

# 修士論文

## 移動ロボットのための深層学習を用いた 歩行者の位置予測とナビゲーションへの応用

Pedestrian Position Prediction Using Deep Learning for Mobile Robots  
and Its Application to Navigation

2024 年 12 月 16 日 提出

指導教員 林原 靖男 教授

千葉工業大学 先進工学部 未来ロボティクス学科

23S1030 藤原 柊



# 概要

移動ロボットのための深層学習を用いた  
歩行者の位置予測とナビゲーションへの応用

キーワード: 自律移動ロボット, Navigation, ROS, 深層学習

abstract

title

keywords:

# 目次

第 1 章	序論	1
1.1	背景 . . . . .	1
1.2	関連研究 . . . . .	3
1.3	目的 . . . . .	8
1.4	本論文の構成 . . . . .	8
第 2 章	要素技術	9
2.1	ナビゲーション . . . . .	9
2.2	ニューラルネットワーク . . . . .	10
2.3	再帰型ニューラルネットワーク . . . . .	11
2.4	Attention 機構 . . . . .	12
2.5	グラフニューラルネットワーク . . . . .	14
第 3 章	提案手法	15
3.1	本章の概要 . . . . .	15
3.2	学習アルゴリズム . . . . .	15
3.2.1	ネットワーク構造 . . . . .	15
3.2.2	学習方法 . . . . .	15
3.3	設計したネットワークの予備実験 . . . . .	15
3.3.1	実験方法 . . . . .	15
3.3.2	結果 . . . . .	15
3.3.3	考察 . . . . .	15

目次	vi
参考文献	16
付録	19
謝辞	20

# 目次

1.1	Overview of Social-LSTM method.* <sup>1</sup>	3
1.2	Framework overview of proposed SR-LSTM.* <sup>2</sup>	4
1.3	Illustration of the pedestrian-wise attention.* <sup>2</sup>	4
1.4	Illustration of the pedestrian-wise attention.* <sup>3</sup>	5
1.5	The Social-STGCNN Model.* <sup>4</sup>	5
1.6	Encoder structure.* <sup>5</sup>	6
2.1	Neural Network	10
2.2	Neural Network	11
2.3	Neural Network	13
2.4	Neural Network	14

# 表目次



# 第 1 章

## 序論

### 1.1 背景

近年，ロボット技術の進歩により，製造業，物流，サービス業などさまざまな分野で自律移動ロボットの活用が進んでいる．特に，人間とロボットが共存する環境では，ロボットは安全かつ効率的に動作するために，周囲の環境，特に人間の行動を理解し，それに適切に対応する必要がある．歩行者は動的な障害物であり，その予測不可能な動きはロボットのナビゲーションにとって大きな課題となっている．

この問題に対して，Social Force Model[1] などのルールベースの手法が提案されている．これらの手法は，歩行者の動きを物理的な力に基づいてモデル化し，歩行者同士や障害物との相互作用を考慮することで，より現実的な動きを再現することを目指している．しかし，これらの手法は依然として限界があり，特に複雑な環境や多様な人間の行動を完全に表現することは困難である．

これに対して，近年では深層学習によって歩行者の動きを予測する手法が注目されている．特に，リカレントニューラルネットワーク（RNN）[2, 3] やその変種である長短期記憶（LSTM）[4] ネットワークを用いた手法が有望視されている．これらの手法は，過去の歩行者の動きの履歴を考慮し，未来の動きを予測することができる．また，グラフ畳み込みニューラルネットワーク（GCN）[5] やグラフアテンションネットワーク（GAT）[6] を用いた手法も注目されている．これらの手法は，人間をノード，人間同士の相対距離をエッジとしてグラフ化し，相互作用を効率的に学習することができる．

しかし，これらの手法を採用している研究では，移動ロボットを対象としたものが少なく，移動ロボットのナビゲーションに予測結果を応用した例の報告が少ない．また，移動ロボットを対象にしたものでも，環境に介入しており，ロボットのナビゲーション性能を評価するための実験環境が整っていない場合が多い．

