

統計的モデルと帰納的学習との融合による ユーザ適応型行動予測手法に関する研究

学位論文内容の要旨

近年、情報技術の発達に伴い、人工物と人との共生的な関係形成やコミュニケーションのための基盤技術に関する研究が盛んである。中でも、各ユーザに動的に適応が可能な高い予測精度を持つ行動予測システムを実現することが、実践的な分野への応用に期待されている。このような知的システムの実現において、最も重要なことは、人間の知識獲得メカニズムとして人間の予測能力を有する原理などをどの程度まで解明することができるかという点であると考えられる。

本研究の目的は、それぞれのユーザに動的に適応可能なシステム、いわゆる「個人適応型システム」に関する研究を通じて、そのメカニズム、工学的な実現方法及び原理を解明することである。

従来の行動予測手法は、統計的なアプローチと解析的なアプローチによるものに分けられる。統計的なアプローチは、データを解析する方法によって Point-Based 予測システムと Pass-Based 予測システムに分類される。Point-Based 予測システムは、観測されたデータの頻度情報のような少ない相関情報を用いることにより、予測を行い、その予測精度が約 30～50%となっている。これに対して、時系列データをシーケンスとし、N-gram を利用した Pass-Based 予測システムが提案されている。この手法では、十分な学習データが与えられた場合、その予測精度は約 70%前後となっている。一方、解析的なアプローチにおいては、システムに与えられたルールが適用された場合には、高い確率で予測できるが、人手により矛盾なく予め与える規則には限界があり、各々のユーザに対し、全てのルールを用意するには多くの労力がかかる。また、これらの手法において、予測となるターゲットの頻度情報が不足している時に、ユーザに適応できないという問題が残っている。

一方、人間の行動の特徴は規則性と多様性などが挙げられる。規則的な行動において、N-gram や SVM などの確率モデルを利用することにより、規則性を解明することができるが、多様性のある行動パターンの予測には、ユーザに適応した利用方法を考える必要があることを示している。本研究においては帰納的学習を用いることにより、システムが与えられたデータから自分自身で予測ルールを獲得していく方法を提案している。提案手法では、予めシステムに与えるユーザの行動履歴を少なくすることで、ユーザに依存した表現をルールとして獲得することが可能となる。さらに、N-gram を利用して、帰納的学習により獲得したルールをモデ

ル化することにより、ユーザの行動の多様性と規則性に共に適用可能な Point-Pass-Based 予測手法を提案している。

この提案手法に基づき、人間の生活パターンや嗜好などを自動的に獲得できる部屋を想定した。ユーザの行動をこの部屋を構成しているテレビや電灯など、人間の行動により状態を変化させることが可能な要素の状態がユーザの動作によって変化した時に、その要素の状態に示されているユーザの動作と限定した。実験システムを構築し、人間の実生活に基づき収集したデータを用いて実験を行った結果、89.3%の予測精度が確認された。

また、帰納的学習では、予測結果に対し、ユーザの判定結果を教師情報として、フィードバック学習を用いることにより、システムの予測精度を向上させることができる。そこで、この特徴を利用して、「確率モデルとフィードバック学習の融合によるテレビビデオ自動予約システム」を提案し、N-gram や SVM などの確率モデルとフィードバック学習との融合について試みた。

N-gram とフィードバック学習を融合したテレビビデオ自動予約システムでは、N-gram を用いて、ユーザのテレビ番組の閲覧履歴から、各番組のカテゴリ情報を分類情報として、各カテゴリに従って閲覧履歴を分類し、単名詞や未知語を解析の対象とし、複合名詞を含めた重み付けのキーワードベクトル空間を抽出する。構築されたキーワードベクトル空間はユーザの興味空間を示しているため、これらの情報を用いて各新番組に対し、ベクトル内積法により、新番組にマッチする全要素をクエリーとして獲得できる。さらに、このクエリーを用いて、tf-idf 法を利用することにより、その番組の tf-idf 重みが求められる。そして、N-gram により求めた重みと tf-idf 重みをそれぞれ正規化して統合することにより、ユーザの嗜好に合う新番組を推薦可能になる。また、推薦結果に対してユーザは判定し、判定結果をフィードバック情報として学習させ、ユーザの興味空間を更新することにより、予測の精度を向上させる。本システムのフィードバック学習では、システムの即応性と頑健性を持たせるために、ユーザの嗜好パターンを示すフィードバックタスクの自動抽出とユーザの閲覧履歴を適切に分割した上で、興味空間を構築する方法を考慮した。実験の結果、86.9%の予測精度が確認された。

また、SVM を用いたフィードバック学習によるテレビビデオ自動予約システムでは、SVM を利用して、ユーザの閲覧履歴から、ユーザの嗜好を示す行動パターンをモデル化し、フィードバック学習を行うことで、モデルを更新させ、ユーザの嗜好に適応した番組を推薦することが可能になる。具体的には、インターネットから抽出されたテレビ番組から、時間(曜日、開始及び終了時刻)、チャンネル名、カテゴリ(日、英)、番組の説明部分に含まれている名詞と未知語などを素性として抽出し、ユーザが見た番組をクラス1、見ない番組をクラス-1、新番組をクラス0と分類して、線形 SVM や多項式 SVM を用いる方法である。各素性の重み付けはブル形式または tf-idf 値を利用する。実験の結果、線形 SVM では約 82%の予測精度で、多項式 SVM では約 92%の予測精度が得られた。

以上の結果から、統計的モデルと帰納的学習の融合によるユーザ適応型行動予測手法に関しては、帰納的学習によりルールを獲得する方法とN-gramとの融合、N-gramによるユーザの興味空間を構築する方法とフィードバック学習との融合、そして、SVM を用いてユーザの行動パターンをモデル化する手法とフィードバック学習との融合の三つの面からシステムを構築し、実験結果により、ユーザ適応型行動予測手法としての有効性が確認された。

