

第 1 章

要素技術

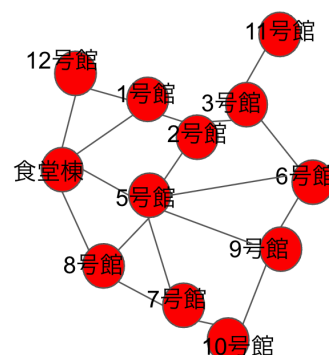
1.1 トポロジカルマップ

私たちの身の回りには様々な種類の地図があり，活用されている．例えば， Fig. 1.1a に表すメトリックマップと呼ばれる地図は，普段人が目的地まで移動する際に用いられる．しかし，本研究で用いているトポロジカルマップは Fig. 1.1b のような形をしている．メトリックマップがやや複雑な形をしているのに対し，トポロジカルマップはより簡潔に，環境を抽象的に表現することができる．

トポロジカルマップは，大きく分けてノードとエッジの 2 つの要素により構成されている． Fig. 1.1b では，赤い丸の図形で表現されているのがノードである．ノードには，地図の作成者が好きな情報を入れることができる．もう 1 つの要素であるエッジは，それぞれのノード同士を接続するのに用いられる．ノード同士に関係性がある場合，ノードとノードはエッジにより接続される．



(a) Map representing Chiba Institute of Technology Shin-Narashino Campus with metric map (出典：国土地理院より一部を加工して作成 [?])



(b) Map representing Chiba Institute of Technology Shin-Narashino Campus with topological map

Fig. 1.1: Example of metric and topological map

1.2 シナリオ

本研究におけるシナリオとは、人間の道案内を文章で表現したものである。例えば、Fig. 1.2 に表すような道案内の場合、シナリオは「次の角まで直進・右折・突き当たりまで直進・停止」のようになる。また、シナリオは「次の角」や「突き当たり」のような条件と「直進」、「右折」のような行動の組み合わせにより作成される。

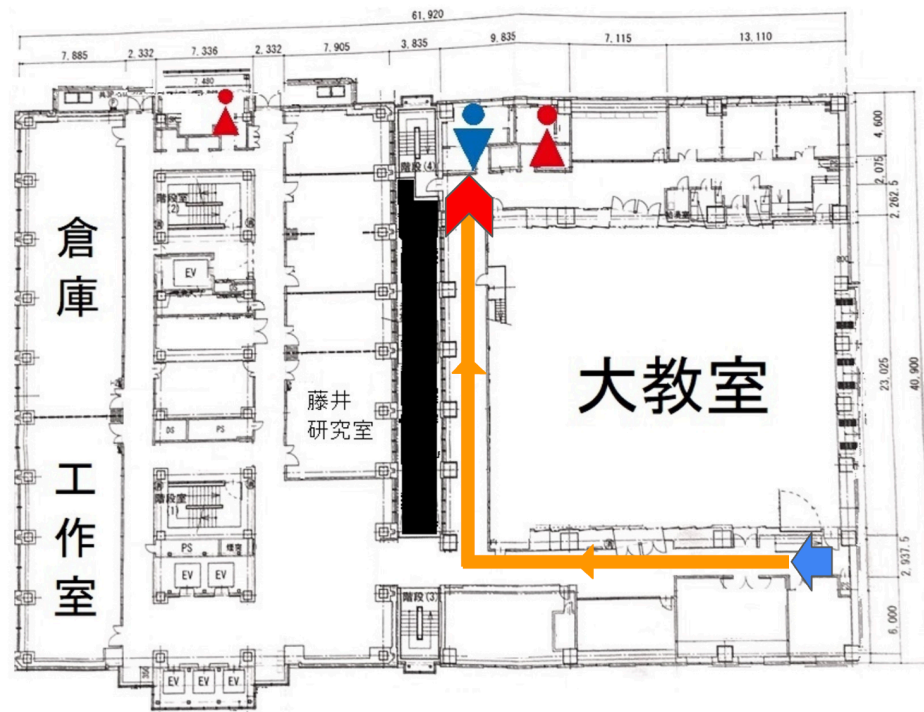


Fig. 1.2: An example of human directions

1.3 Neural Network

ニューラルネットワークとは、人間の脳内の神経細胞（ニューロン）のネットワーク構造を模して作られた数式的なモデルである。ネットワークは、Fig. 1.3 に示すように入力層、出力層、1 つ以上の隠れ層（中間層）により構成されており、円で表されているものはニューロンと呼ぶ。また、線により結ばれているニューロンとニューロンの間には重みを持っており、重みは接続されているニューロン間のつながりの強さを表現している。ニューラルネットワークを用いることにより、複雑な回帰問題や分類問題などを解くことができる。