また，移動ロボットのナビゲーションに予測結果を応用する場合，歩行者同士の相互作用の他にロボットと人間にも相互作用が存在する．そのため，歩行者だけでなく，ロボットの行動も考慮した軌道予測が必要である．

我々は，前述の 2 つの課題を克服するために，ロボットの行動を考慮した歩行者の軌道予測を行い，その予測結果を用いたナビゲーションシステムを提案する．そして，そのシステムによるナビゲーション性能を検証する．

## 1.2 関連研究

歩行者の複雑な動きを予測するために、深層学習を応用しようという研究が、近年注目を集めている。深層学習を歩行者の軌道予測に応用した例として、Alexandre ら [1] は、人間の動きを学習し、未来の軌跡を予測できる LSTM モデルを提案している。この研究は、歩行者の軌道予測に着目した最も初期の深層学習モデルの一つである。Social-LSTM は、Fig. 1.1 のように各軌跡に 1 つの LSTM を使用し、プーリング機構を用いて、LSTM 間で情報を共有している。このシステムにより、複数の個人にわたって相互作用から生じるさまざまな非線形行動をうまく予測することに成功した。

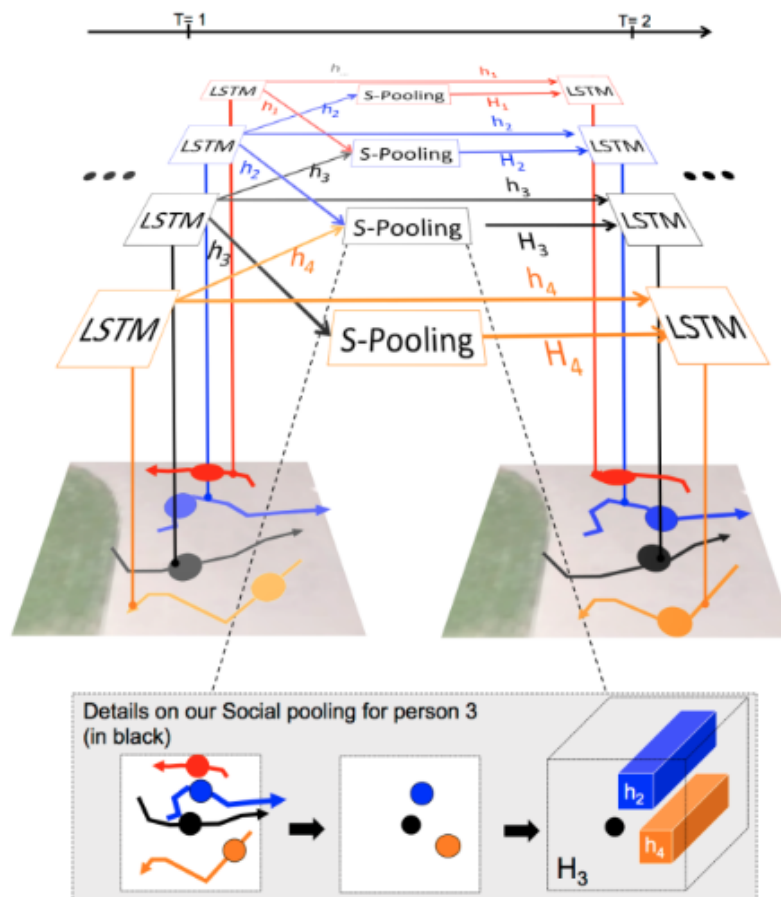


Fig. 1.1 Overview of Social-LSTM method.\*<sup>1</sup>

\*<sup>1</sup> [1] より引用

Zhang ら [7] は, LSTM のシーケンス学習能力にエージェント間の相互作用をモデル化するため, Fig. 1.2 のような LSTM ネットワークの情報精密化モジュールを提案している. この研究では, Fig. 1.3 のように Attention メカニズムにより, 各歩行者の他者への影響を重み付けしている.

