

アダマール符号を用いたグラフの類似度計算およびその性能評価

情報科学科 研究室 番号 名前

1 はじめに

グラフマイニングとは頂点と辺で表現できるグラフデータの分析を対象とするデータマイニングの一種である．グラフマイニングによるデータのクラス分類問題はこれまで多くの手法が提案されてきたが，いずれも計算時間と分類精度とのトレードオフの壁に直面した．そこで本研究ではグラフデータのクラス分類問題において既存手法よりも高いスケーラビリティと表現力を持つ手法を考案，構築する．

2 Support Vector Machine とグラフカーネル

Support Vector Machine(SVM) とは正例と負例が混在する事例集合を分類する分類器である．SVM はある特徴空間上にプロットされた事例集合を，超平面によって正例と負例で線形に分離する．事例集合を線形分離可能な空間に写像するために2つの事例 x_i と x_j の内積 $K(x_i, x_j)$ を計算する関数を定義する．この関数はカーネル関数と呼ばれ，とくに x_i と x_j がグラフ構造で表現されているデータの時，カーネル関数はグラフカーネルと呼ばれる．

3 既存手法

グラフカーネルの入力として得られるグラフの特徴付けは非常に重要である．そのためにグラフの頂点をラベル付けし，頂点の比較によってグラフカーネルを定義する．更に各頂点について自身の頂点 v のラベルと隣接頂点集合 $N(v)$ に属する頂点のラベルを引数とし，新しいラベルに更新する関数 $\ell(v)$ を定義する．近傍ハッシュカーネル (Neighborhood Hash Kernel: NHK) はその1つであり，NHK によるラベルを更新する関数 $\ell_1(v)$ は以下で定義される．

$$\ell_1^{(h)}(v) = ROT(\ell_1^{(h-1)}(v)) \oplus \left(\bigoplus_{u \in N(v)} \ell_1^{(h-1)}(u) \right) \quad (1)$$

ここで ROT はビット回転であり，後半 $L-1$ ビットを1ビット左に移動させ，前半の1ビットをその後ろに移動させる．

NHK はハッシュ衝突のリスクを持つ．ここでのハッシュの衝突とは，ある頂点 v_i と v_j において $\ell(v_i)$ と $\ell(v_j)$ が等しくない，かつ $N(v_i)$ と $N(v_j)$ に属する頂点のラベルが等しくないにもかかわらず更新後のラベルが一致することである．これは分類精度の低下の原因の1つである．

4 アダマール符号カーネル

ハッシュの衝突を解決するアプローチとしてアダマール符号カーネル (Hadamard Code Kernel: HCK) を提案する．この手法で使われるウォルシュ・アダマール行列は任意の2行の内積が0の正方行列である．HCKはこのウォルシュ・アダマール行列の各行を頂点ラベルとして保持し，ラベルを更新する関数 $\ell_2^{(h)}(v)$ は以下で定義される．

$$\ell_2^{(h)}(v) = \ell_2^{(h-1)}(v) + \sum_{u \in N(v)} \ell_2^{(h-1)}(u) \quad (2)$$

HCK によるラベルの更新の様子を図1に示す．

HCK も NHK と同様にラベルを固定長のビット列で表現するが，頂点ラベルの出現率が等しいならば，第2要素以降の期待値は0である．よって第2要素以降は小さいビット列 ρ のみ与え，残りを第1要素に与えることで高い表現力が期待できる．

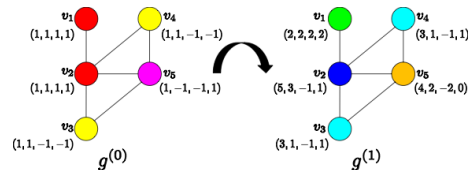


図1 HCK によるラベルの更新操作

5 評価実験

化合物のデータセット MUTAG を用いて NHK と HCK で化合物のクラス (変異原性の有無) を予測した．表??より HCK は NHK はより高い精度で予測できることを示した．

表1 MUTAG における予測精度 (%)

MUTAG における	平均 (± 標準偏差)	最大値
予測精度 (%)		
NHK	86.04(±2.09)	88.30
HCK($\rho = 1$)	86.22(±1.87)	88.30
HCK($\rho = 2$)	86.62(±1.73)	88.30
HCK($\rho = 3$)	87.47(±2.48)	89.89

6 まとめと今後の課題

HCK はラベルの出現確率が等しい前提でより効果を発揮するものであった．課題としてラベルの出現確率が一律でないグラフにも効果を期待できる手法を考案するべきである．

参考文献

[1] 著者名, 書名, In *Proceeding* とか, pp. 179-188, (2009).