氏　名	藤野 昭典

（論文内容の要旨）

本論文は，コンピュータを用いて文書やＷｅｂページなどの各種データを自動的にカテゴリ分けする自動分類の問題に対して，異種データを利用した機械学習によって自動分類の精度を向上させる手法（学習アルゴリズム）を検討したものである．本論文では，異種データを利用する課題として，複数の構成要素から成るデータを分類する課題，複数のカテゴリにデータを分類する課題，異なる種類のデータ（ラベルあり・なしデータ）を用いて学習させる課題，の３つに焦点を当て，これらの課題に対処するための分類器設計法を提案している．また，実データを用いた分類実験によって提案法の有用性を確認している．提案法は，代表的な機械学習法である生成モデルと識別学習の両アプローチを用いることを特長としている．両アプローチの利点を活かしたハイブリッド法の導入によって，異種データを効果的に用いた分類器を設計している．本論文は６章からなる．

１章は序論であり，本研究の背景，対象とする課題，本論文の研究分野への寄与および論文構成を示している．

２章では，複数の構成要素から成るテキストデータの分類問題に対して，構成要素に含まれる情報を効果的に利用して分類精度を向上させる手法を提案している．Ｗｅｂページや技術論文などのテキストデータは，主要な要素である本文と，タイトルや引用といった付加的な要素から構成されている．提案法は，構成要素ごとに設計した生成モデル（構成要素モデル）を識別学習に基づいて重み付き統合することによって分類器を構築することを特長とする．具体的には，データが属するカテゴリを２つ以上の候補の中から１つ選択する多クラス単一分類問題に対して，最大エントロピー原理を用いた構成要素モデルの統合により，データが各カテゴリに属する確率を与える条件付確率モデルを設計している．実データを用いた実験によって，とくに生成モデルと識別学習それぞれのアプローチで同等の分類精度が得られる場合に，提案法が両アプローチよりも高い精度を与えることを確認している．

３章では，多クラス単一分類問題に対して，異なる種類のデータ（ラベルあり・なしデータ）を用いて学習させることで分類精度を向上させる半教師あり学習の一手法を提案している．高精度な分類器を得るには正解のカテゴリが既知のデータ（ラベルありデータ）を大量に用いて学習させる必要があるが，ラベルありデータを収集するには高いコストを要する．一方，正解のカテゴリが未知のデータ（ラベルなしデータ）は容易に取得できる．それ故，ラベルなしデータをラベルありデータと同時に学習に用いて分類器の精度を向上させる半教師あり学習は実用上，重要な課題である．この課題に対して，提案法では，ラベルあり・なしデータでそれぞれ学習させた2種類の生成モデルを識別学習に基づいて重み付き統合することで分類器を構築する．実データを用いた実験によって，従来の生成モデルと識別学習それぞれのアプローチによる半教師あり学習法で同等の精度が得られる場合において，とくに提案法が有用であることを明らかにしている．

４章では，複数の構成要素から成るデータの分類問題に対して，半教師あり学習により分類器を構築する手法を提案している．提案法は，ラベルあり・なしデータ

でそれぞれ学習させた構成要素モデルを統合することで，複数の構成要素と半教師あり学習の課題を同時にハイブリッド法で扱うことを特長とする．実データを用いた実験で，両課題を同時にハイブリッド法で扱うことが，高い分類精度を得るのに有効であることを示している．

5章では，多重分類問題に対して半教師あり学習により分類器を構築する手法を提案している．各データに複数の異なるカテゴリラベルが付与される多重分類問題は，多クラス単一ラベル分類問題と比べてより一般的で複雑なタスクである．提案法では，ラベルありデータで学習させる識別モデルをラベルなしデータで学習させる生成モデルと統合して多重分類器を構築することで,従来の生成・識別の各アプローチに基づく半教師あり学習法を応用した場合と比べてより高い分類精度を得られることを実験的に確認している．

6章は結論であり，本研究で得られた成果を総括し，今後の研究課題を展望している．

氏　名	藤野 昭典

（論文審査の結果の要旨）

本論文は，データを自動的にカテゴリ分けする自動分類の問題に対して，異種データを利用した機械学習によって自動分類の精度を向上させる手法を検討したものであり，導かれた主な成果は以下のようにまとめられる．

1. 複数の構成要素（本文，タイトル，引用など）から成るテキストデータの分類問題に対して，構成要素に含まれる情報を効果的に利用して分類精度を向上させる手法を提案した．提案法では，構成要素ごとに設計した生成モデル（構成要素モデル）を識別学習で重み付き統合することによって分類器を構築する．実データを用いて，生成モデルと識別学習それぞれのアプローチで同等の分類精度が得られる場合に，提案法が有用であることを実験的に明かにした．
2. 低コストで高い自動分類を実現するために，正解のカテゴリが未知のデータ（ラベルなしデータ）を正解のカテゴリが既知のデータ（ラベルありデータ）と同時に学習に用いる半教師あり学習の一手法を提案した．提案法では，ラベルあり・なしデータでそれぞれ学習させた2種類の生成モデルを識別学習で重み付き統合することで分類器を構築する．従来の生成モデルと識別学習それぞれのアプローチによる半教師あり学習法で同等の分類精度が得られる場合にとくに，提案法が両アプローチより高い分類精度を与えることを実験的に明らかにした．
3. 複数の構成要素から成るデータの分類問題に対して，半教師あり学習により分類器を構築する手法を提案した．提案法では，ラベルあり・なしデータでそれぞれ学習させた2種類の構成要素モデルの識別学習に基づく統合により分類器を構築する．複数の構成要素と半教師あり学習の課題を同時に扱う提案法が，高い分類精度を得るのに有効であることを実験的に示した．
4. 各データに複数の異なるカテゴリラベルが付与される多重分類問題に対して半教師あり学習により分類器を構築する手法を提案した．提案法は，ラベルありデータで学習させた識別モデルをラベルなしデータで学習させる生成モデルと統合して多重分類器を与えることを特長とする．従来の生成・識別の各アプローチに基づく半教師あり学習法を応用した場合と比べてより高い分類精度を得られることを実験的に確認した．

以上要するに，異種データを利用する機械学習の主要な課題に対して，本論文は生成モデルと識別学習の両アプローチの利点を取り入れた分類器設計法を提案し，実際の分類問題への適用によって有用性を確認したものである．よって，本論文は博士（情報学）の学位論文として価値あるものと認める．また，平成20年11月25日実施した論文内容とそれに関連した試問の結果合格と認めた．