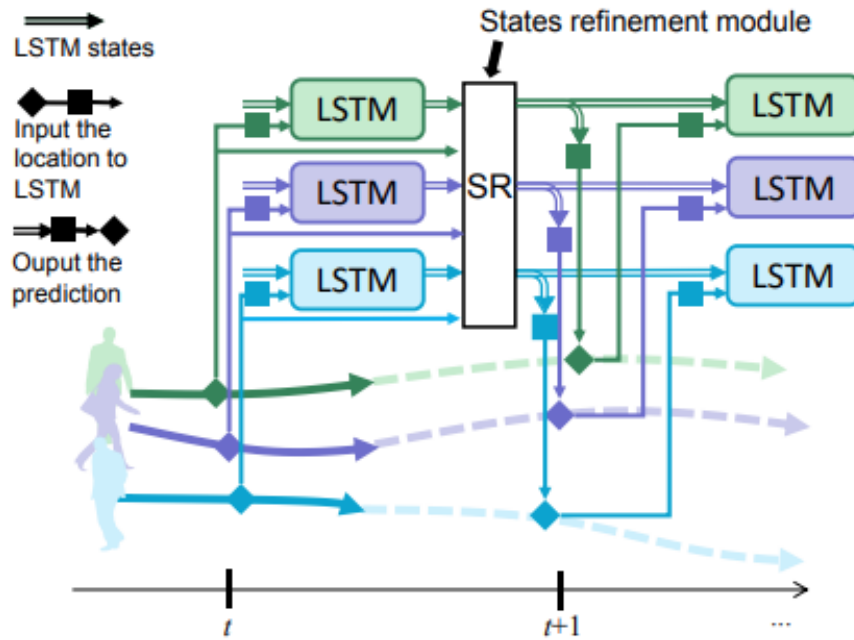


Fig. 1.2 Framework overview of proposed SR-LSTM.\*<sup>2</sup>

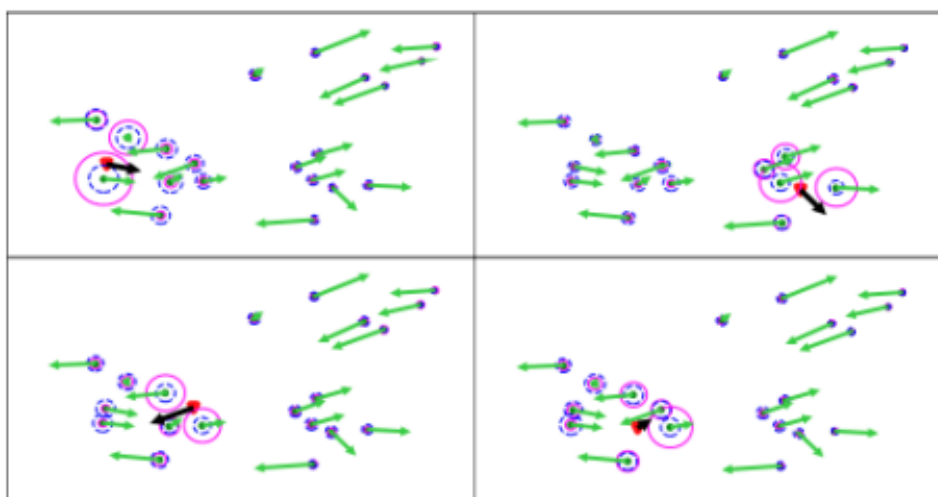


Fig. 1.3 Illustration of the pedestrian-wise attention.\*<sup>2</sup>

\*<sup>2</sup> [7] より引用

同様に、歩行者の注意に着目した研究を Kosaraju ら [8] も行っている。この研究は、LSTM を用いて各歩行者の軌跡をモデル化し、グラフアテンションネットワーク (GAT) を用いて、現実的な歩行者の軌道予測を実現している。この研究では、グラフは Fig. 1.4 のように、歩行者をノード、距離をエッジとしてモデル化される。このグラフ構造を用いることで、歩行者間の相互作用を効果的に捉え、より正確な軌道予測が可能となる。

