

令和2年度

人工知能研究助成交付伝達式

少量の学習データを対象としたショートテキストの自動分類

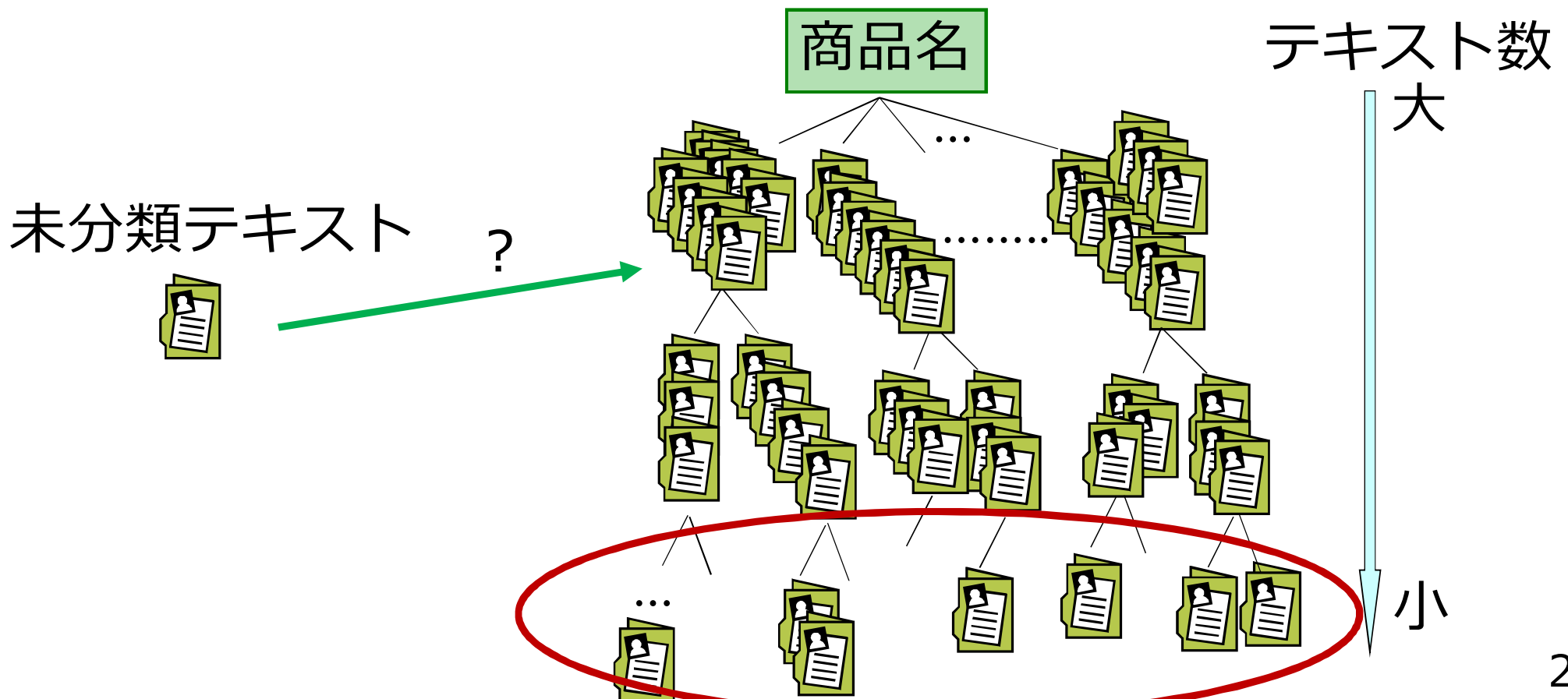
山梨大学大学院総合研究部 工学域

福本 文代



- ショートテキスト分類の問題

- ショートテキスト (タイトル, 商品名など)の情報は, 少ない
- インターネットディレクトリでは, 下位に属するショートテキスト数は, 少ない

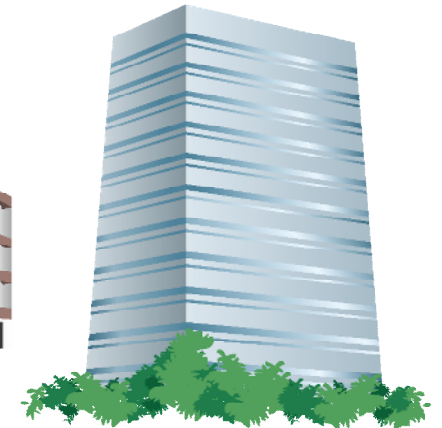
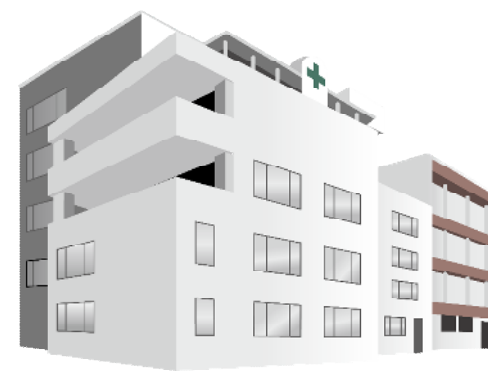


