

研究室紹介

徳山研究室 = 計算理論研究室

構成メンバー

- 教授：徳山 豪
- 大学院生は、修士1年生が5名
 - 野口 真、藤原 直紀、藤野 準平、山田 康晴、武元 祐磨
- 4年生9名
 - 全員就職決定なので、来期はM1はいません
 - 今年はオンラインが主体のハイブリッドセミナーでした
- 学科秘書： 堀口さん （研究室に常駐）

自己紹介

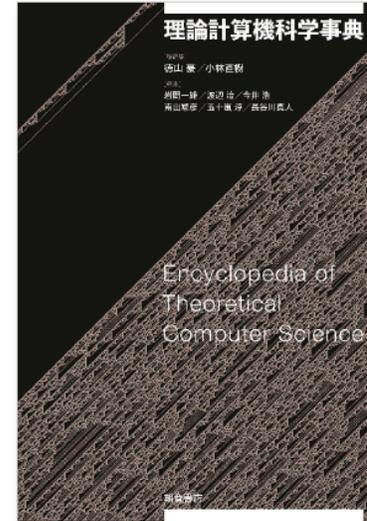
- 徳山の専門は理論計算機科学

- アルゴリズムと計算の理論
 - 「世の中の役に立つ数学」を研究しています

- 経歴

- 東大数学科博士（当時の専門は代数群論）
 - 博士号を取って1年半 河合塾で講師をして
- 日本IBM東京基礎研究所(1986-1999)
- 東北大学(1999-2019)
- 関学(2019-)
- 研究や著作やら： `takeshi tokuyama` でネットで調べてください
 - Researchmapかdblpが判りやすい
- 授業： 離散数理、グラフネットワーク理論+実習、計算幾何学でお会いしている
- 性格は、楽観派で緻密性には欠ける（のんきでうっかり者ということ）
 - 学生には優しいと思うが、学会ではどうも怖い「ボス猿」らしい

理論計算機科学事典



徳山 豪・小林 直樹(総編集)

定価 19,800 円（本体 18,000 円 + 税）

A5判 / 816ページ

発売予定日：2022年01月11日

ISBN：978-4-254-12263-3 C3504

近日発売予定です

セミナーの雰囲気

- 学生たちが共同で、でも主体性をもって学ぶ
 - プロジェクトではなく、全員個人でテーマを持つ
 - それぞれの研究から、全員が学ぶ
 - できるだけ質問することを歓迎します
- 卒研は、新しい成果が出るという事が目的ではない
 - 頑張っ、て、「新しい成果は得られなかった」でも、それでよい
 - 社会に出てから役立つ基礎を学ぶ
 - 卒論は、研究のまとめ方や学術文書の書き方を学ぶもの
- 時間：基本は週2日セミナー(対面の予定、時間は短い)
- 場所：4号館2階 W205 (広いです)
- イベント等：コロナのおかげで今年はなし。来年は？



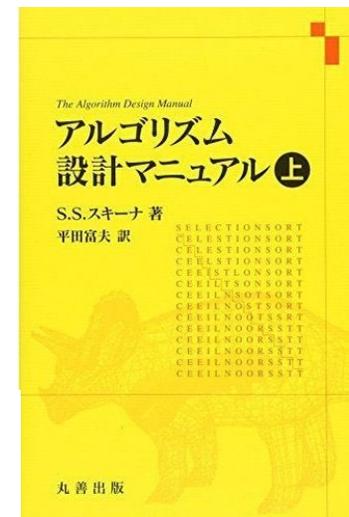
写真は2年前のもの

去年のセミナー(卒研前期)

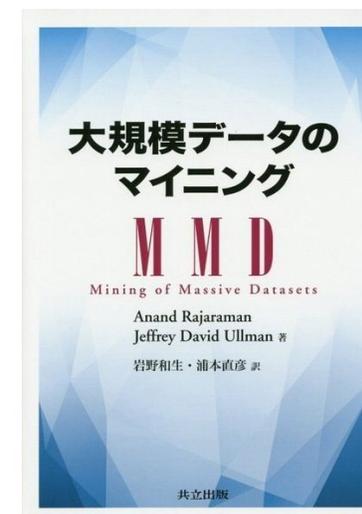
- 今年のテキスト： Pythonで儲かるAIを作る
 - 英文講読は、Programming Pearls (Bentley)



