氏　　名	Tareeq Saifuddin Md.
学位（専攻分野）	博士（情報学）
学 位 記 番 号	総研大甲第 1427 号
学位授与の日付	平成２３年３月２４日
学位授与の要件	複合科学研究科　情報学専攻 学位規則第６条第１項該当
学 位 論 文 題 目	Rapid Behavior Adaptation for Human Centered Robots Through Demonstration
論 文 審 査 委 員	主　　査　准教授　稲邑　哲也 教授　山田　誠二 教授　佐藤　健 准教授　市瀬　龍太郎 名誉教授　上野　晴樹

論文内容の要旨

Robots have proven powerful tools in the predictable environments of factories and manufacturing plants. However, they have been far less successful in human environments characterized by a higher degree of uncertainty and change. Each response of today's industrial robots has to be programmed in advance. This approach is ill suited for robots in human environments, which require a vast amount of knowledge and the specification of a wide set of behaviors for successful performance. Typically robots in human environments are placed in very restricted worlds because then the environment can be controlled. If a robot is taken in a unknown home, that approach just doesn't hold anymore. Moreover when the user or environment changes frequently the robotic system should be able to adapt to new user or environment rapidly to take correct action. This has introduced the need for building robotic systems able to adapting to user and environment in an engaging way by using their observed sensory information.

The recent trend in robotics is to develop a new generation of robots that are capable of adapting to new user, interacting with user and participate in our daily lives. Adaptive behavior plays an important role in the assistance of different user with different needs. Therefore, such robots should be able to rapidly adapt to user preference, user policy and have interaction skills to communicate with user. In this thesis user's preference indicates variation of behavior decision by the user even though identical sensor is observed. And user's policy is defined by the mapping from observation to action. The problem of learning a policy, a task representation mapping from world states to actions, lies at the heart of many robotic applications. One approach to acquiring a task policy is learning from demonstration, an interactive learning approach based on human-robot interaction that provides an intuitive interface for robot programming. In this approach, a teacher performs demonstrations of the desired behavior to the robot. The robot records the demonstrations, typically as state to action mappings, and learns a policy imitating the teacher's behavior.

Learning from demonstration is an incremental online learning process in which the robot begins with no knowledge about the task, and acquires training data until a fully autonomous policy representing the complete task is learned. If the user changes his preference or policy the system should adapt to the new preference or policy rapidly. This thesis contributes an interactive approach to demonstration learning that enables the robot to rapidly adapt to user preference or policy. These algorithms enable the robot to identify the need for and request demonstrations for specific parts of the state space based on confidence thresholds characterizing the uncertainty of the learned policy. In our evaluation, we show that this approach significantly reduces the number of demonstrations and can rapidly follow user preference or policy.

Demonstrations provide the robot with a dataset consisting of state-action pairs representing examples of the desired behavior. The robot's goal is to use this information to adapt to a policy, which enables the robot to select an action based upon its current world state. Our policy should map from the robot's state to a discrete set of action primitives. And due to the interactive nature

of learning from demonstration, policy adaptation must occur in real time.

The state-action mapping represented by a policy is typically complex. One reason for this complexity is that the desired observation-action mapping is unknown. A second reason for this complexity is the complications of policy adaptation in real world environments. Traditional approaches to robot control model the domain dynamics and derive policies using mathematical models. Though theoretically well-founded, these approaches depend heavily upon the accuracy of the model. Not only does this model require considerable expertise to develop, but approximations such as linearization are often introduced for computational tractability, thereby degrading performance. Other approaches, such as (reinforcement learning), guide policy learning by providing reward feedback about the desirability of visiting particular states. To define a function to provide these rewards, however, is known to be a difficult problem that also requires considerable expertise to address. Furthermore, building the policy requires gathering information by visiting states to receive rewards, which is non-trivial for a mobile robot learner executing actual actions in the real world. We chose Bayesian network for rapid policy adaptation because it can represent degree of confidence for behavior decision as probability and can provide a confidence even with a small number of observations. Also Bayesian network is suitable for online interactive learning.

This thesis presents a Bayesian network based framework to address rapid behavior adaptation. The performance of Bayesian learning strongly depends on the quality of the demonstration dataset. When the dataset included significant data, the learning would be a success. But it is difficult to evaluate data to be insignificant because when the data become insignificant for learning process is not known a priori. We propose a method for evaluating significance of data based on a concept of change in the degree of confidence. A small change in the degree of confidence can be regarded as an insignificant data for learning, so that data will be evaluated as insignificant.

For evaluating the significance of demonstration, the experience data is assigned to distribution parameters. The distribution represents not only event probability among behaviors, but also degree of confidence for the output probability. The system calculates the degree of confidence by integrating the area around peak of the distribution after each demonstration. The change in the two consecutive degrees of confidence can be regarded as the importance of the observation to the learning process. When the change in the degree of confidence in two consecutive time steps is small, this situation is regarded as familiar; the experience data is considered insignificant for learning and discarded. In contrast, when the robot detect a large change in the degree of confidence in two consecutive time steps, this situation is considered unfamiliar; the experience data is considered significant for learning and be accepted.

