E-mail: zushinaoya@gmail.com

商品レビューデータを用いた日常的に経験するにおいの収集・分類

○図師　直弥[1]，竹内　元気[1]，小川　緑[1]，後藤　なおみ[2]，掛谷　英紀[1]，小早川　達[2]，綾部　早穂[1]

1. 筑波大学　　2. 産業技術総合研究所

背景

日常的に経験するにおいは文化や時代によって異なる

食習慣や好まれる香水など，日々接するにおいは文化によって異なる。
また，例えば同じ日本であっても，食のグローバル化や香害という概念の出現など様々な要因によって日常的なにおいは時代によって変化している。

嗅覚能力の検査に最適なにおいの選定

特定の化学物質の検知能力を測定する検査であれば世界標準による測定が可能
しかし，日々接するにおいのほとんどは複数の化学物質の化合物であり，
特定の化学物質の検知能力は，日常的に必要な嗅覚能力とは一致しない。

→日常的に必要な嗅覚能力を測定する嗅覚検査では，その文化や時代に適したにおいを選定し，それについての検知能力を測定するべき。

本研究の目的

Web上に投稿されたテキストデータから
現代日本人が経験するにおいデータを収集，分類する方法の提案

方法

① “楽天市場データセット”からにおい表現を含むレビューの抽出

Mecabとシステム辞書mecab-ipadic-NEolog を用いてレビューデータを形態素解析
“の/のような/みたいな” + “におい/ニオイ/臭い/匂い/かおり/カオリ/香り”
を含むレビューのみを抽出

