

教授 山本 章博

助教 小林 靖明

代表教員研究室：総合研究7号館 324室

Web : <http://www.iip.ist.i.kyoto-u.ac.jp/> Email : akihiro@i.kyoto-u.ac.jp

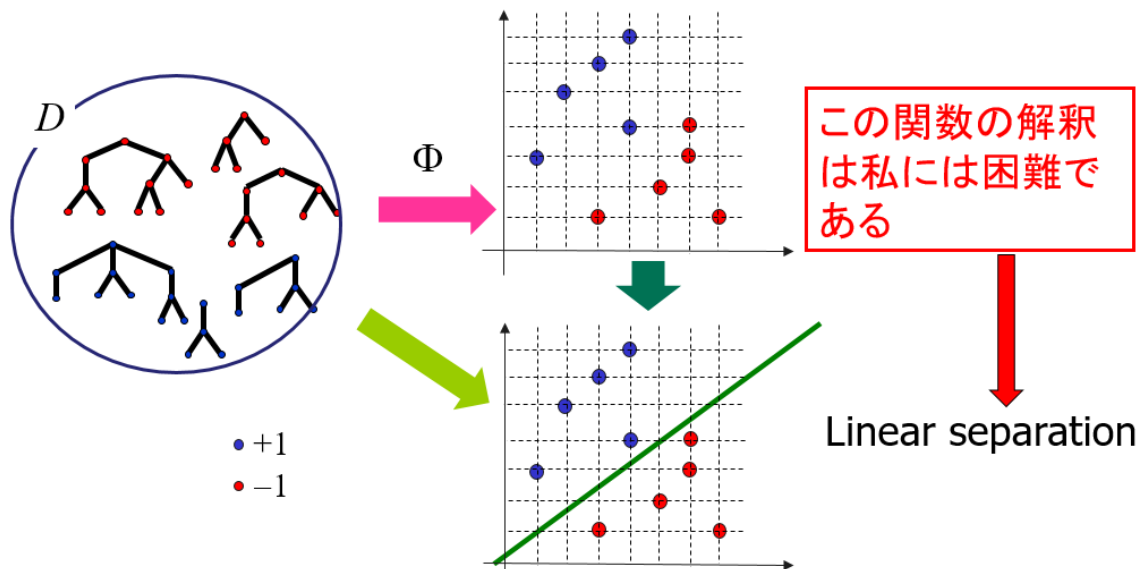
当研究室では、**機械学習・知識発見**を中心にして人間の高次推論機構の性質を解明し、またそれらを用いて、与えられたデータから適切な情報を取出すための計算機構を構築することを目標に研究を行っている。データ構造、最適化手法、形式言語理論、グラフ理論を応用し、**数理論理・計算数学**との関係の解明へと展開し、知能情報学における新たな基盤の構築を目指している。

研究のこだわり所

- **解釈可能な機械学習**
 - 数学・数理論理は「**解釈可能な正しい説明**」を求める学問
 - 機械学習は「**正しくかつ解釈が可能な結果**」を求めるべき
- 複雑な**データ構造**に着目した機械学習
 - **説明**は構造を持っている(自然言語の構文解析木など)
 - 半構造化文書(web pageなど), 自然言語文書の構文解析木, 物語の展開

研究の方針

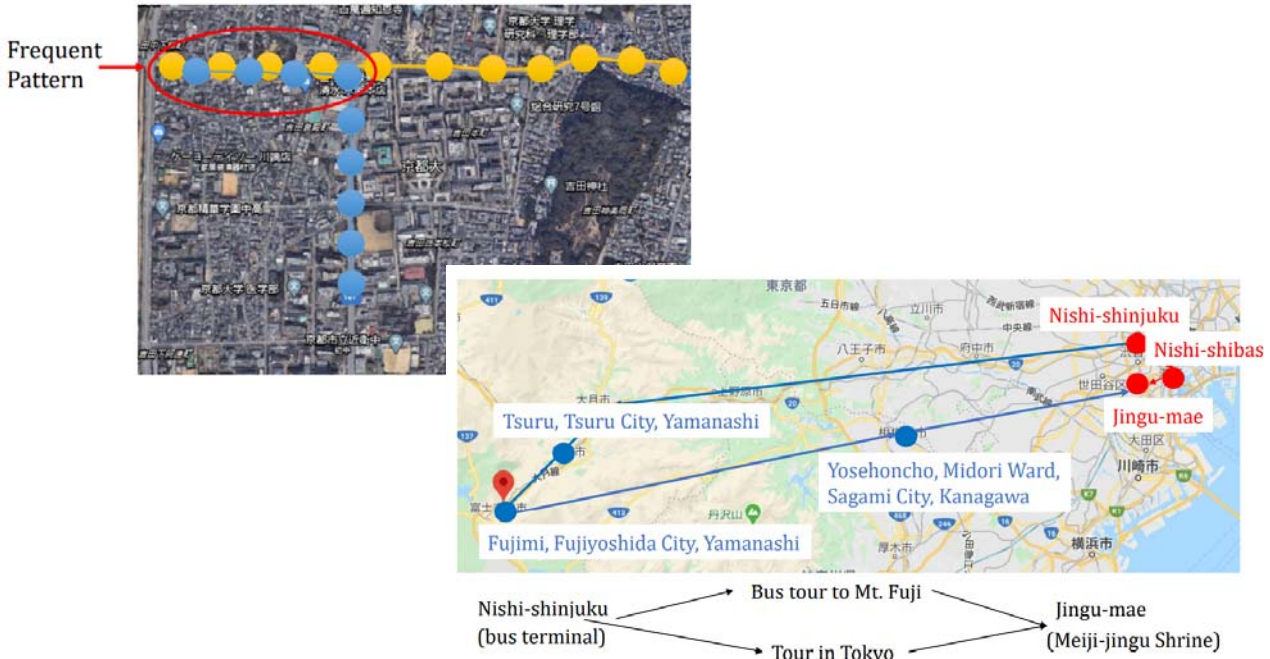
“**機械学習**” そのものを一般化した枠組みを設定した上で
離散構造データ用の機械学習アルゴリズムを開発する。