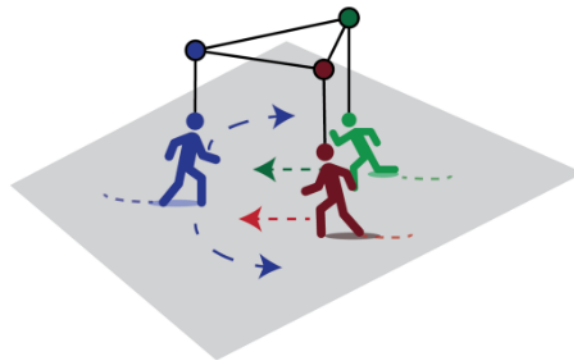


Fig. 1.4 Illustration of the pedestrian-wise attention.\*3

また、Mohanmed ら [9] は軌跡をグラフとしてモデル化し、歩行者間のユークリッド距離で重み付けされたエッジは歩行者間の相互作用を表している。Fig. 1.5 のように、モデルはグラフ畳み込みニューラルネットワークと時間畳み込みネットワーク (TCN) を用いて、時空間グラフ上で動作し、一度にシーケンス全体を予測できるようにしている。

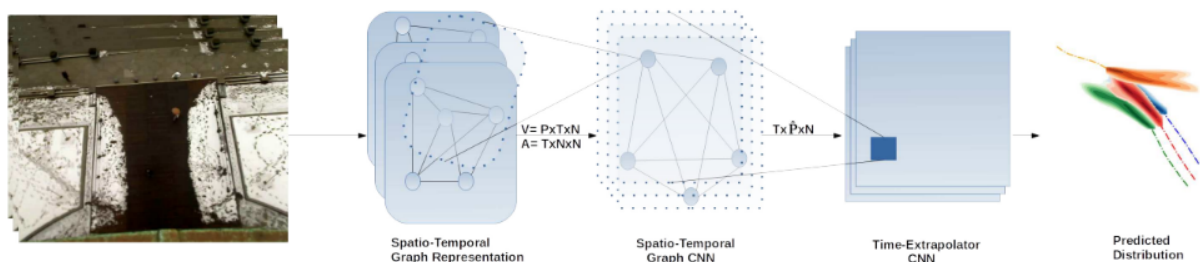


Fig. 1.5 The Social-STGCNN Model.\*4

\*3 [8] より引用

\*4 [9] より引用

移動ロボットで歩行者の軌道予測を行う場合，歩行者同士の相互作用の他にロボットと人間にも相互作用が存在する．そのため，歩行者だけでなく，ロボットの行動も考慮した軌道予測が必要である．

丹野ら [10] は，歩行者の軌道予測に過去の動きだけでなく，Fig. 2 のように，ロボットが選択する将来の動きを考慮して予測をしている．実験では，ロボットの将来の行動を考慮しない場合と比較して，ロボットの動きに合わせて変化する歩行者の動きを予測できることが確認されている．しかし，学習器の事前学習でデータセット内の人間をロボットと見なして学習している．つまり，歩行者の行動を移動ロボットの行動に近似できるという前提の上で学習している．それならば，モデル構造のエンコード部分を除外し，ネットワークを人間とロボットで共有化させることができる可能性がある．これは，ネットワークの肥大化を防いだり，学習の手間を低減させるねらいがある．4 章では，この論文を参考に実験を行う．

