

契約ネットプロトコルにおける入札ルールの学習

池田 圭吾[†] 小川 祐樹^{††} 末田 直道^{††} 新田 恒美^{†††}

[†]大分大学大学院工学研究科知能情報システム工学専攻

^{††}大分大学工学部知能情報システム工学科

^{†††}デジタルアンネット(株)

1 はじめに

近年、インターネット環境の普及に伴い、エージェントと呼ばれる、人間の代理としてネットワーク上で自律的・協調的に仕事をこなすソフトウェアが注目を浴びている。複数のエージェントの行動により問題を解決するようなシステムはマルチエージェントシステムと呼ばれており、マルチエージェントシステムにおいて、エージェントは個々の問題と、システム全体の問題を解決することを目的とするが、システム全体の問題を解決するために合意形成が必要である。マルチエージェント合意形成問題は多くの研究がなされており、マルチエージェントシステムにおける重要な問題の一つとなっている。

本研究では、合意形成問題解決手法の一つである契約ネットプロトコルをヘルパースケジューリング問題に適用させ、学習の要素を組み込むことで、効率的なタスク割り当てを行うことを試みる。また、学習には強化学習を用いる。

2 問題設定

本研究では問題設定として、居宅介護支援におけるヘルパースケジューリング問題を考える。現在、高齢化に伴い、自宅で生活するための介護サービスの需要が高まってきている。介護サービスのスケジュール作成は、ケアマネージャと呼ばれる介護支援専門員が行っている。しかし、被介護者の希望日時やヘルパーの労働条件、また被介護者が必要とする介護支援レベルに適したヘルパーの派遣といった様々な条件を考慮しなければならないため、被介護者、ヘルパー共に満足するスケジュールを作成することが困難になっている。そこで、この問題を、ケアマ

ネージャ、ヘルパーをエージェントとするマルチエージェントシステムとして捉え、それぞれのエージェントが学習を行うことで、被介護者が十分に満足でき、かつヘルパーの希望を考慮できるようなスケジュールの自動立案を行うことを目的とする。

3 利用技術

3.1 契約ネットプロトコル

契約ネットプロトコルは、複雑なタスクを独立した個々のタスクに分割して、それらを複数のエージェントに割り当てるためのプロトコルである。分割されたタスクは独立に解かれるものとして契約が結ばれる。人間社会における契約入札の仕組みを基にしているため、契約ネットプロトコルと呼ばれており、タスクのアナウンス、入札、落札といった手順を用いてタスクの割り当てを行う。契約ネットプロトコルを用いることで、両者の意思を反映した相互選択の解を得ることができる。

3.2 強化学習

強化学習とは、環境との相互作用を通じて学習する制御の枠組みである。教師あり学習とは異なり、状態入力に対する正しい行動出力を明示的に示す教師が存在しない。かわりに報酬というスカラーの情報を手がかりに学習を行う。与えられた目標を達成すれば正の報酬、失敗した場合は負の報酬を与えることで、エージェントはより多くの報酬を得られるように自らの行動パターンを変化させ、最適な行動を学習することが可能となる。

4 実験

4.1 システム概要

最初に、ケアマネージャと被介護者との間でアセスメントが行われ、ケアマネージャによりタスク（介護計画）が作成される。これを全て

Learning of Tender Rule in Contract Net Protocol
Keigo Ikeda[†] Yuki Ogawa^{††} Naomichi Sueda^{††}
Tsunemi Nitta^{†††}

[†]Graduate School of Engineering, Oita University

^{††}Department of Computer Science and Intelligent Systems, Oita University

^{†††}DigitalianNet & Co.,Ltd

の被介護者に対して行い、作成された全てのタスクを日付ごとに分類し、各ヘルパーに送信する。ヘルパーは、受信したタスクに対し入札を行う。ケアマネージャは各ヘルパーから受け取った入札から、タスクごとに適したヘルパーを決定し、落札通知を送信し、タスクが処理される。また、ヘルパーは時間帯が重なるタスクには、一つしか入札できないものとする。

