制限付き識別ランダムウォークによるグラフベースのラベル拡張

木村 正成

現実の分類問題において，ラベル付きデータやデータへのアノテーション作業は一般にコストがかかる．一方でラベル無しデータであれば，インターネットからのクローリングやカメラ映像，センサなどから自動でかつ大量に入手することができる．そこで，ラベル無しデータを活用して，ラベル付きデータの数やアノテーション作業のコストを抑えつつ性能の良い分類器を学習することが求められている．こうした問題設定に対応するため，近年では半教師あり学習と能動学習という二種類のアプローチが広く研究されている．

多くの半教師あり学習の目標は，ラベル付きデータとラベル無しデータをうまく組み合わせて分類性能の高いモデルを作ることである．半教師あり学習でよく用いられる手法の一つにラベル伝播があるが，確信度の低いデータにもラベルをつけてしまうため性能が落ちてしまうという問題がある．本研究では，データ集合から生成した有向グラフ上でのランダムウォークにいくつかのルールを課した新しい手法を提案し，深層学習などのより強力な分類器を学習することによって，ラベル付きデータが極端に少ないケースであっても高い性能のモデルを学習することができることを示す．実験では，提案手法をベンチマークデータに適用し，既存の単純なラベル伝播を行ってラベルを増やした手法と比較してそれを上回る結果が得られた．

一方で能動学習は，アノテーションコストを最小化しながら分類器の学習を行うことを目指している．これを達成するため，多くの能動学習の研究は，ラベル無しデータ集合の中から，ラベル情報が得られた際に最も分類器の性能向上に寄与するようなデータ集合を選択し，アノテータに問い合わせることで学習を行なっている．本研究では提案するラベル拡張を，こうした能動学習に適用した新しい手法を提案する．ベンチマークデータに対する実験から，無作為にデータを選択してラベルを付与した場合よりも良い性能の分類器を学習できることを示す．

（指導教員　若林 啓）