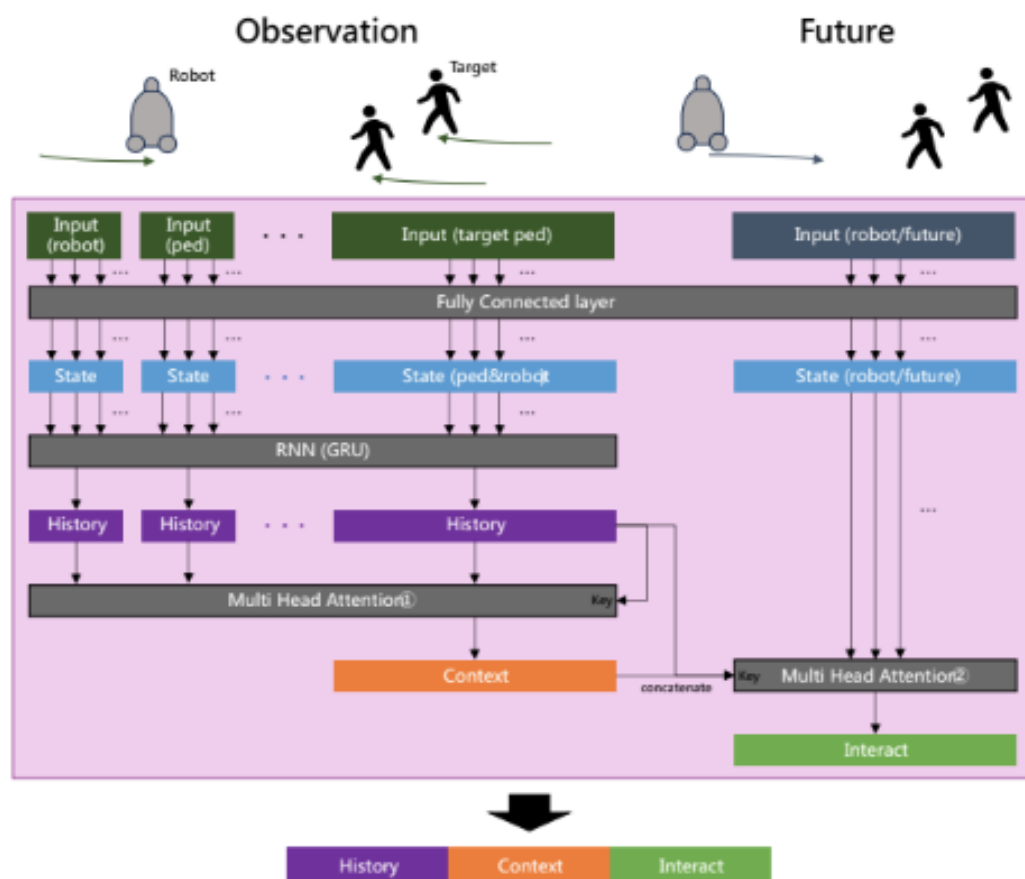


Fig. 1.6 Encoder structure.\*5

\*5 [10] より引用

このように，深層学習を歩行者の軌道予測タスクに応用する例は少なくない．しかし，その多くは学習器の訓練を鳥瞰図のデータセットを用いて行っている．それに伴い，移動ロボットを対象にしたものが少ない．また，移動ロボットを対象している場合においても，予測対象にトラッキングセンサなどを取り付け，環境に介入している場合やシミュレータで予測対象の真値を取得しており，移動ロボットのセンシングしたデータだけで完結していないものが多い．そのため，深層学習を応用した歩行者の予測結果を移動ロボットのナビゲーションで利用しようとする研究例は現時点ではあまり報告されていない．

### 1.3 目的

1.2 章で紹介した深層学習によるアプローチは，歩行者の複雑な動きを予測する上で有用である．しかし，環境に介入せずに予測結果を移動ロボットのナビゲーションに応用している事例は少ない．また，前述のように歩行者の軌道予測を移動ロボットで活用する際，ロボットと人間の相互作用も考慮すべきである．そこで本論文では，環境に介入することのないロボットの行動を考慮した歩行者の軌道予測を提案する．また，提案するシステムを移動ロボットのナビゲーションに応用し，安全かつ効率的にロボットが移動できるかという観点の検討を行う．

### 1.4 本論文の構成



## 第 2 章

# 要素技術

### 2.1 ナビゲーション

LiDAR やオドメトリなどのセンサを用いて自律走行する移動ロボットの多くは、ROS Navigation stack[11, 12] を利用している。例えば、つくばチャレンジと呼ばれるつくば市内の遊歩道等を移動ロボットが自律走行する技術チャレンジにおいて、原らが行った技術調査[13]によると、少なくとも 76 チーム中 22 チームが自律走行で使用していた。これは、オープンソースのソフトウェアに限れば最多であった。ROS Navigation stack は以下のような主要な機能を含んでいる。