研究の具体例(1)

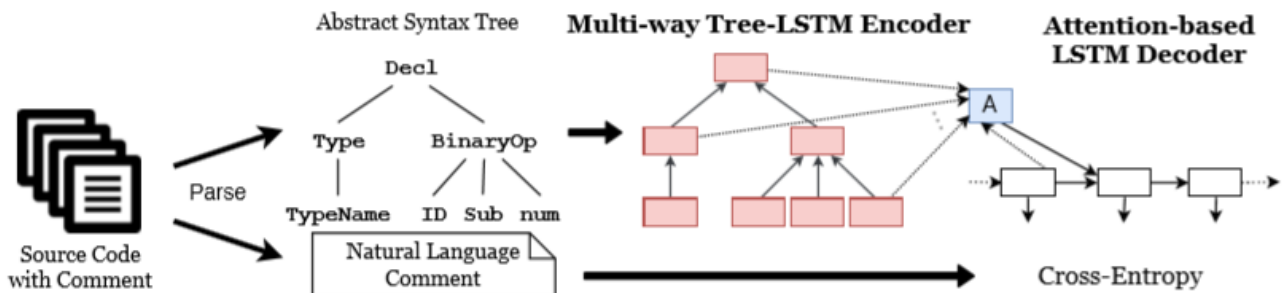
旅行者移動データからの頻出経路発見

外国からの旅行者の移動記録データにたいして、**位置情報を住所表記に変換した上で**、頻出**記号列**発見手法を適用し、旅行者のクラスと経路のクラスの対応関係を閉集合として抽出することに成功



NNを用いたソースコード要約へ応用

ソースコード要約を「ソースコードから自然言語文への翻訳」とみなして自然言語文の翻訳用のNNであるTree-LSTMを理論的に整備した上で、**抽象構文木**に適用し、従来手法よりも適切な要約を生成

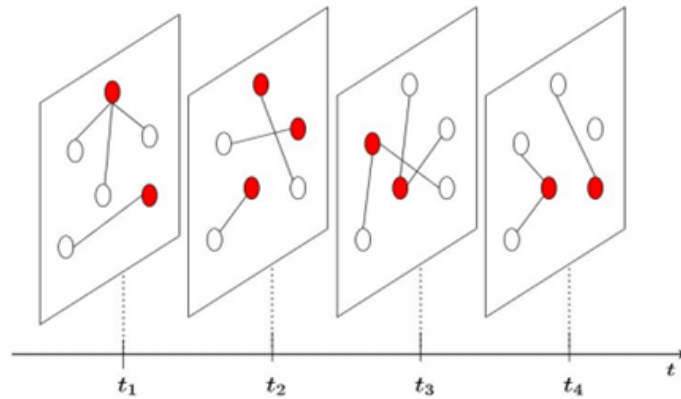


Source code	<pre>public boolean more() throws JSONException { next(); if (end()) { return false; } back(); return true; }</pre>
Gold	Determine if the source string still contains characters that next() can consume
Generated	Determine if the source string still contains characters that next() can consume
Source code	<pre>@JsonIgnore public boolean isDeleted(){ return state.equals(Experiment.State.DELETED); }</pre>
Gold	Signals if this experiment is deleted
Generated	Returns true if this session has been deleted

研究の具体例(2)

多段階最適化問題

- 目的：
- ・ 時間グラフ $G=(V, E, T)$ の各層に対する解を求める
 - ・ 時間経過とともに辺集合のみが変化する T 個のグラフ
 - ・ 隣り合う解の相違の大きさ(対称差)をできるだけ小さくしたい



以下のすべてを卒論1編で解決

- ・ 多段階 d-Hitting Set 問題
- ・ 木幅限定グラフ上の多段階頂点被覆問題
- ・ 木幅限定グラフ上の多段階最適化問題の固定パラメータ容易性
- ・ ロバスト修復可能最適化問題

木構造データの主成分分析

実数値ベクトルデータに対する主成分分析が距離を用いて特徴的な部分空間を求めるように、**木構造データ**に対しても距離を用いて特徴的な部分空間を求める