ここで、ケアマネージャ、ヘルパー共に学習を行う。ケアマネージャは、各々のタスク（被介護者、日時、サービス内容）に対し割り当てたヘルパーが良かったかどうかを、予め用意された評価モデルによって評価される。評価により報酬を得ることで、タスクごとに適したヘルパーの割り当てルールを学習する。各ヘルパーは、入札したタスクに対して、落札できたら正の報酬、できなかつたら負の報酬をケアマネージャから得ることで、どのタスクに入札したらよいかという入札ルールを学習する。つまり、ケアマネージャは、用意された評価モデルの学習、ヘルパーは、ケアマネージャの割り当てルールの学習を行うことになる。

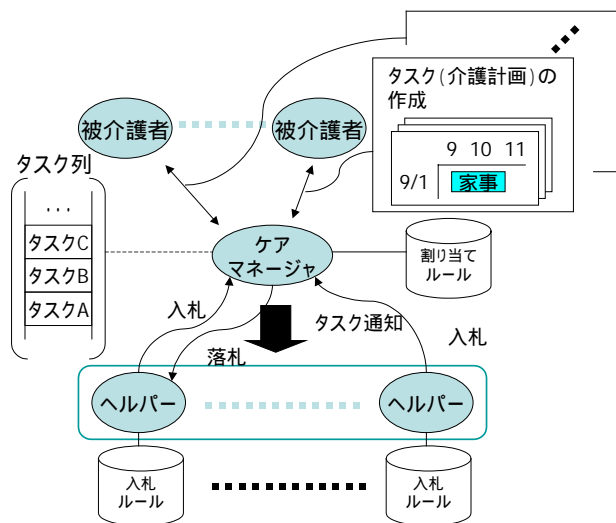


図1 システム概念図

4.2 学習について

本稿では、ヘルパーの学習に焦点を当てる。ヘルパーは、タスクの受信、入札、落札といった手順を繰り返しながら、ケアマネージャから報酬を得、各タスクに対する採択度を更新する。採択度とは、ヘルパーがそのタスクに入札したときの落札しやすさを示す値である。

ヘルパーの学習を行うための状態空間を考えると次のようになる。

- サービス内容（10通り）
- サービス要求レベル（3通り）
- 被介護者の性別（2通り）
- 被介護者の体重（3通り）
- 被介護者の年齢（2通り）

状態数は360（10×3×2×3×2）である。学習は、入札後のケアマネージャからの報酬によって行われるが、タスクの難度や入札人数により報酬の与え方を変える。入札人数が多いときに落札できたときほど、そのタスクに対する適正があると言えるからである。この報酬の違いにより、入札ルールを学習することで、時間帯の競合する複数のタスクが合った場合でも、より落札しやすいタスクに入札することが可能となる。

4.3 評価基準

ヘルパーの学習の評価は、落札したときにケアマネージャから与えられる報酬の和によって行う。1日分の報酬の和が多いほど、より落札しやすいタスクに入札していると言えるので、効率的な入札ができていると考えることができる。

5 おわりに

本稿では、契約ネットプロトコルにおける入札ルールを学習する手法について述べた。ヘルパースケジューリング問題に適用させ、各タスクに対する採択度を学習し、入札の際に考慮することで、効率の良い入札を行うことが可能となる。今後、実験により、学習を行わなかった場合との比較、また、学習手法、行動選択手法の違いによる比較、クラスタリングを用いて状態を分割した場合等の比較により、性能評価を行う予定である。

参考文献

- [1]Richard S.Sutton and Andrew G.Barto, 訳：三上貞芳・皆川雅章 共訳「強化学習」（森北出版株式会社,2000）
- [2]Watkins, C. J. C. H. and Dayan, P.: Technical Note: Q-Learning, Machine Learning 8, pp. 279-292(1992)