- 昨年のテキスト：アルゴリズム設計マニュアル
- 一昨年のテキストは、「大規模データのマイニング」



- 来年のテキストは、相談して決める予定



卒研後期は、各自で研究テーマの調査

- 学生の研究テーマは自由
 - テーマを自分で探すことが研究の大半です
 - 学生にとって楽ではないが、テーマ探しは良い経験になります
 - もちろん、やりたいことに対して「こんな方法や論文があるよ」というガイドはします
 - 「手を動かす」ことが大切
 - アイデアややり方をノートに書いてみる
 - プログラムしてみて、データで試してみる
 - 結果をみて、どうすればいいのか、またアイデアを考える
 - セミナーで進捗を発表して、みんなの意見を聞く
 - 今年は、データ解析・機械学習6名、最適化1名、折紙工学1名
 - どうも前期で読んだテキストに「引きずられる」学生が多い
 - 去年は、組合せアルゴリズム2名、計算幾何学2名、機械学習3名
 - 一去年は、組合せアルゴリズム1名、データ解析6名
 - 過去のテーマは、他にWEB解析、最適化、金融工学、離散数学など
 - まあ、アルゴリズム関連のテーマなら何でも対応します

「失敗しない研究」は良い研究ではない

- 「高い確率で成果が望める研究テーマ」は、お金と人員が多い方が勝ち→ 中国などに勝てない、歴史にも残らない
- 優れた研究（仕事）は好奇心と失敗から。他人がやらないこと、自分がやりたいことをやろう（若者の特権です）
 - コペルニクスやアインシュタインやメンデルやダーウィンやチューリングや…
 - 江崎玲於奈（1973年ノーベル物理学賞：トンネルダイオードの発見）
 - ソニー（当時は小さい会社）で半導体の研究。トランジスタの不良品ができてしまう。なぜ不良品になる？
 - 半導体を薄くしすぎて、今までにない物理現象が起きていた→トンネル効果
 - 白川秀樹（2000年ノーベル化学賞：導電性高分子の重合構造の解明）
 - 学生がm gとgを間違えて千倍の試薬を使って実験したら、大失敗だが、不思議な膜ができた
 - これは何だろうと好奇心で調べると→導電性高分子薄膜（捨てちゃダメ!）
 - 小林誠・益川敏英（2008年ノーベル物理学賞：素粒子の新理論）
 - 理論が実験に合わない。仮想的に、3つしかないクォークの数を6つに増やしてみるとどうなるかなあ（好奇心）。
 - まあ、そんなのは実際あり得ないよね。20年後に6つのクォークが存在することが実見される
 - 山中伸弥（2012年ノーベル賞医学・生理学賞：IPS細胞の発現）
 - 山中先生「根性と体力で、できるまでやれ！」→ 研究員の高橋さん「そんなの無理」と非常識な発想で実験を行う
 - 真鍋叔郎（2021年ノーベル物理学賞:地球温暖化理論の提唱と推進）
 - 最新のIBMコンピュータが使える米国気象学研究所の研究員に採用された(1958年)。興味本位でいろいろシミュレーションをやってみよう
 - もし空気中の成分が変化したら、気象はどうなるだろうか？ 二酸化炭素を増やすシミュレーションをすると…
 - 情報科学でもたくさんそういう例があります。今のGAFAの創業者は皆そうです。
- こんな幸運は期待はできないが、したがって、「安全なテーマを指導教員が与える」ことは、できるだけしない。
- 研究は（人生も） 運・鈍・根 といいます。「頭が良い」必要はありません。