研究目的

- 商品名などのショートテキストを対象
 - 少量の学習データを対象
- > 高精度な分類手法を開発

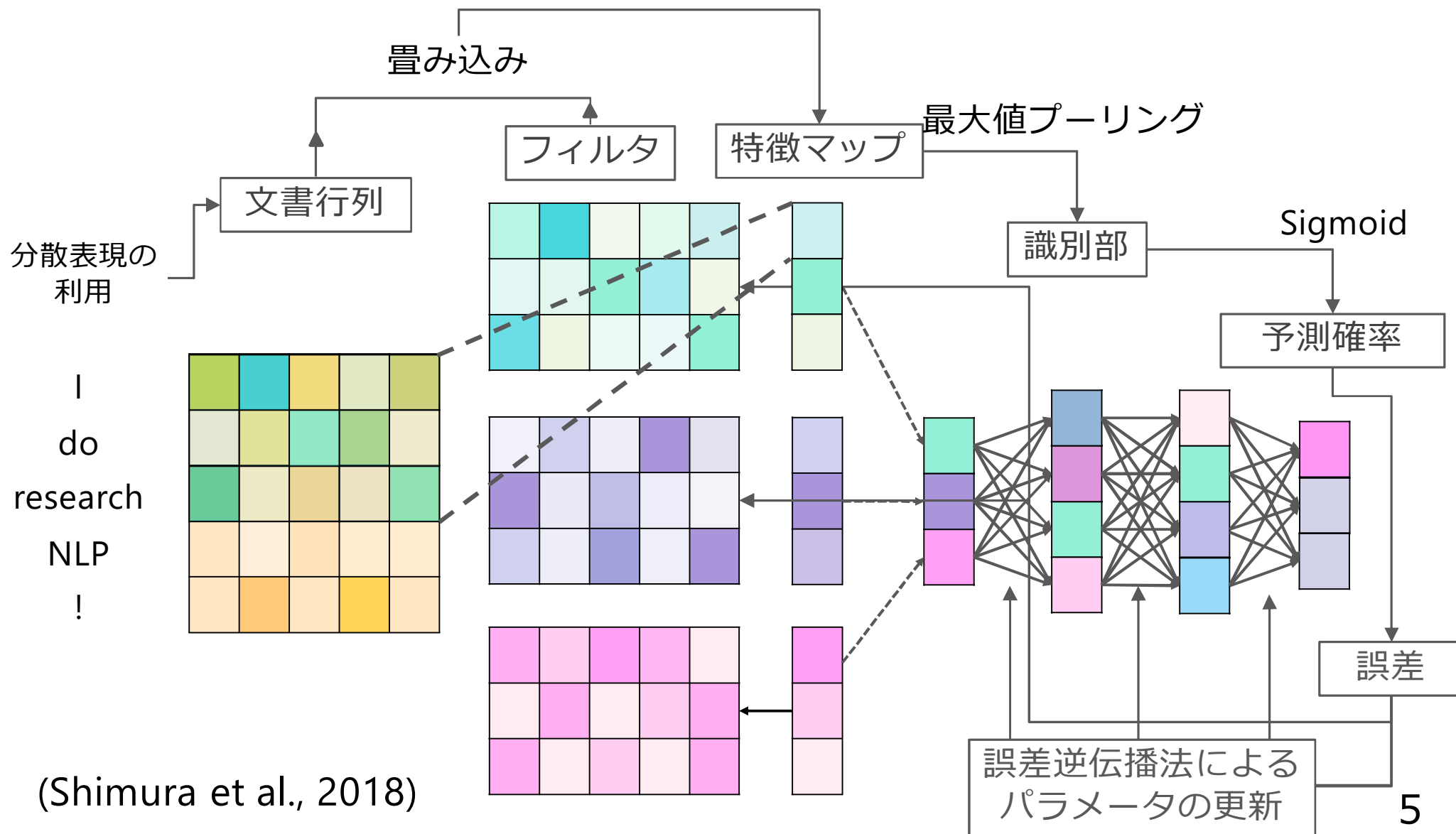
期待される効果

- インタネット検索のコア技術
- 図書館の情報検索システム
- 病院や企業での社内文書管理システム
- 各種情報配信サービスにおける情報フィルタリングとしての利用が可能.



