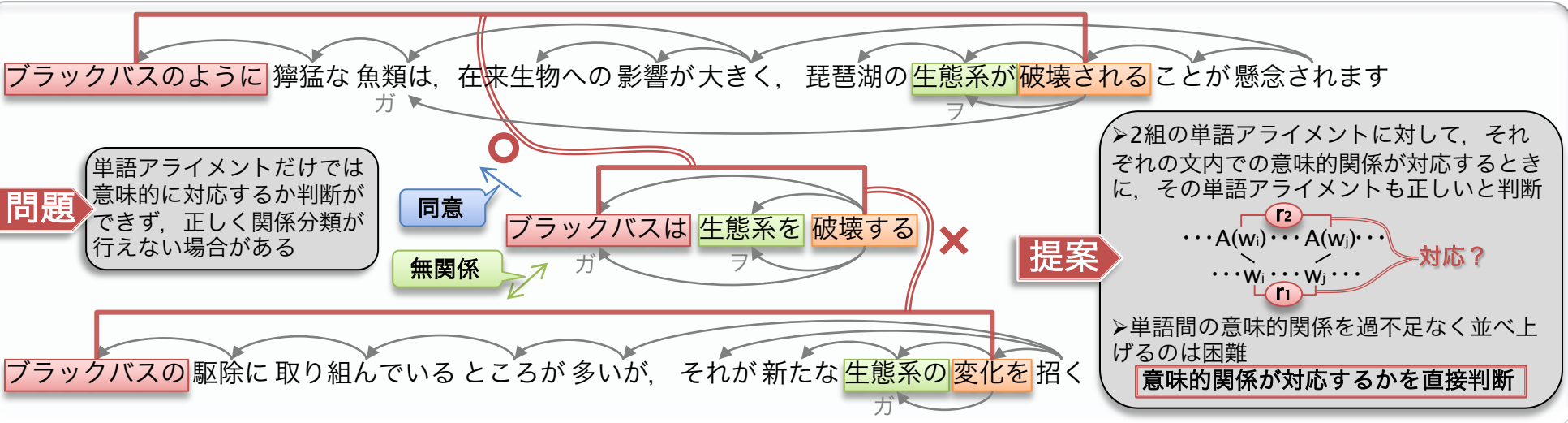


- 多文書要約や情報アクセスに有用
 - ワークショップTAC[Iftene+09]をはじめ、広く研究されている [Hickl+05, MacCartney+06,08, Glickman+05 etc.]

- アプローチはTransformationとAlignmentに大きく分けられる [Mirkin+10]
 - Transformation: 受動態 ↔ 能動態, 上位語 ↔ 下位語などの変換を行い、一方の文が他方の文に変形可能かどうかを判断
 - 多様な関係への分類は困難
 - Alignment: 文間で類似・関連する単語間に対応付けを行い、文間のどこに注目することが関係分類に有効かを判断 [Sammons+09]
 - 基本的なアプローチ [Hickl+05, MacCartney+08 etc.]
 - 2文を解析 → 文間で類似する単語へ対応付け(単語アライメント) → 関係分類



局所構造アライメント



1. 文ごとの解析

- 係り受け解析 [Kudo+02]
- 述語項構造解析 [Watanabe+10]
- 拡張モダリティ解析 [江口+10]
 - 真偽判断(成立～不成立)が述語の対立関係の認識に重要
- 評価極性判定
 - 対極性にある単語は対立関係を示唆するが多い
 - 日本語評価極性辞書[小林+05, 東山+08]を用いて判定

2. 単語アライメント

- 表層的な類似度
 - WordNet[Bond+09]やALAGINなどの語彙知識に基づく類似度
- バイオエタノール - 混合燃料
 イソフラボン - サプリメント
 守る - 維持する
 予防する - 防ぐ

3. 単語アライメントの動的推定

- 構造的に類似する単語に対応付け
 - 語彙知識不足やドメインに依存した単語に対して有効
- ブラックバスは生態系を破壊する
 ブラックバスは生態系を 変化させる
- 係り元の文節や項が2文節(項)以上単語対応する場合

4. 局所構造アライメント

- 対応を判断する4文節を入力し、対応するか二値分類
- ブラックバスのように...琵琶湖の生態系が破壊される...
- ブラックバスは生態系を破壊する
- SVM
- 単語間にいくつ係り受けが存在するか
 - 述語項の関係かどうか
 - 品詞・表層 etc.

文間関係認識における有効性

- 文間関係認識における提案手法の有効性を検証するために、4種類のアライメント結果を関係分類に適用・比較

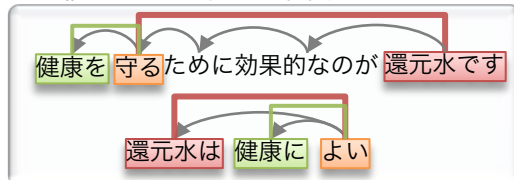
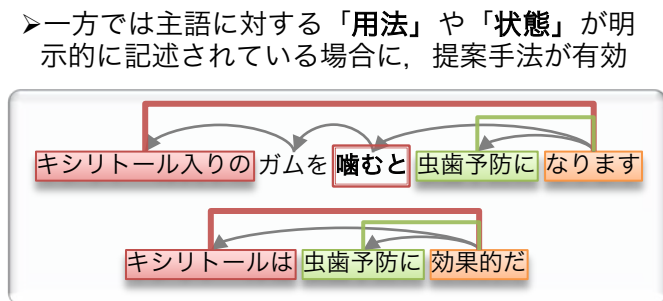
- 実験設定
 - 「同意/対立/その他」の関係が付与された966文対(367:210:389) [村上+10]
 - 5種類のトピックからなる検索質問をTSUBAKIで検索して獲得
 - 局所構造アライメントの正解データは作業員1名によって付与
 - 関係分類器



- 結果
 - [Baseline1-Proposed] 提案手法の有効性を確認
 - [Baseline2-Proposed] 2単語間の意味的關係の対応を判断するために、依存構造、述語項構造だけでは不十分
 - [Proposed-Upper Bound] 提案手法の性能向上は関係分類にとって重要

エラー分析

- 単語アライメントが困難 (バイオエタノールは)環境によい - カーボンニュートラルが健康維持に効果的だ - 健康効果を認定する
- 文構造、特に語順が大きく変化する場合に、局所構造対応の判断が困難



実験設定	同意			対立		
	P	R	F	P	R	F
Baseline1	54.23	41.96	47.31	58.97	10.95	18.47
Baseline2	48.13	62.94	54.55	61.73	23.81	34.36
Proposed	58.44	49.05	53.33	67.09	25.24	36.68
Upper Bound	63.41	95.37	76.17	88.98	53.81	67.06

- Baseline1: 単語アライメントのみで動的推定も行わない
- Baseline2: 2単語間の構造が依存構造または述語項構造である場合のみ局所構造対応 [Chang+10]のedge-mappingに相当
- Proposed: 提案手法
- Upper Bound: 局所構造アライメントの正解データを利用

今後の課題

- 提案手法の改善
 - 語彙知識の追加
 - 素性の改善
 - グラフカーネルなど、機械学習手法の改善
- 多様な文対に対するアライメントの正解付与
 - 正解付与基準や付与済みデータの公開
- TAC RTE(英語)など他言語への適用