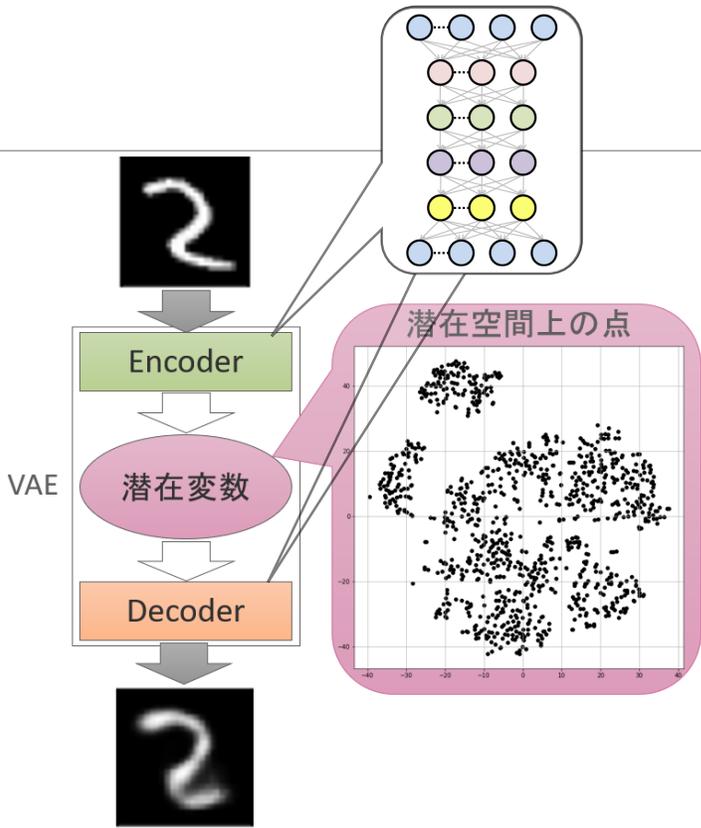


徳山 江崎 白川

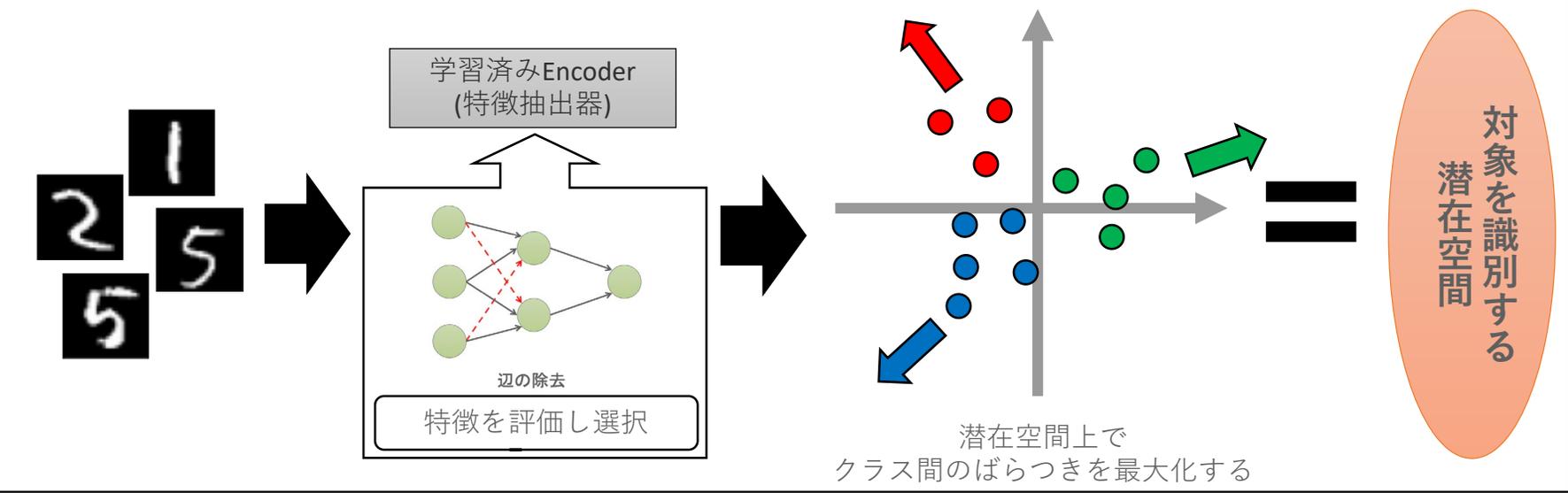
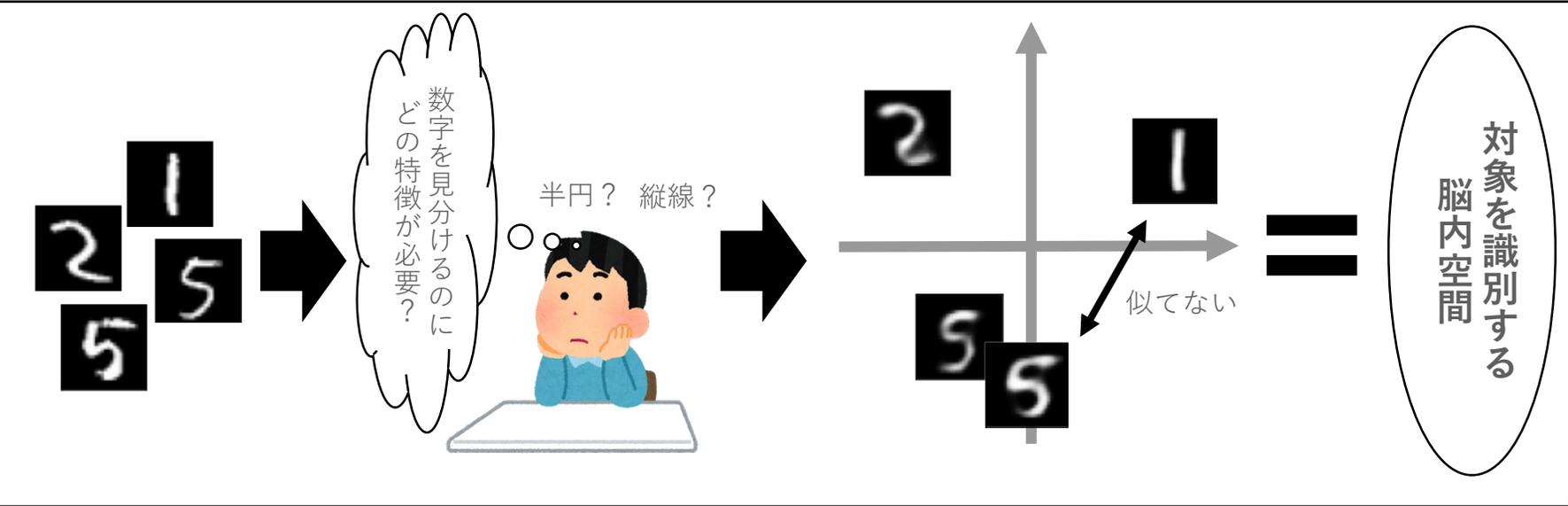
M1の学生の研究テーマは