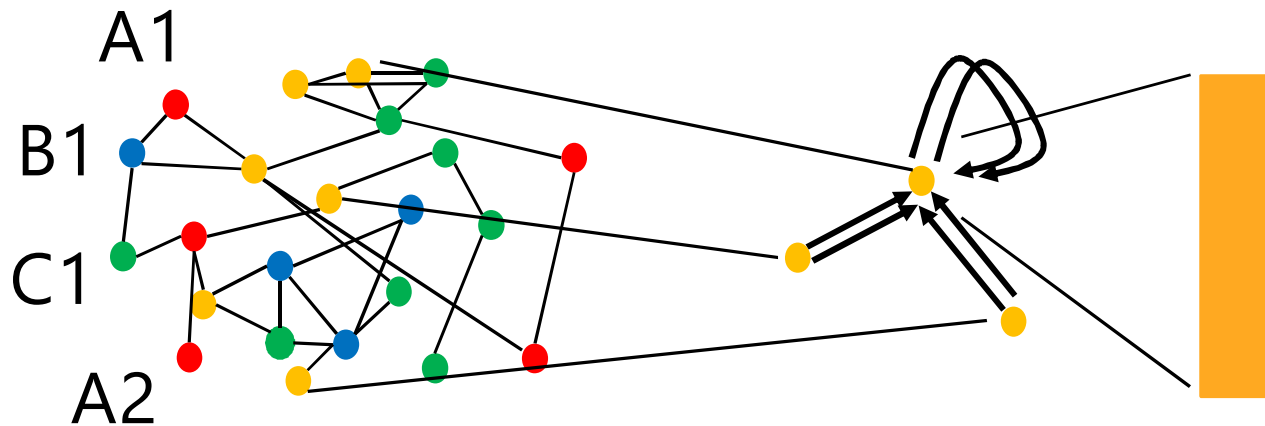
研究内容

1. 深層学習によるテキスト分類

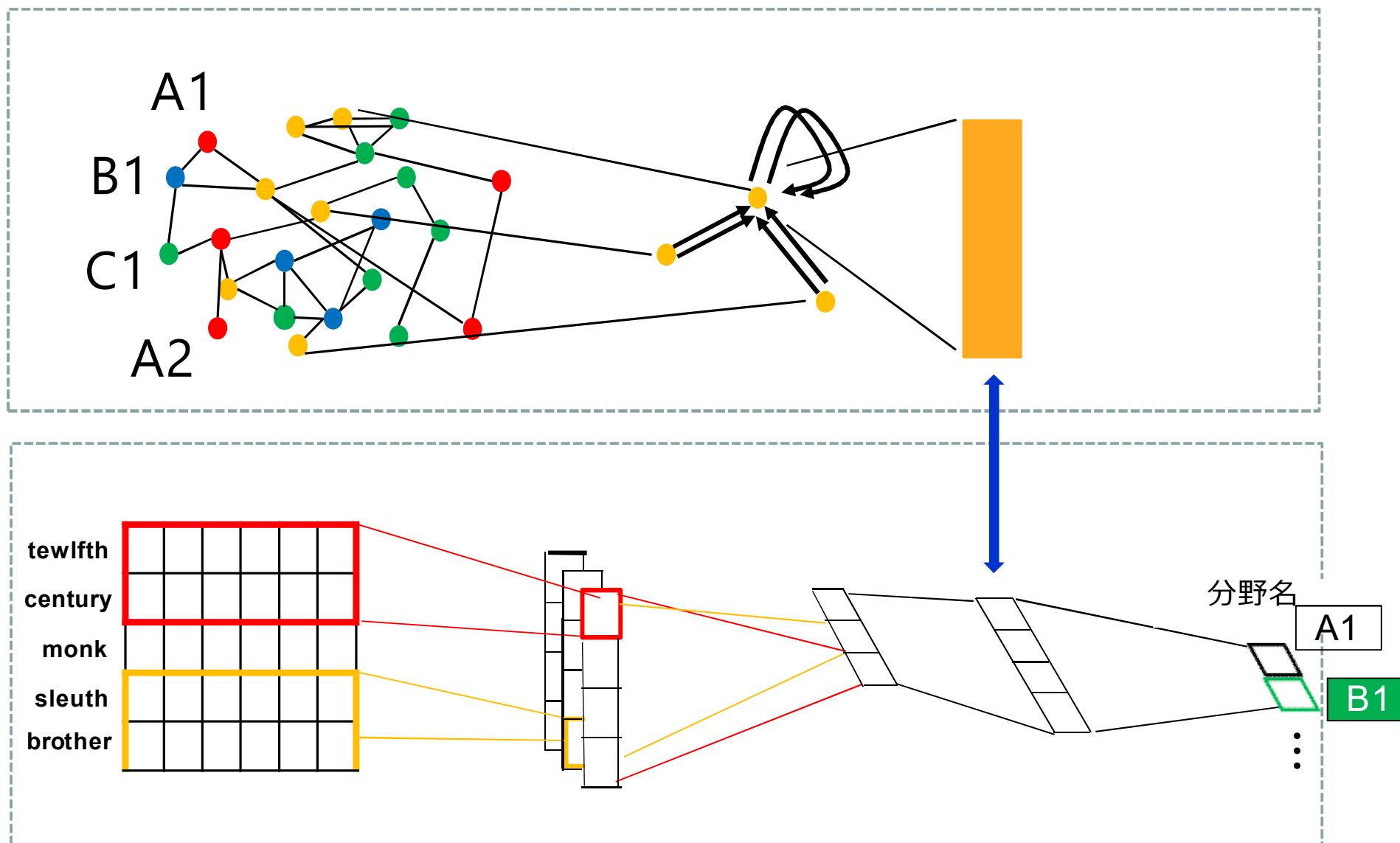


2. 分野間の意味表現学習

- Graph Attention Networks を用いた分野間の意味表現学習



3. テキスト情報と分野情報の統合, 及びテキスト分類



4. New York Timesを用いた実験

- データ, 階層, 分野数
 - 訓練データ数 : 23,345
 - バリデーシオン数 : 5,834
 - テストデータ数 : 7,292
 - 8階層, 166分野
- 比較手法
 - HFT-CNN (Shimura et al., 2018)
 - SGM (Yang et al., 2018)
 - HiLAP-RL (Mao et al., 2019)
 - HiAGM-TP (Zhou et al., 2000)