- 自己位置推定 (Localization)

AMCL (Adaptive Monte Carlo Localization) アルゴリズムを利用して、ロボットの現在位置を推定する。

- 地図生成 (Mapping)

SLAM (Simultaneous Localization and Mapping) 技術を活用し、環境のマッピングを行う。

- 経路計画 (Path Planning)

Dijkstra 法 [14] や A\*アルゴリズム [15] を利用して、最適経路を計画する。

- 障害物回避 (Obstacle Avoidance)

センサから取得したデータに基づいて、障害物を回避しながら移動する。

## 2.2 ニューラルネットワーク

ニューラルネットワーク (Neural Network) は、人間の脳の神経回路に着想を得た計算モデルであり、機械学習や人工知能の分野で重要な役割を果たしている。大量のデータから学習し、パターン認識や予測などの複雑なタスクを遂行する能力を持つ。パーセプトロンを入れる

ニューラルネットワークは、多数のノード (ニューロン) が層状に接続された構造を持つ。各ノードは、入力信号を受け取り、重み付けを行い、活性化関数を適用することで出力信号を生成する。これらのノードは、入力層、隠れ層、出力層の3つの層に分けられる。

- 入力層

入力データを受け取る層であり、各ノードは入力データの各特徴に対応する。

- 隠れ層

入力層と出力層の間にある層であり、複数の隠れ層を持つことができ、複雑な非線形関係を学習することが可能になる。

- 出力層

ネットワークの最終的な出力を生成する層である。

各層のノードは、次の層のノードと接続されており、接続には重みが割り当てられている。学習プロセスでは、これらの重みを調整することで、ネットワークが目的のタスクを実行できるように最適化される。多層構造を持つことで、複雑なパターンや特徴を学習し、従来では困難であった高精度な予測を実現した学習手法を深層学習と呼ぶ。深層学習は、画像認識、自然言語処理、音声認識といった多岐にわたる分野で活用されており、YOLO[16] や ChatGPT[17, 18, 19] などの成功例が多数存在し、高い性能が評価されている。



Fig. 2.1 Neural Network

## 2.3 再帰型ニューラルネットワーク

再帰型ニューラルネットワーク (Recurrent Neural Network, RNN) [2, 3] は、時系列データや系列データを処理するために設計されたニューラルネットワークの一種である。RNN は、隠れ層の出力を次の時間ステップの入力として再利用することで、データの時間的依存性をモデル化する特徴を持つ。この仕組みにより、従来のニューラルネットワークでは難しかった順序情報を考慮した学習が可能になる。

RNN の基本的な構造は、入力層、隠れ層、および出力層から成る。時系列データを扱う際には、系列データの各時点での入力が増加する RNN に与えられ、隠れ層が現在の入力と過去の隠れ状態を組み合わせることで次の隠れ状態を生成する。この隠れ状態は過去の情報を保持する役割を担うため、時系列の依存関係を学習する際に重要である。

数式で表すと、RNN の隠れ状態  $h_t$  は以下のように計算される。

$$h_t = f(W_{ih}X_t + W_{hh}h_{t-1} + b_n)$$

ここで、 $x_t$  は時刻  $t$  の入力、 $h_{t-1}$  は時刻  $t - 1$  の隠れ状態、 $W_{ih}$  と  $W_{hh}$  はそれぞれ入力重みと隠れ層の重み行列、 $b_n$  はバイアス項、そして  $f$  は非線形活性化関数である。

RNN には、勾配消失問題 [20, 21, 22] や勾配爆発問題といった課題がある。これらの問題は、長い系列データを扱う場合にネットワークが十分な長期依存性を学習できない原因となる。この課題を解決するために、LSTM (Long Short Term Memory) [4] や GRU (Gated Recurrent Unit) [23] といった改良モデルが提案されている。これらのモデルは、ゲート機構を用いて重要な情報を選択的に保持または忘却することで、長期依存性を効率的に学習できるように設計されている。