実は、徳山の現在の研究テーマで研究している学生はいません

深層学習モデルVAEを用いた，分類問題に対する脳の「思考」のモデル化 (山田君)



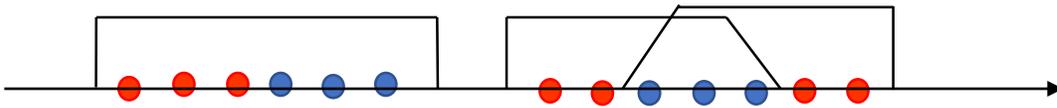
図：VAEの構造
(教師なし深層学習モデル，特徴抽出と生成が可能)



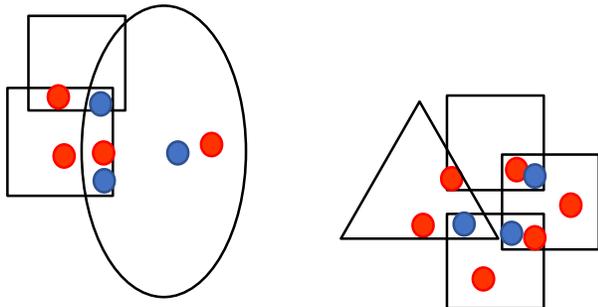
図：「思考」のモデル化

最大差赤青セットカバー問題 (藤原君)

赤点と青点の数の差の数を最大にするようなカバーリングを求めよ



一次元：領域は区間 (多項式で解ける)

