

要旨

製品生産を行う物流センターにおける水平倉庫において、使用されるマップカートの製品搬送稼働率を上げるためには、複数台のマップカート走行方法が大きく影響する。本研究では、水平倉庫内のマップカートの走行行動規則を強化学習によって求める。マップカートおよび走行レーンの状態は、Bluetooth 技術によって通信する手段を、実際にマップカートに積載する設計指針を得ることを目的とする。

1. はじめに

製品生産を行う物流センターでは、製品を倉庫に待機させ個別の注文票に従って区分けし出荷する。その際、区分けを行う者がどのように仕事をこなすかで、製品搬送の稼働率が左右される。

本研究では水平倉庫が置かれているセンター内での製品搬送を想定し、製品を倉庫から運び出す役目を持つマップカートの(準)最適行動規則を強化学習の一つである Q 学習を用いて求める。この行動規則を得るために必要な工場内のカートおよび走行レーンの状態表現は、Bluetooth 技術によって通信する手段を前提とする。

2. マップカート自律走行問題

生産された多数の製品を収納した倉庫から、工場の作業員はあらかじめ決められている仕分け票に従ってカートに入れ、個別にまとめて出荷する。この時に倉庫内の各セルへ短時間で到達する経路の決定、及び倉庫内に複数のカートが稼働している時カート同士の衝突およびデッドロックの回避がマップカートの自律走行において重要な課題となる。これらの問題を解決できれば、製品搬送の効率を格段に上昇させることができる。

2.1 倉庫の設計

本研究では水平倉庫内でのシミュレーションを行いカートの行動規則を獲得する。図のように倉庫内は製品が収納されているセルと、カートが走行するレーンで構成されている。レーンは簡略化のため走行を一方方向とする。

2.2 Bluetooth の搭載

Bluetooth とは、モバイル PC または通常の固定 PC、携帯電話、及びその他の周辺機器との間を、短距離ラジオを使用し無線でつなぐ新しい技術に付けられた呼び名である。Bluetooth 技術を使用することで、様々なコンピュータ機器と通信機器を、ケーブルを使用することなく無線的に通信することができる。これにより、様々なコンピュータ機器と電話を含めた通信機器間に無意識のうちにも場所や時間を選ばず簡単にコネクションが確立できる。

倉庫内のカートの行動を認識するために、各カートに 1 機、倉庫内の 3ヶ所に Bluetooth をそれぞれ設置する。カートの行動(走行中・停止中・セルから荷出し中 etc)を信号として 3ヶ所の Bluetooth に送信する。そして集まった情報を整理し各カートに伝達して、倉庫内の状況を判断する方法をとる。

3. Q 学習による最適行動規則の獲得

3.1 倉庫モデルの設定

倉庫内を走行するマップカートは 5 機。倉庫に配置されている 36 セル(6 列×6 セル)に格納されている製品の中から各カートに割り当てられている仕事すなわち指定された製品を収集し出口へ向かう。(図 1)

3.2 マップカートのモデル化

マップカート 1 台をオートマトン A_i と表現すると、全カートの集合は集団オートマトン A を、

$$A = \{A_i : i=1, 2, \dots, n\} \quad (1)$$

とする。ここで n はカートの総数である。

個々のオートマトン A_i を、

$$A_i = (I_i, O_i, S_i, F_i, G_i) \quad (2)$$

の 5 項目で記述する。ここで、 I_i, O_i, S_i, F_i, G_i はそれぞれ入力、出力、状態、状態遷移関数、出力関数である。

3.3 Q 学習

3.3.1 入力 I

各マップカートには製品搬送の内容、すなわち目的地となる倉庫内のセルの情報 J_i を入力とする。

$$I_i(t) = \{J_i\} \quad (3)$$

とする。ここで t は時刻を表す。

3.3.2 出力 O

出力は時刻 t におけるマップカートの行動を $O_i(t)$ とする。各 A_i の行動 O_i は、

$$O_i(t) = \{ \text{forward, right, stop, carry, exit} \} \quad (4)$$

直進中・右折・停止・荷出作業中・退場の 5 種類とする。

3.3.3 状態 S

カートと工場内に設置される Bluetooth は個々のセンサ範囲にいる他の Bluetooth を探知できるので、 $i+j$ 台の情報を組み合わせることで工場の状態 S_i を表現することができる。

$$S_i(t) = \{ B_k(t), B_j(t) \} \quad (5)$$

ここで、 $B_k(t)$ は他のマップカートの位置、 $B_j(t)$ は Bluetooth の信号である。

3.3.4 状態遷移関数 F

A_i の状態は時間毎に遷移する。この状態の遷移関数として Q 学習で用いられる Q 値の更新式を採用する。

$$Q_{i,t+1}(s_i(t), o_i(t)) = (1-\alpha)Q_{i,t}(s_i(t), o_i(t)) + \alpha \left[f_c + \gamma \max_{b \in O} Q_{i,t}(s_i(t+1), b) \right] \quad (6)$$

ここで、 $Q_{i,t}(s_i(t), o_i(t))$ は、状態-行動間に与えられる Q 学習値、 α, γ, f_c, O はそれぞれ学習率、減衰率、環境報酬値、行動集合である。