学位論文審査の要旨

主 査 教 授 荒 木 健 治

副 査 教 授 北 島 秀 夫

副 査 教 授 山 本 強

学 位 論 文 題 名

統計的モデルと帰納的学習との融合による ユーザ適応型行動予測手法に関する研究

著者は、それぞれのユーザに動的に適応が可能なシステム、いわゆる「個人適応型システム」の実現に関する問題を解決するために、統計的モデルと帰納的学習との融合によるユーザ適応型行動予測手法の提案を行い、予測性能の向上に関する研究を行った。また、著者は、近年の情報技術の向上に伴う個人ユーザ適応システムの需要を考慮し、人間の知識獲得メカニズムそして人間の予測能力を有する原理などに着目し、行動予測における一般的な学習の枠組みを作り上げる方法に関する検討、及び人間のように学習能力を有するシステムの工学的な実現方法や原理の探索を研究の目的としている。

従来の行動予測手法は、統計的なアプローチと解析的なアプローチによるものに分けられる。統計的なアプローチは、データを解析する方法によってPoint-Based 予測システムと Pass-Based 予測システムに分類される。これらの手法においては、予測精度が低い、動的適応が不可能、そして、データの収集やルールの用意などに多くの労力がかかるという問題が残されている。

一方、人間の行動の特徴としては規則性と多様性などが挙げられる。規則的な行動においては、N-gram や SVM などの統計的モデルを利用することにより、規則性を解明することができるが、多様性のある行動パターンの予測には、ユーザに適応した利用方法を考える必要がある。著者は、これらのユーザの行動の特徴を考慮した上で、「N-gram を用いた帰納的学習によるユーザの行動予測手法」を提案している。提案手法では帰納的学習を用いることにより、システムが与えられたデータから自分自身で予測ルールを獲得し、ユーザに依存したルールを獲得することが可能となる。さらに、N-gram を利用して、ユーザの行動予測に適用されたルールをモデル化することにより、ユーザの行動の多様性と規則性に共に適用可能な Point-Pass-Based 予測手法を提案している。

この提案手法に基づき、人間の生活パターンや嗜好などを自動的に獲得できる部屋を想定した。ユーザの行動をこの部屋を構成しているテレビや電灯など、人間の行動により状態を変化させることが可能な要素の状態がユーザの動作によって変化した時に、その要素の状態に示されているユーザの動作に限定した。実験システムを構築し、人間の実生活に基づき収集したデータを用いて実験を行った結果、89.3%の予測精度が確認された。

また、著者は、帰納的学習は、フィードバック学習を含んでいるという特徴を利用して、「確率モデルとフィードバック学習の融合によるテレビビデオ自動予約システム」を提案し、N-gram や SVM などの確率モデルとフィードバック学習との融合について試みた。N-gram とフィードバック学習を融合したテレビビデオ自動予約システムでは、N-gram を用いて、ユーザのテレビ番組の視聴履歴から、各番組のカテゴリ情報を分類情報として、各カテゴリに従って視聴履歴を分類し、単名詞や未知語を解析の対象とし、複合名詞を含めた重み付けのキーワードベクトル空間を抽出する。構築されたキーワードベクトル空間はユーザの興味空間を示しているため、これらの情報を用いて各新番組に対し、ベクトル内積法により、新番組にマッチする全要素をクエリーとして獲得できる。さらに、このクエリーを用いて、tf・idf 法を利用することにより、その番組の tf・idf 重みが求められる。そして、N-gram により求めた重みと tf・idf 重みをそれぞれ正規化して統合することにより、ユーザの嗜好に合う新番組を推薦可能となる。また、推薦結果に対してユーザが判定し、判定結果をフィードバック情報として学習させ、ユーザの興味空間を更新することにより、予測の精度を向上させる。本システムのフィードバック学習では、システムの即応性と頑健性を持たせるために、ユーザの嗜好パターンを示すフィードバックタスクの自動抽出とユーザの視聴履歴を適切に分割した上で、興味空間を構築する方法を考察した。実験の結果、86.9%の予測精度が確認された。

また、SVM を用いたフィードバック学習によるテレビビデオ自動予約システムでは、SVM を利用して、ユーザの視聴履歴から、ユーザの嗜好を示す行動パターンをモデル化し、フィードバック学習を行うことで、モデルを更新させ、ユーザの嗜好に適合した番組を推薦することが可能になる。具体的には、インターネットから抽出されたテレビ番組から、時間（曜日、開始及び終了時刻）、チャンネル名、カテゴリ（日、英）、番組の説明部分に含まれている名詞と未知語などを素性として抽出し、ユーザが視聴した番組をクラス 1、視聴しなかった番組をクラス -1、新番組をクラス 0 と分類して、線形 SVM や多項式 SVM を用いる方法である。各素性の重み付けはブール形式または tf・idf 値を利用する。実験の結果、線形 SVM では約 82 % の予測精度、多項式 SVM では約 92 % の予測精度が得られた。

以上を要約すると、著者は統計的モデルと帰納的学習の融合によるユーザ適応型行動予測手法に関しては、帰納的学習によりルールを獲得する方法と N-gram との融合、N-gram によるユーザの興味空間を構築する方法とフィードバック学習との融合、SVM を用いてユーザの行動パターンをモデル化する手法とフィードバック学習との融合の三つの面から実験システムを構築し、性能評価実験を行った。実験結果よりユーザ適応型行動予測手法としての有効性が確認された。本研究を通じて、情報化社会におけるユーザ適応型行動予測手法の確立に貢献するところ大なるものがある。よって、著者は北海道大学博士（工学）の学位を授与される資格あるものと認める。