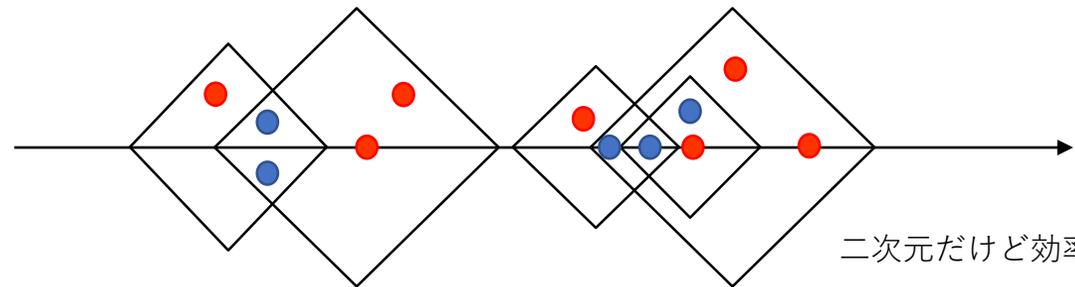


二次元：領域は任意の平面図形 (多項式では解けない)

- 二次元以上では単一正方形領域に限定したとしても, 強NP-困難であることが分かっている.

- インスタンスで与えられる領域に制限を持たせることで, 効率的に解くアルゴリズムを構成できないか?

- たとえば領域集合をL1ノルムでの円集合に限定し, 各円の中心が同一の直線状に乗るという制約では, 効率的に解くアルゴリズムが存在する.



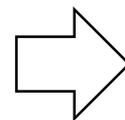
二次元だけど効率的に解ける!

深層学習モデルBERTを用いた記述式問題の自動採点 (藤野君)

- 共通テスト(センター試験)に国語記述式問題が導入

- 約60万人の解答を短時間で採点しなければならない
- 採点者によって点数に偏りが生じてしまう

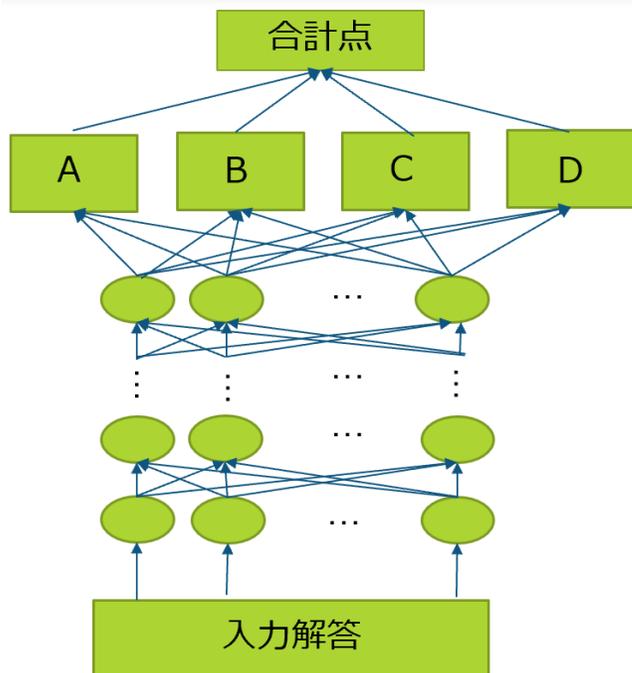
問題



採点を自動化

高速かつ採点に一貫性をもたせる

深層学習モデルBERT



<学生の解答例>

「他人を自分とは異質な考え方をもつ人間と見なす西洋では、自分の意見に同意を得るために、言葉を尽くして他人を説得する技術が培われたということ。」

部分点：真

A : 2点

B : 5点

C : 6点

D : 3点

合計点：16点

部分点：モデル予測

A : 2点

B : 5点

C : 5点

D : 3点

合計点：15点

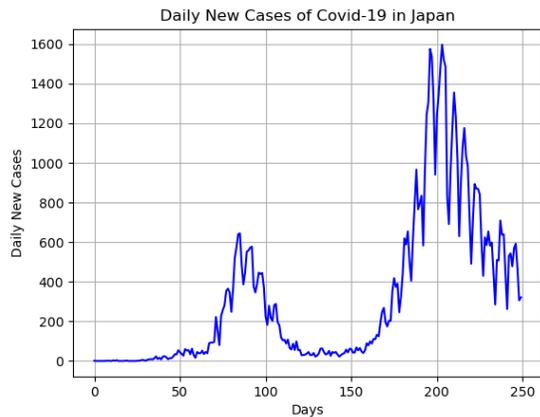
現状

自動採点の精度はそこそこ高いが、文脈を変えたり否定語を導入した場合については採点が難しいことがわかっている (今後)