With this significance evaluation method we introduce multiple rapid behavior adaptation algorithms that enable the robot to evaluate demonstrations based on the change in the degree of confidence. The rapid adaptation algorithm enables the robot to evaluate demonstrations in real time as it interacts with the user.

博士論文の審査結果の要旨

本論文では，ユーザが操縦をするタイプのロボットにおいて，操縦の履歴を観察することで段階的にその自律行動の枠組みを学習する機能に着目し，環境の状況やユーザの操縦の傾向等が動的に変化したとしても，その変動に迅速に対応し，再学習に必要となる時間を短縮する方法が提案された．ユーザの実演に基づいて行動を学習する枠組みは Learning by Demonstration と呼ばれ，ロボットにおける機械学習の応用例として近年着目されているが，学習に必要となるサンプル数が膨大になることが一つの問題点となっており，上記のような迅速な状況の変動に対する再学習を十分に実用的な速さで行う手法についてはあまり議論がされてこなかった．本論文で，Saifuddin Md. Tareeq 氏は，そのような迅速な再学習を実現するためのコンセプトを統計的学習理論をベースに考案し，実際の移動ロボットにおいてその有効性を確認している．まず，少ない数の学習サンプルであってもある程度の確信度で行動決定するための枠組みとしてベイジアンネットワークを採用した．適切なネットワーク構造の中の条件付き確率値を求めるには，取り得る状況の各状態を経験した頻度を用いるが，平凡な状態の経験数とまれで重要な状態の経験数に違いを持たせることで迅速な再学習を行うアプローチを提案した．そのアプローチを実現するために，行動決定の確信度の時間変動に着目し，確信度変動が大きい場合にまれで重要な状態であると判断することで迅速な再学習が可能であることを示し，仮想環境における移動ロボットや実際に廊下を走行する移動ロボットにおける実験を通じてその有効性を示した．

本論文は５章からなる．第１章では，まず．ユーザ中心型の知能ロボットにおける現在の研究動向と今後予想される問題点が述べられ，移動ロボットやユーザを取り巻く状況の変動に伴う迅速な再学習の必要性について議論された．

第２章では，知能ロボットの学問分野における Learning by Demonstration の位置づけとその問題点について議論がなされ，前述した状況の変動に伴う再学習の困難性について述べられた．また，関連研究との比較を行い，ベイジアンネットワークに基づく行動学習と自律動作の選択に枠組みについての利点と問題点が議論された．ベイジアンネットワークの学習を用いることで，少ない数の学習サンプル数でもある程度の精度の学習が可能であることを示し，行動決定の確信度の時間変動を用いた迅速な再学習のアプローチについて提案がなされた．

第３章では，提案する迅速な再学習方法の評価の第一段階として，ユーザの操縦における傾向（Preference）を学習する状況での実験と評価がなされた．具体的には，ユーザが移動ロボットを操縦中に障害物に遭遇した場合，右に避けるか左に避けるかという判断の傾向を対象としたものである．観測された各行動の経験回数を単純にカウントして頻度を求めると迅速な再学習に支障をきたす場合があるので，Dirichlet 分布を用いて経験された新しい学習サンプルが，過去の履歴と比較して重要であるか否かを判断し，経験回数のカウントに含めるか否かを判定する手法が提案された．実際の実験を通じて，提案手法を導入することで，迅速性が向上することを示し，その有効性について議論がなされた．

第４章では，提案する迅速な再学習方法の評価の第二段階として，ユーザの操縦における方策(Policy)を学習する状況での実験と評価がなされた．第３章で議論された Preference の学習と異なり，Policy の学習はユーザが操縦した行動とその時に観測されたセンサ値と

の関係性を学習する必要があるため，複数のセンサノードを持つベイジアンネットワーク構造に対して提案手法を実装する拡張アルゴリズムが示された．Policy の学習の具体例として，動的に移動する対象物を追従する／回避するという二つのユーザの方策（Policy）が予測不可能なタイミングで変更されるという状況を題材とし，実験を通じて提案手法の有効性が示された．

第５章では，本研究の成果をまとめ，今後の研究の課題点や将来展望について述べた．提案手法の本論文における役割は状況変動に伴う迅速な再学習であったが，ユーザへの質問・提案・確認などを通じた対話的な操縦補助を行うためのポテンシャルも持ち合わせており，その実装方法について述べられた．

ユーザの実演をベースに学習を行う知能ロボットでは，その学習の精密さや異なる状況への適応可能性に重点が置かれる傾向が強く，膨大な数の実演を行いデータベースを構築した後にオフラインで学習することが多かった．これに対して，リアルタイムに学習を行い，刻一刻と変動する状況に追従するような再学習を実現した点は学術的に新規性があり，重要な成果であると言える．また，予備審査の段階では，1)迅速な再学習の定量的評価法とそれに基づく実験の評価，2)取り扱う問題の知能ロボット・機械学習分野における位置づけの明確化，3)Preference と Policy の詳細な定義の追加，4)Dirichlet 分布による実装方法の詳細な議論の追加，の４点が本審査時点で満たされるべきであるとの指摘があったが，本審査では上記の４点ともに十分に満たされていることが確認された．

以上の点を総合的に判断して，本論文は学位を授与するに値すると判断した．