

モンテカルロ木探索を用いた 効率的な部分グラフ探索

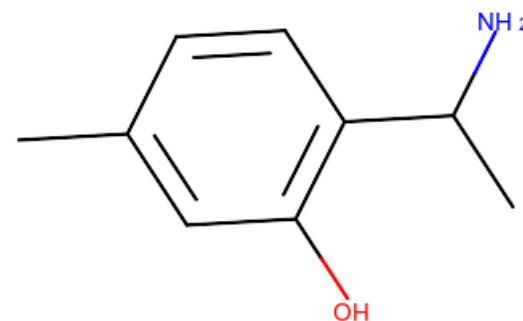
情報認識学研究室 修士2年

白川 稜

背景

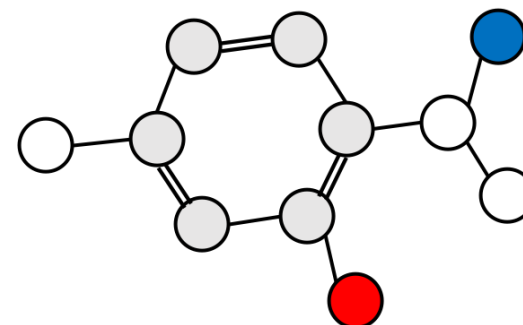
グラフは広く用いられる重要なデータ構造

- 低分子化合物の構造式
- RNA二次構造
- 自然言語処理における構文木




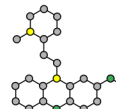
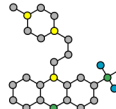
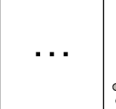
グラフデータからの教師付き学習

- 創薬の分野
- 生命科学や物質化学の分野



グラフ分類回帰問題

input:

y_1	y_2	y_3		y_n
0.1	0.7	1.2	...	0.9
G_1	G_2	G_3		G_n
			...	









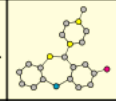
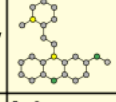
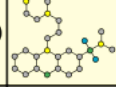
ラベル付きグラフ集合



output:

未知のグラフに対する
ラベル予測器 f

特徴量: 部分グラフ指示子

y	G									...
0.1		1	1	1	1	1	1	1	1	...
0.7		1	1	1	0	1	1	1	1	...
0.9		1	1	1	0	1	1	1	1	...

グラフサイズに対して
部分グラフの総数は組合せ爆発

目的

- 事前制約なしに部分グラフ特徴を利用
- 非線形モデルの構築
- 膨大な特徴空間に対して低コストなモデル構築

提案手法

- 弱仮説：回帰木
- モデル：勾配ブースティング



非線形モデル

- 各分割ごとに部分グラフ特徴探索
有効な枝刈りを設計

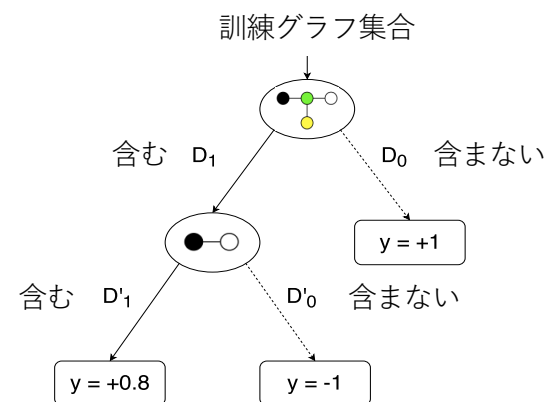


事前制約なし

- モンテカルロ木探索を利用した特徴探索
モンテカルト木探索に適した探索空間の拡張



低コスト

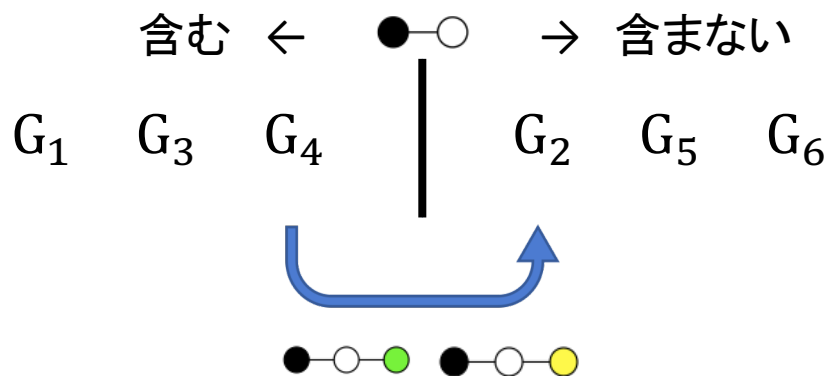


下限値の計算

探索木の特徴：子孫 (x') は親 (x) の拡大グラフ

$$G_i \not\supseteq x \Rightarrow G_i \not\supseteq x', x' \supseteq x$$

含むグラフが含まない側に移る方向性しかない



任意のグラフの組み合わせを含まない側へ移したときの
不純度を全て計算すれば下限値が求まる