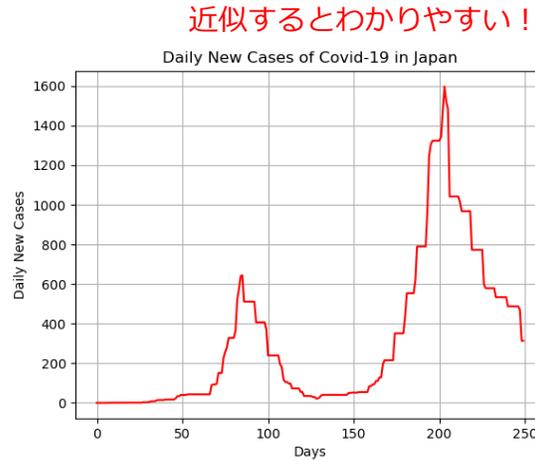
時系列データの最適な簡略化とその応用 (野口君)

1次元の時系列データを曲線として扱おうと...

→ 凸包の近似アルゴリズムが利用できる。



近似



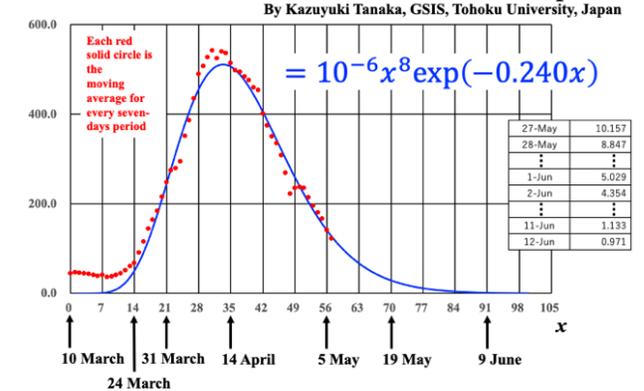
More



他の近似方法は？

- 感染症数理モデルの構築
- 解析的な関数での近似...

New Case Estimations for COVID-19 in Japan



感染拡大率を考慮した近似例

(Private communication with Prof. Kazuyuki Tanaka, Tohoku University)

