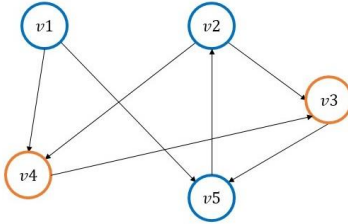


グラフデータに対する特徴ベクトル表現の拡張

グラフデータとは

グラフデータ

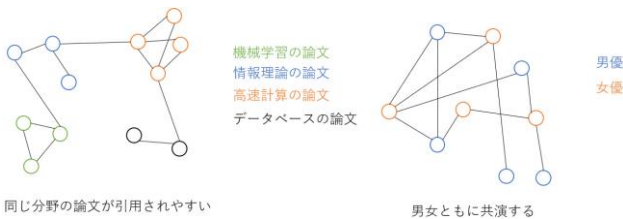
- ・ノードとエッジで構成されたデータ構造
- ・それぞれのノード v は特徴ベクトル X_v を持つ
- ・ノードの中には特定のクラスに属するものがある
- ・ノードのクラスを予測するタスクがクラス分類問題



$X_{v1} : [0, 1, 0, 1, 0, \dots]$
 $X_{v2} : [0, 1, 0, 0, 0, \dots]$
 $X_{v3} : [0, 1, 1, 1, 1, \dots]$
 $X_{v4} : [1, 1, 1, 1, 1, \dots]$
 $X_{v5} : [0, 0, 0, 0, 1, \dots]$

同種グラフ

- ・隣接するノード同士が同じクラスに属する傾向が強いグラフデータ。逆を異種グラフと呼ぶ。



既存手法

情報利得率

- ・重要な情報を抽出する際の指標
- ・情報利得率が大きいほど有益な情報と判断する
- ・特徴 X_v から情報利得率が上位の特徴 n 個を抽出した特徴を $X_v^{\epsilon_n}$ とする

ホップベクトル

- ・近傍の情報を取り込んだ特徴ベクトル
- ・自ノード v の i ホップ先の近傍を表すベクトル表現 $Z_{v,i}$

$$Z_{v,i} = [z_1, z_2, \dots, z_k]$$

$$z_j = \begin{cases} 1 & i\text{ホップ先に頂点}j\text{を持つなら} \\ 0 & \text{それ以外} \end{cases}$$

- ・ホップベクトルのベクトル表現

$$Z_{v,i}^{\epsilon_n} \|X_v\| \lambda Z_{v,2}^{\epsilon_n} \left[\sum_{v' \in N(v)} \text{sgn}(X_{v'}) \right]^{\epsilon_n}$$

1ホップ先の近傍 2ホップ先の近傍 隣接ノードの特徴

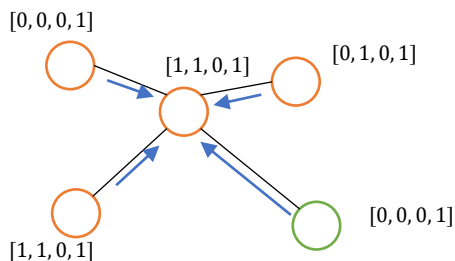
- ・異種グラフにおいては優れているが、同種グラフの分類はGNN系の手法に劣る。

提案手法

- ・ホップベクトルは同種グラフに対して精度があまりよくない
- ・同種グラフは同じクラスに属するノードの間に辺がある傾向が強い



- ・全隣接ノードの特徴の平均を自身の特徴ベクトルに取り込むことで同じクラスのノードが近い特徴を持つようになる



- ・次の式で表した特徴ベクトルを生成する

$$Z_{v,1}^{\epsilon_n} \|X_v + \alpha \text{Ave}(v)\| \lambda Z_{v,2}$$

$$Z_{v,1}^{\epsilon_n} \|X_v + \alpha \text{Ave}(v)\| \lambda \left[Z_{v,2} \left\| \sum_{v' \in N(v)} \text{sgn}(X_{v'}) \right\| \right]$$

$$Z_{v,1}^{\epsilon_n} \|X_v + \alpha \text{Ave}(v)\| \lambda \left[Z_{v,2} \left\| \frac{\sum_{v' \in N(v)} (X_{v'} + \alpha \text{Ave}(v'))}{|N(v)|} \right\| \right]$$

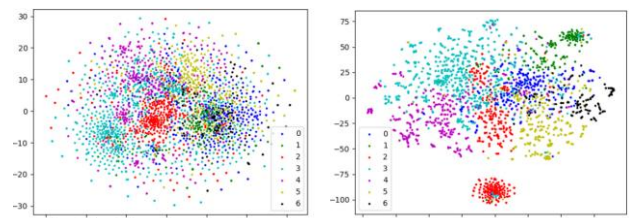
自身のノードに周りに特徴に割引定数をかけたものを取り込む

- ・割引定数を使うことで、同種グラフ、異種グラフの両方に対応できるようにする

実験結果

- ・多層パーセプトロン(MLP)を用いてクラス分類実験

	異種グラフ								同種グラフ		平均
	Texas	Wisconsin	Actor	Squirrel	Chameleon	Cornell	Citeseer	Pubmed	Cora		
HopVecMean1	0.8514	0.8549	0.3516	0.6027	0.7561	0.8459	0.7473	0.8878	0.8641	0.7513	
HopVecMean2	0.8676	0.8431	0.3507	0.6359	0.7518	0.8162	0.7506	0.8772	0.8584	0.7502	
HopVecMean3	0.8270	0.8607	0.3482	0.5923	0.7430	0.8189	0.7501	0.8848	0.8626	0.7437	
HopVec1-MLP	0.8270	0.8471	0.3470	0.6141	0.7232	0.7865	0.7397	0.8701	0.8312	0.7318	
HopVec1-SVM	0.8243	0.8588	0.3173	0.5641	0.7292	0.8297	0.7964	0.8729	0.8056	0.7222	
HopVec2-MLP	0.8541	0.8314	0.3484	0.6227	0.7333	0.7737	0.7411	0.8667	0.8384	0.7359	
HopVec2-SVM	0.8595	0.8294	0.3197	0.5703	0.7221	0.8279	0.7110	0.8725	0.8161	0.7253	
H2GCN-1	0.8486	0.8667	0.3586	0.3642	0.5711	0.8216	0.7707	0.8940	0.8692	0.7072	
H2GCN-2	0.8216	0.8588	0.3562	0.3790	0.5939	0.8216	0.7688	0.8959	0.8781	0.7082	
GCNII	0.6946	0.7412	-	-	0.6061	0.7486	0.7708	0.8957	0.8849	-	
GCNII*	0.7784	0.8157	-	-	0.6248	0.7649	0.7713	0.9030	0.8801	-	
GEOM-GCN*	0.6757	0.6412	0.3163	0.3814	0.6090	0.6981	0.7799	0.9005	0.8527	0.6405	
MLP	0.8189	0.8529	0.3576	0.2968	0.4636	0.8108	0.7241	0.8665	0.7475	0.6599	



元の特徴ベクトルと拡張したホップベクトルでのノードの分布

- ・同種グラフでホップベクトルを上回る正答率を、異種グラフにおいてホップベクトルと同等の正答率を示した
- ・同種グラフにおいてはGNNの手法より正答率が低かった

結論と今後の課題

- ・同種グラフにおける正答率のさらなる向上
- ・ベクトル生成にかかる計算量の削減
- ・最適なパラメータの決定方法