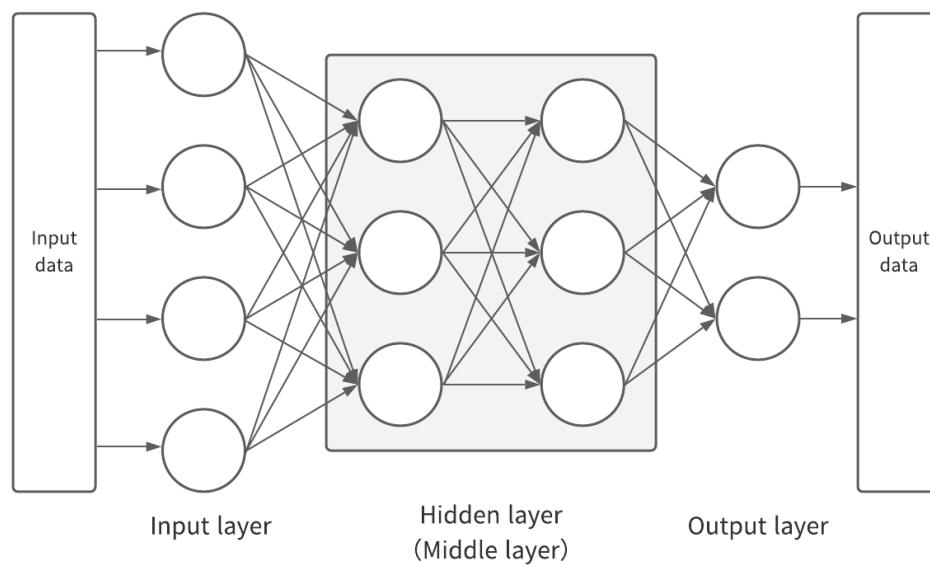


Fig. 1.3: Neural network architecture

1.4 Convolutional Neural Network(CNN)

CNN はディープラーニングの手法の 1 つであり，ニューラルネットワークを元に作られている．ニューラルネットワークは入力層，中間層，出力層により構成されているが，CNN はニューラルネットワークの中間層に畳み込み層，プーリング層を組み込んだネットワークである．CNN により画像認識をする場合の各層の役割を以下に記す．

- ・畳み込み層：入力した画像データにフィルタをかけ，特徴マップを出力することで特徴抽出を行う
- ・プーリング層：特徴マップをいくつかの領域に分割し，各領域で最大値または平均化を行うことで重要な情報を残しながら画像を縮小する．

CNN は機械学習分野において主に画像・動画認識の分野で高い性能を発揮する．LeCun ら [?] は CNN を用いて手書きの画像データから数字を識別できることを示した．また，使用されたネットワーク構造を Fig. 1.4 に表す．

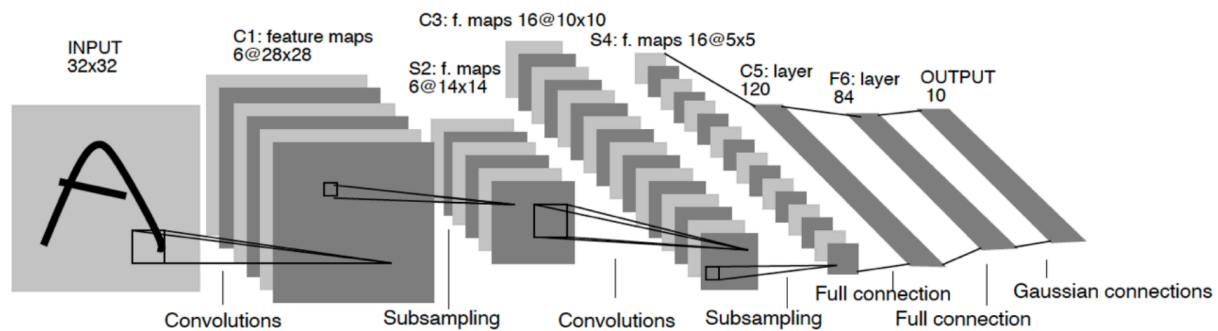


Fig. 1.4: Architecture of CNN (出典 : [?])

1.5 You Only Look Once(YOLO)

本研究で用いる YOLO[?] [?] [?] は，リアルタイム物体検出アルゴリズムである．YOLO は，画像の RGB データの配列を CNN に入力し，画像中のどの範囲に物体が存在しているのかを表すバウンディングボックスの情報と，ボックス内の物体がどのクラスに属しているのかを確率とともに表すクラス確率の情報を出力する． Fig. 1.5 は，YOLO を用いて画像中の物体を検出している様子である．

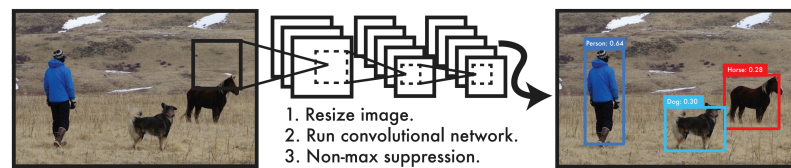


Fig. 1.5: The YOLO Detection System.(出典:[?])

左の画像データを入力した結果，画像からは 3 つの物体が検出されている．また，それぞれの物体が Person 0.64 , Dog 0.30 , Horse 0.28 の確率で予測されていることが確認できる．