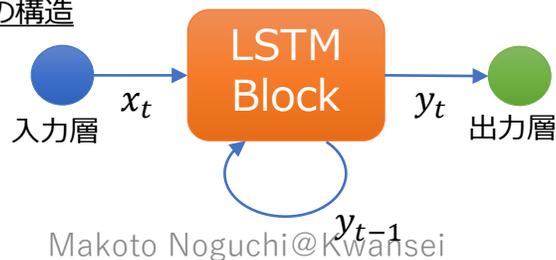
More



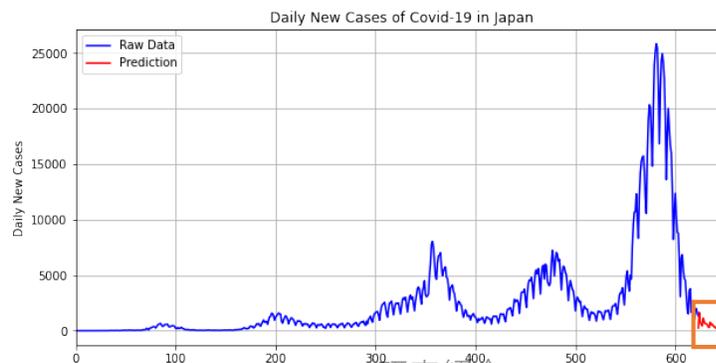
近似データの応用

- 深層学習を用いた予測
- Ex. LSTMによる予測

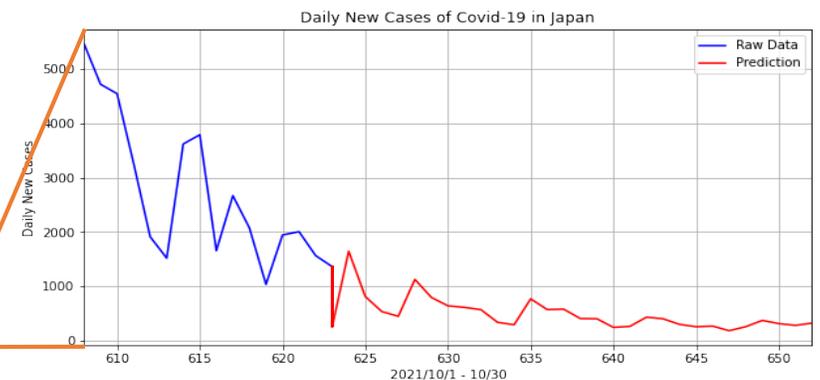
LSTMの構造



Makoto Noguchi@kwansai



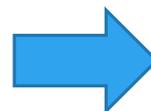
研究紹介



マイクロクラスタリングを用いた楽天データの解析(武元君)

近年の情報化による問題

情報量が膨大となり、そのデータの構造を理解するのが難しい



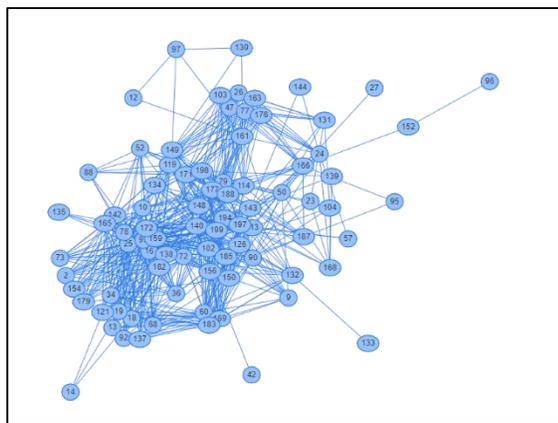
データがどのような構造
になっていればいい？



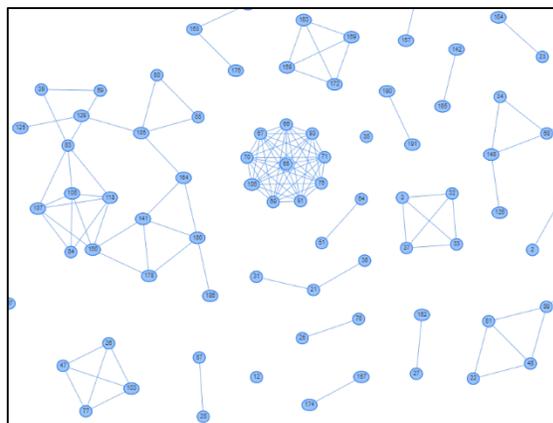
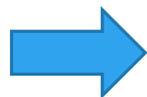
グラフ研磨という手法を用いて、ノイズの少ない明確な構造のグラフへと変換

マイクロクラスタリングとは

グラフ研磨(各ノードの類似度を用いてグラフを再構成すること)を行い、得られたグラフで極大クリーク列挙を行うこと



グラフ研磨



構造を明確にすることによって、
あるノードが作成したグラフでは
どのような関係を持ち、
どのグループに含まれるのかを
理解しやすくすることができる。

二年生（領域実習A）へのガイド

- 領域実習は、輪読形式です
 - 今のところ、対面指導の予定
 - アルゴリズムに関するパズルをみんなで考える
 - 計算理論と数理、問題解決の手法を学ぶ
 - 当番制で、学生が解説します
 - 領域実習Bは、よりプログラミングに近いパズル
- 大学院進学志望の学生への注意
 - 徳山は、後3年で退職します。来年の卒研究生は大学院修了まで指導できますが、来期領域実習の現2年生は大学院では指導できません



興味のある学生は、

- 卒研希望学生は、メールで連絡ください
 - 研究室（教授室にいるかもしれないが、話は研究室で）に来れば、お話しできます（できるだけまとまってきてください）
 - 12月20日－23日 は、時間が取れます。
- 領域実習は自動的に配属が決まるので、事前連絡は不要
 - ただし、もちろん見学はできます