② 抽出したレビュー文からにおい表現の直前の名詞を抽出

“これはとてもいい木の香りがする。”
↓
“これ / は / とても / いい / 木 / の / 香り / が / する。”　→におい単語”木”を抽出

③ 表記のみ異なる名詞は出現数の多い表記に統一し，出現回数をカウント

e.g. “ひのき”と”ヒノキ”と”檜”は”ヒノキ”に統一；バラ(薔薇)とローズは区別

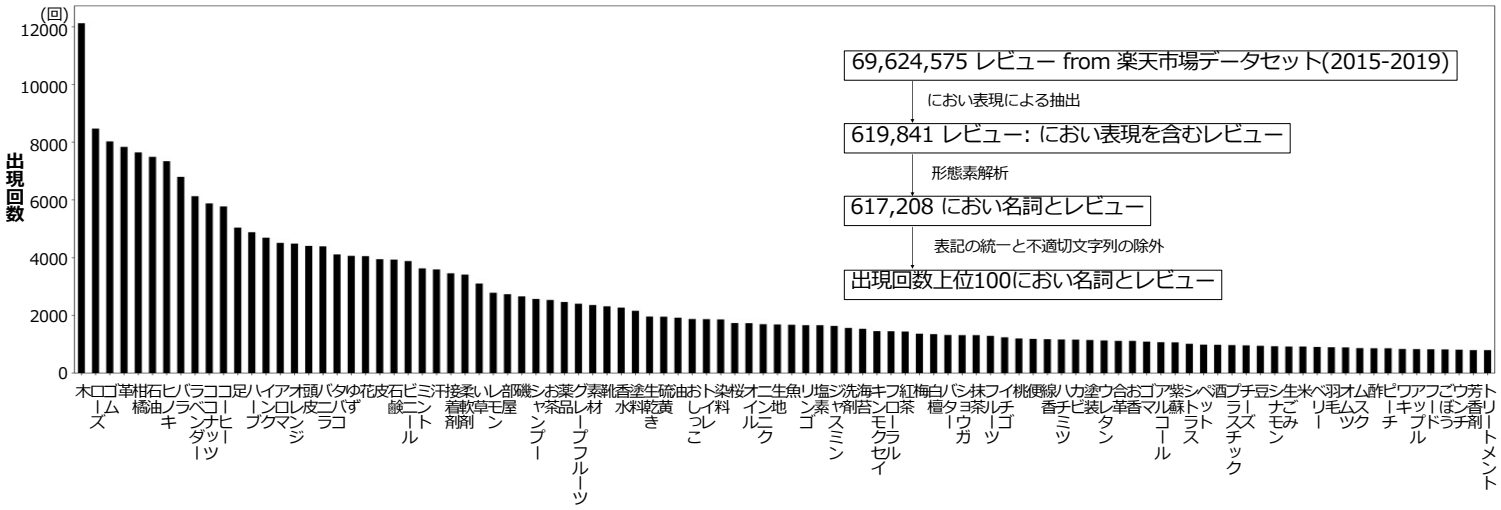
④ 具体物を表す名詞でない文字列を除外

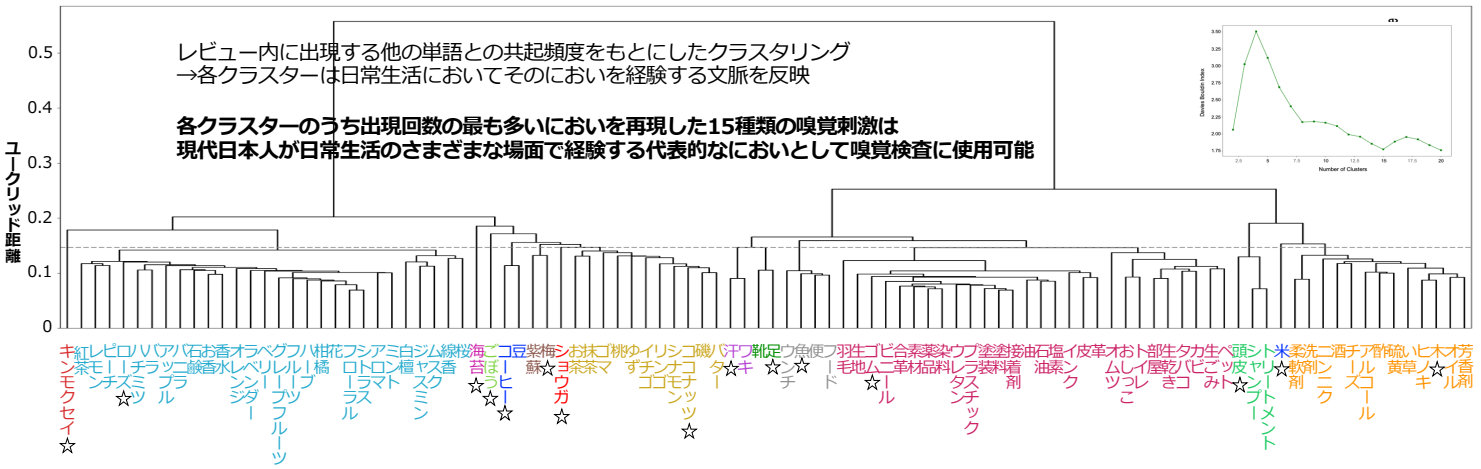
e.g. “独特”，”好み”，”こ”，”そ”....計27個を除外

⑤出現回数上位100単語についてクラスター分析

レビュー内の他の名詞との共起頻度の特徴ベクトルによるにおい単語間のユークリッド距離を算出し，Davies Bouldin Indexを参照して15クラスターに分類

結果





本研究の成果

地域やデータ数の限定される対面調査と比較して，日本全国から2015-2019年に投稿された大規模なにおい経験データを収集，分類することができた。

今回使用したデータセットは，におい経験に関する記述を意図的に求めたものではなく，より自然に表出されたにおい経験を収集することができた。

同様の分析を継続的に行うことで，時代ごとに異なる日常的なにおいを知ること，その時代に適した嗅覚検査の作成に貢献できる。

限界と今後の展望

今回使用したデータセットは，楽天市場上で売買が可能な商品に関する記述に限定され，日常生活のすべての経験を網羅できるものではなかった。
→より多様なにおい経験を収集できるデータと合わせて解析を行う必要。

においい単語として不適切な文字列の除外を人為的に行なった。
→自動的にフィルタリングを行うシステムの構築によって，経時的な変化に一貫した対応が可能。