Fig. 2.2 Neural Network

## 2.4 Attention 機構

Attention 機構 [24] は、ニューラルネットワークの文脈で導入された手法であり、特に自然言語処理や画像認識などのタスクにおいて顕著な成果を上げている技術である。この手法の主な目的は、入力データの中から重要な情報に焦点を当て、それ以外の情報を適度に無視することである。これにより、モデルはより効率的かつ効果的に学習を進めることが可能となる。

Attention 機構の基本概念は、入力に含まれる各要素がどの程度重要であるかをスコア付けし、そのスコアを基に加重平均を計算することである。これにより、モデルは特定のタスクにおいて重要な部分に「注意」を向けることができる。このスコア付けは、キー (Key)、クエリ (Query)、バリュー (Value) の3つの成分を用いた計算で表現される。以下に、それぞれの構成要素について説明する。

- クエリ (Query)

クエリは、特定の出力を生成する際に、どの部分の情報が重要であるかを判断するための基準となるベクトルである。入力シーケンス全体に対して「どこを注目すべきか」を問う役割を果たす。

- キー (Key)

キーは、入力データの各部分がどの程度重要であることを示す指標となるベクトルである。クエリと比較されることで、入力内の異なる部分の関連性が評価される。

- バリュー (Value)

バリューは、実際に計算される情報の内容を保持するベクトルである。計算された注意スコアに基づいて加重平均が適用され、最終的な出力が生成される。

Attention 機構の計算では、クエリとキーの内積を用いて関連性を評価し、ソフトマックス関数で正規化することで注意スコアを得る。その後、これらのスコアを用いてバリューに加重平均を適用し、最終的な結果を出力する。

この手法は、特に Transformer アーキテクチャにおいて重要な役割を果たしている。Self-Attention として知られる拡張では、入力シーケンスの各要素が他のすべての要素とどのように関連しているかを同時に評価する。これにより、従来のリカレントニューラルネットワーク

(RNN) や畳み込みニューラルネットワーク (CNN) [25, 26] の限界であった長距離依存性の問題を効果的に解決することが可能となる。

Attention 機構は，自然言語処理のみならず，画像認識や音声処理といった他の分野にも応用されている．例えば，画像セグメンテーションタスクでは，画像内の特定の領域に焦点を当てることで，より精度の高い結果が得られる．このように，Attention 機構は汎用的かつ強力なツールとして，多くの応用領域でその価値を証明している．



Fig. 2.3 Neural Network

## 2.5 グラフニューラルネットワーク

グラフニューラルネットワーク (GNN) は、グラフ構造データを処理するために設計されたニューラルネットワークの一種である。GNN は、ノード (頂点) とエッジ (辺) からなるグラフを入力として受け取り、各ノードの特徴を学習し、ノード間の関係性を考慮した上で予測や分類を行う。

GNN の基本的なアイデアは、各ノードの特徴ベクトルをその隣接ノードの特徴ベクトルと組み合わせて更新することである。このプロセスは、メッセージパッシングと呼ばれ、複数の層を通じて繰り返される。各層では、ノードの特徴ベクトルが隣接ノードからの情報を集約し、更新される。

代表的な GNN のアーキテクチャには、以下のようなものがある：

- **Graph Convolutional Networks (GCN):** グラフ畳み込みネットワークは、畳み込み操作をグラフデータに適用することで、ノードの特徴を更新する。
- **Graph Attention Networks (GAT):** グラフアテンションネットワークは、各エッジに異なる重みを付けることで、重要な隣接ノードからの情報を強調する。
- **GraphSAGE:** サンプリングと集約の手法を用いて、効率的に大規模なグラフデータを処理する。

GNN は、ソーシャルネットワーク分析、化学分子の特性予測、知識グラフの補完など、さまざまな分野で応用されている。これらのネットワークは、グラフ構造データの複雑なパターンを捉える能力があり、従来の手法では難しかった問題を解決することができる。