3.3.5 出力関数 G

出力関数 G_i は時刻 t におけるマップカートの状態 S_i から行動 O_i を確率的に決定する関数である。本研究ではボルツマン分布式によって以下の式に基づいて決定される。

$$G(s, o) = \arg \text{prob} \frac{\exp\{Q(s, o)/H\}}{\sum_{o_i \in O} \exp\{Q(s, o_i)/H\}} \quad (7)$$

ここで、 H は計算温度である。

3.4 数値計算実験

本アルゴリズムの有効性検証のため数値計算実験を行った。

各マップカートには次のような制約の下で走行を行

- う.
- 1) マップカートは 0~n 番まで番号付けされており, 番号順で出発する.
 - 2) マップカートは一方通行のレーンを進み, 分岐点では Q 値に従って決定されたレーンを選んで走行する.
 - 3) 倉庫セルでの荷出し作業には 5~10 単位時間を要する.
 - 4) マップカートの番号の低い順に優先権を持ち, 衝突の可能性がある場合は優先権の低いカートが停止する.

これらの制約は, 走行レーン上で各マップカートが衝突を起こさないために設定したものである. すべてのマップカートの走行を終えた時点で作業完了時間とする.

Q 値の更新式に用いられる学習パラメータ α, γ はそれぞれ 0.3, 0.8 とする. 環境報酬値 f_c は, マップカートが周回を終えた場合, 出発から終了までの行動に対して +0.1, 逆に途中で他のマップカートと衝突を起こして停止した場合は -0.1 とする. 計算温度 H は 1.0 とする.

その結果, 図 2 のような学習曲線を得た. 約 1000 回の学習で学習が収束している.

4. ガイドソフトウェアの設計

これまでのアルゴリズムに基づいて, 強化学習によって走行するマップカートのシミュレータを開発する.

4.1 シミュレータの機能

シミュレータのプロセスは以下のようになる.

- 1) 倉庫のモデル設計
倉庫の形状, セルの配置, カート用の走行レーンの設置を決定する.
- 2) ジョブの決定
各マップカートに製品の荷出しの命令情報を割り振る.
- 3) シミュレーション
強化学習による最適経路を計算した後, 実際に稼働したときの様子を表示する.

4.2 実装例

上記の設計に基づいて Windows で起動するシミュレータを図 3 に示す. 利用するユーザは水平倉庫の形状設計とマップカートの個数を入力するだけで, その場合の最適な製品搬送ルートを示すことができる.

5. まとめ

本研究では強化学習を用いて水平倉庫における製品搬送用マップカートの最適行動を獲得するとともに, そのプロセスをユーザに伝えるソフトウェアの開発を行った. 現段階では試作状態であるが, これをさらに現実の搬送工場に見立ててゆくことでさらに柔軟なソフトウェアを制作する.

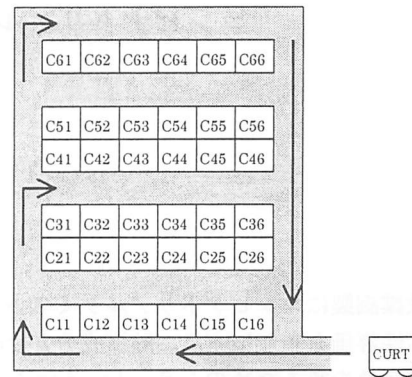


Fig. 1 水平倉庫

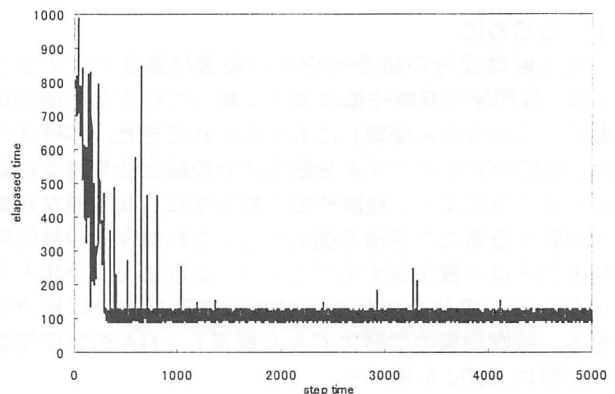


Fig. 2 学習曲線

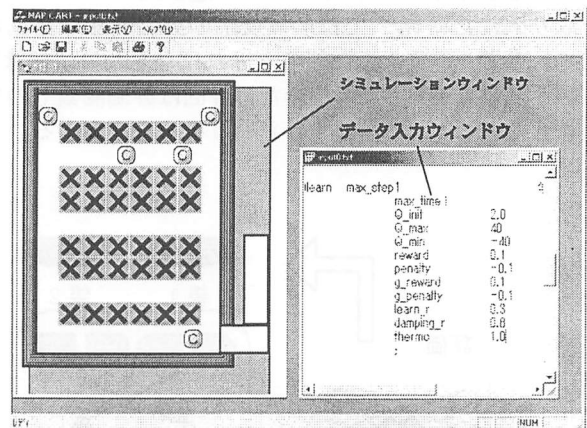


Fig. 3 シミュレータの実行例

参考文献

- (1) A.G.Barto, S.J.Brake and S.P.Singh: Learning to Act Using Real-time Dynamic Programming, Artificial Intelligence, 72(1995)81.
- (2) 古川正志・渡辺美知子・玉山裕也・嘉数侑昇, 一方向レーンを持つ FMS 工場における複数 AGV の運行スケジューリング, 日本機械学会講演論文集, (1996), 407-412.