別記様式第３号－１

学 位 論 文 の 要 旨

氏　　名	ケティ　ティオウン KATHY THI AUNG
学位論文題目	連続的な状態空間のボロノイ分割を用いた強化学習に関する研究

本論文は，ボロノイ分割の概念を用いて状態空間を分割するVQE(Voronoi Q-value Element) を提案し，その有効性を検証した．さらに，このVQEの追加や統合手法を提案し，状態空間を格子状に分割するQ-Tableとの性能比較を行い，それぞれの検証を行った結果をまとめたものである．

第１章は，本研究の背景と目的，関連研究に対する位置付け，及び本論文の構成を述べた．

第２章は，まず2.1節で，本論文で使用する記号の定義を行った．2.2節では，人間が自然に行っている学習能力と同様の機能をコンピュータで実現させるための技術．手法である又，未知の環境内においてエージェントが将来の報酬の期待値を最大にするためにどのような行動をとればよいかという問題を扱う機械学習の一種である強化学習について述べた．2.3節では，強化学習の問題について述べ，2.4節では，この強化学習の代表的な手法であるQ学習という各状態において取れる行動ごとにその行動価値を示すQ値という値を持たせ，環境から得られた報酬をもとにこのQ値を更新することで学習を行う学習式について述べた．2.5節では，エージェントの行動選択方法について説明し，2.6節では，状態空間を離散化する又は格子状に分割するQ-Tableについて述べた．2.7節では，高次元の環境において状態数が指数的に増加する次元の呪いについて述べ，2.8節ではその問題を解決ためQブロックを使った場合，実行時間はQ-Tableより約2倍かかることを示した．

第３章は，以上の問題を解決するために，ボロノイ分割の概念を用いて状態空間を分割するVQEを提案した．Q-Tableはあらかじめ全ての状態に対してQ値を準備しておく必要があるが，VQEは必要に応じて状態空間に追加できる．3.1節では，空間上に設置された母点に対して任意の点がどの母点に一番近いかによって空間を分割する方法であるボロノイ分割について述べた．3.2節では，VQEの作成方法，参照方法，またVQEを用いたQ学習の利点を実証した．

第４章は，前節で述べたQ-TableとVQEの性能比較実験として連続状態空間の実験モデルにおいてシミュレーションを行った．また，それらについて考察した．

第５章では，行動空間と状態空間が一致していない場合と一致している場合，エージェントの行動向きとVQEの回転角度を0度から90度まで5度刻みで回転させて実験を行った．Bug　Pos実験モデルを用いて検証した結果，VQEの角度と行動の角度が45度ずれている場合が最も性能が良く，0度の場合は最も悪いことが分かったことを考察した．

第６章は，VQEの配置によってQ学習の性能が本当に変化していることから，ボロノイ領域を実現するためにVQEの位置を決める方法としてVQEの追加方法を提案した．また，Bait　View　World実験モデルにおいてシミュレーションを行い，性能を検討した．

第７章では，VQEを統合することで，学習性能を保ったまま，状態数を少なくでき，さらにメモリ使用量や学習時間を抑えることができるのではないかというVQEの統合手法の検証を行った．

第８章では，全体のまとめと今後の課題について述べた．

別記様式第13号

論 文 審 査 の 要 旨

報告番号	理工研 第 378 号		氏 名	Kathy Thi Aung
審査委員	主 査	渕田 孝康		
	副 査	中山 茂	森 邦彦	

学位論文題目 Study on reinforcement learning using Voronoi diagram in continuous state space
(連続的な状態空間のボロノイ分割を用いた強化学習に関する研究)

審査要旨

提出された学位論文及び論文目録等を基に学位論文審査を実施した。本論文は，ボロノイ分割の概念を用いて状態空間を分割するVQE(Voronoi Q-value Element) を提案し，その有効性を検証した．さらに，このVQEの追加や統合手法を提案し，状態空間を格子状に分割するQ-Tableとの性能比較を行い，それぞれの検証を行った結果をまとめたものである．

第１章は，本研究の背景と目的，関連研究に対する位置付け，及び本論文の構成を述べた．

第２章は，まず2.1節で，本論文で使用する記号の定義を行った．2.2節および2.3節で，機械学習の一種である強化学習、およびその問題について述べ，2.4節では，この強化学習の代表的な手法であるQ学習の学習式について述べた．2.5節では，エージェントの行動選択方法について説明し，2.6節では，状態空間を離散化する又は格子状に分割するQ-Tableについて述べた．2.7節では，高次元の環境において状態数が指数的に増加する次元の呪いについて述べ，2.8節ではその問題を解決するためにQブロックを使った場合，実行時間はQ-Tableより約2倍かかることを示した．

第３章は，以上の問題を解決するために，ボロノイ分割の概念を用いて状態空間を分割するVQEを提案した．Q-Tableはあらかじめ全ての状態に対してQ値を準備しておく必要があるが，VQEは必要に応じて状態空間に追加できる．3.1節では，空間上に設置された母点に対して任意の点がどの母点に一番近いかによって空間を分割する方法であるボロノイ分割について述べた．3.2節では，VQEの作成方法，参照方法，またVQEを用いたQ学習の利点を実証した．