Fig. 2.4 Neural Network

## 第 3 章

# 提案手法

### 3.1 本章の概要

本章では、2 章で述べたようにロボットの行動を考慮した歩行者の軌道予測を行う手法を提案する。学習に用いたアルゴリズムに関する内容を 3.2 章、

### 3.2 学習アルゴリズム

以下では、予測を行うネットワーク構造とその学習方法に関して説明する。

#### 3.2.1 ネットワーク構造

#### 3.2.2 学習方法

### 3.3 設計したネットワークの予備実験

#### 3.3.1 実験方法

#### 3.3.2 結果

#### 3.3.3 考察

## 参考文献

- [1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 961–971, 2016.
- [2] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, Vol. 323, No. 6088, pp. 533–536, 1986.
- [3] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation, parallel distributed processing, explorations in the microstructure of cognition, ed. de rumelhart and j. mcclelland. vol. 1. 1986. *Biometrika*, Vol. 71, No. 599-607, p. 6, 1986.
- [4] S Hochreiter. Long short-term memory. *Neural Computation MIT-Press*, 1997.
- [5] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [6] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, Vol. 1050, No. 20, pp. 10–48550, 2017.
- [7] Pu Zhang, Wanli Ouyang, Pengfei Zhang, Jianru Xue, and Nanning Zheng. Sr-lstm: State refinement for lstm towards pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12085–12094, 2019.
- [8] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, Hamid Rezaatofghi, and Silvio Savarese. Social-bigat: Multimodal trajectory forecasting



- using bicycle-gan and graph attention networks. *Advances in neural information processing systems*, Vol. 32, , 2019.
- [9] Abdullah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14424–14432, 2020.
- [10] 丹野壮一郎, 田村雄介, 平田泰久. 移動ロボットが選択する将来の軌道が歩行者に与える影響を考慮した歩行者軌道予測. 第 24 回計測自動制御学会システムインテグレーション部門講演会, pp. 2423–2425, 2023.
- [11] ros planning. Github - ros-planning/navigation: Ros navigation stack. code for finding where the robot is and how it can get somewhere else. <https://github.com/ros-planning/navigation>. (Accessed on 2024-12-16).
- [12] ros navigation. Github - ros-navigation/navigation2: Ros 2 navigation framework and system. <https://github.com/ros-navigation/navigation2>. [Online; accessed 2024-12-16].
- [13] 原祥堯, 萬礼応, 富沢哲雄, 伊達央, 大川一也, 大矢晃久. つくばチャレンジ 2023 全チームの技術動向調査. ロボティクス・メカトロニクス講演会 2024 講演論文集, pp. 1P1–N02(1)–1P1–N02(4), 2024.
- [14] Edsger W Dijkstra. A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: his life, work, and legacy*, pp. 287–290. 2022.
- [15] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, Vol. 4, No. 2, pp. 100–107, 1968.
- [16] J Redmon. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
- [17] Alec Radford. Improving language understanding by generative pre-training. 2018.
- [18] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, Vol. 1, No. 8, p. 9, 2019.

- [19] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, Vol. 33, pp. 1877–1901, 2020.
- [20] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.
- [21] Thomas Weinlein. Diplomarbeit im fach informatik.
- [22] J Schmidhuber. Deep learning in neural networks: An overview, 2015.
- [23] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- [24] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- [25] Kunihiro Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, Vol. 36, No. 4, pp. 193–202, 1980.
- [26] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, Vol. 1, No. 4, pp. 541–551, 1989.
- [27] Dirk Helbing and Peter Molnar. Social force model for pedestrian dynamics. *Physical review E*, Vol. 51, No. 5, p. 4282, 1995.
- [28] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, Vol. 65, No. 6, p. 386, 1958.

## 付録

# 謝辞

本研究を進めるにあたり，1年に渡り，熱心にご指導を頂いた林原靖男教授に深く感謝いたします．