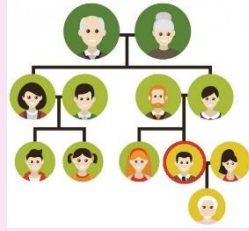


ビッグデータから社会を把握する 名前データから民族を推定する研究

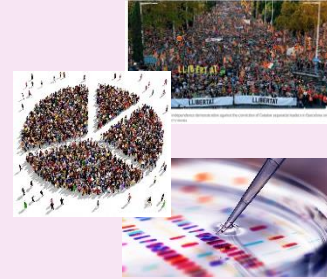
どんな研究？

名前は人間の基本的な情報インデックスであります。名前は社会、文化、時代の特徴を反映して作られているため、名前からは、その人物の色々な特徴をよみとることができます。本研究では特に苗字のデータを用いて、民族の分類モデルを構築し、社会の現象を理解しようとしています。



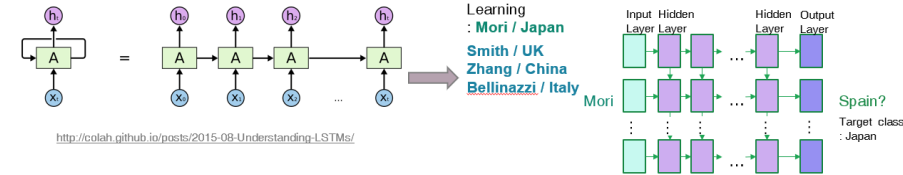
何がわかる？

- 苗字データから民族を推定できる分類モデル
- 分類に時間や費用がかかる社会構成員の把握が容易になる
- 応用分野：社会分断問題、バイオテクノロジー、人口統計など



機械学習で苗字から民族を推定

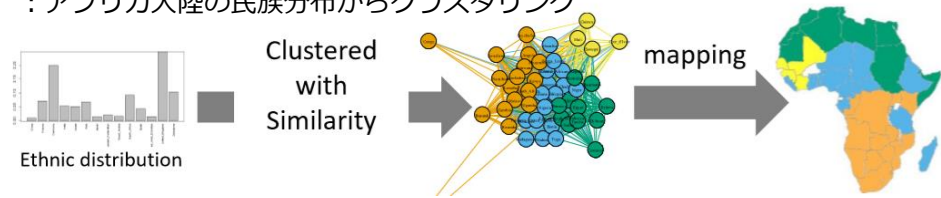
RNN(Recurrent Neural Network)で大規模の苗字データを学習し、苗字から民族を推定できるモデルを構築



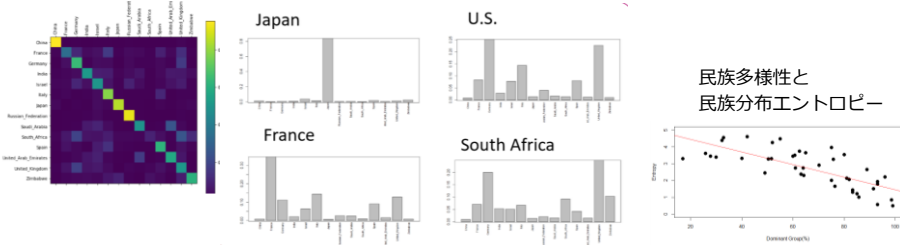
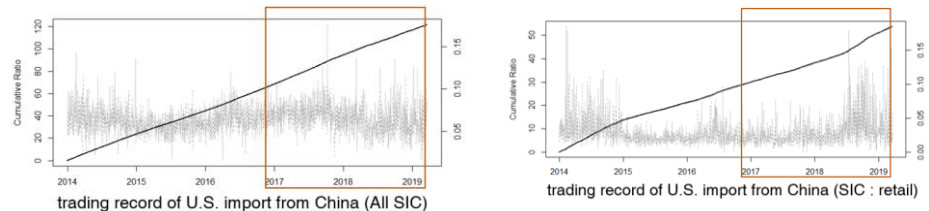
<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

研究内容

- 民族の空間的な分布
：アフリカ大陸の民族分布からクラスタリング



- 民族ネットワークの活性化
：米中貿易戦争(2017-2019年)の間、観察された民族ネットワークの活性化



苗字・民族の分類モデルを用いて各国の民族分布を把握する。
民族分布のエントロピーが高い国は民族多様性を持っている国

貿易障害の発生期間