第４章は，前節で述べたQ-TableとVQEの性能比較実験として連続状態空間の実験モデルにおいてシミュレーションを行った．また，それらについて考察した．

第５章では，行動空間と状態空間が一致していない場合と一致している場合，エージェントの行動向きとVQEの回転角度を0度から90度まで5度刻みで回転させて実験を行った．BugPos実験モデルを用いて検証した結果，VQEの角度と行動の角度が45度ずれている場合が最も性能が良く，0度の場合は最も悪いことが分かったことを考察した．

第６章は，VQEの配置によってQ学習の性能が本当に変化していることから，ボロノイ領域を実現するためにVQEの位置を決める方法としてVQEの追加方法を提案した．また，BaitViewWorldの実験モデルにおいてシミュレーションを行い，性能を検討した．

第７章では，VQEを統合することで，学習性能を保ったまま，状態数を少なくでき，さらにメモリ使用量や学習時間を抑えることができるのではないかというVQEの統合手法の検証を行った．

第８章では，全体のまとめと今後の課題について述べた．

以上、本論文は、連続状態空間におけるQ学習に関する研究で、ボロノイ分割を利用したQ値の追加と統合について検討を行い，報酬獲得数が向上可能であることを明らかにした。これは、知能ロボット等における学習理論に大きく寄与する。

よって，審査委員会は博士（工学）の学位論文として合格と判定する。

別記様式第14号

最終試験結果の要旨

報告番号	理工研　第 378 号	氏　名	Kathy Thi Aung
審査委員	主　査	渕田 孝康	
	副　査	中山 茂	森 邦彦

主な質疑応答は、以下の通りであった。

質問1: エピソードとは、エージェントが生まれてからreward得るまでが1 episodeか？ターンとは何か？

回答1: いいえ、連続行動を100万回行動を1エピソードとして実験している。1ターンは、強化学習の状態観測から学習までの一通りの流れを言う。

質問2: 最後の実験結果でintegration methodの結果が急に良くなって、そしてepisodeが増えると下がるのはなぜか？Number of rewards というのはreward を得た回数のことか？

回答2:縦軸は報酬を得た回数でもありVQEの数も表している。この研究では、同じ数のVQEで性能を良くするというのを目的しているため、グラフの最後のほぼ同数のVQEになった時点での結果を見る必要がある。

質問3: Integration手法で後半の報酬獲得数が減少しているのはintegrationをやり過ぎていることが原因か？ Integrationをすることの目的はVQEの数または状態数を下がるためか？

回答3: はい。Integrationの目的は状態数を減らすためである。Over integration problemは残っているが、Integration法により、最終的には同数のVQEで報酬数をわずかではあるが向上させることができている。

質問4: Model Iでreward areaが回っていたが、そのときに学習して、別の動きしたら報酬獲得数はどうなるのか？

回答4: 報酬エリアが円に動く状態で学習し、他の動きの実験は行っていない。しかし、基本的には、餌場の距離を縮めて角度を0に持って行くような学習をしているため、餌場の動きが変わったとしても報酬を獲得できると考えられる。若干の性能低下がある可能性はある。

質問5: 学習時のランダム行動確率を30%としているが、その比率は実験の最初と最後のではずっと変化しないのか？ ランダム行動率を減少させるような実験は行っていないのか？

回答5: はい。その比率は一定で実験している。ランダム行動率を徐々に減少させるという方法は良くやられている方法ではあるが、今回は、いつ学習が終わるかはっきり分からないのでこれを使っていない。

質問6: 参考文献中で、連続空間でQ-Tableを使った別な方法は提案されているか？ その方法と提案手法の実験の比較はあるのか？

回答6: ボロノイ分割を用いた連続空間での強化学習の研究は存在している。しかしそれは、常に評価関数を得られるような環境で行っている研究あり、通常の強化学習の条件とは異なっているモデルを採用している。このため、提案手法との実験条件を合わせられないため比較できなかった。

質問9: VQEを使うことの利点は何か？ なぜそれを使うのか？

回答9: VQEを使うことにより、状態空間の無駄な分割を減らすことができる。また、重複領域がないため、Q値を有効に利用できる。さらに、Q値の位置の自由度が高く、統合しやすいという利点もある。

質問10: 実験Model IでVQEを回転したら、報酬獲得数が45度のとき落ちているのはなぜか？

回答10: 状態空間の分割が、回転することで行動方向に対して斜めになると、無駄な行動が起きる。例えば、状態が変わるとagentが違う行動を取っても同じ状態に行くことがある。Q学習は同じ状態で行動するとQ値が下がる。

など、約20の質問が出され、いずれも適切に回答した。

以上の結果から、審査委員会は申請者が博士（工学）の学位を与えるに十分な学力と見識を有するものと認定した。