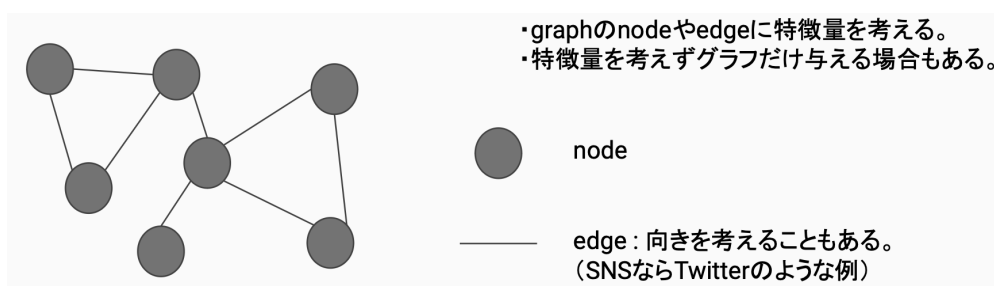


1 グラフニューラルネットワーク概論

abstract 世の中にはグラフ構造を持ったデータがたくさんあります。グラフニューラルネットワークは、このようなグラフ構造を持ったデータのための機械学習モデルです。この講義では、グラフとは何か、グラフ構造を持つデータの機械学習課題にどのようなものがあるのか、最も基本的なグラフニューラルネットワークであるKipfとWellingの空間畳み込みの紹介の3点を解説します。

1.1 グラフとは何か

グラフ (graph) は、**頂点** (node) と**辺** (edge) で構成され、いくつかの頂点の間を辺で結んだものです。特に今回は、頂点や辺に特徴量 (feature) を対応させたものを考えます。頂点の特徴量は**node feature**、辺の特徴量は**edge feature**とよばれます。このようなグラフを特に**attributed graph**といいます。



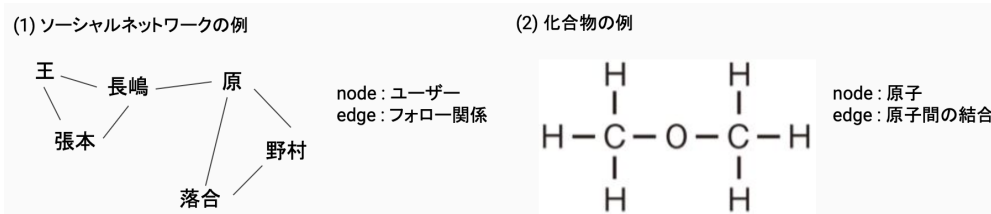
グラフ構造を持つデータの例には、ソーシャルネットワークサービスの友達関係 (例: Facebook) や化学式があります。ソーシャルネットワークサービスの友達関係の場合、頂点をユーザー、辺をユーザー間の友達関係とし、それぞれの特徴量として

- node feature : 各ユーザーの性別、年齢、職業
- edge feature : 友達関係になって何年目か

を考えることができます。また化学式であれば、頂点を原子、辺を原子間の結合とし、それぞれの特徴量として

- node feature : どの原子かを表すone-hot表現
- edge feature : 何重結合か

を考えることができます。



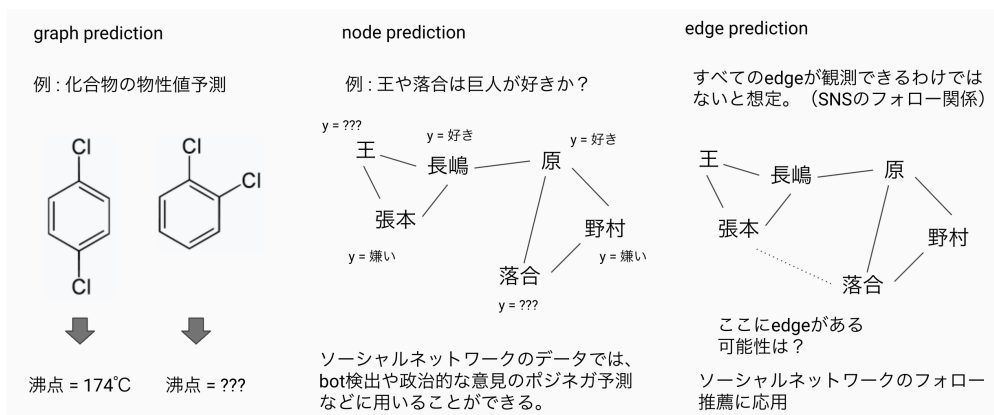
辺には向きがついている場合とそうでない場合を考えることができ、それぞれ**有向グラフ** (directed graph)、**無向グラフ** (undirected graph) といいます。Twitterのフォロー関係やリプライ関係、論文の引用関係、Wikipediaの関連項目は有向グラフのデータと考えることができます。

1.2 グラフ構造を持つデータの機械学習課題

グラフ構造を持つデータの機械学習課題には主に3種類のものがあります。

- graph prediction
- node prediction
- edge prediction

graph predictionは、グラフ全体の性質を予測する課題です。例えば、化合物の物性値を予測するような課題が該当します。**node prediction**は、グラフの各頂点の性質を予測する課題です。例えば、Facebookのあるユーザーの職業を予測するような課題が該当します。**edge prediction**は、グラフの各辺の性質を予測したり、存在を予測するような課題です。例えば、Facebookの友達関係にない二人のユーザーの間に友達関係を推薦するかという課題が該当します。



1.3 最も基本的な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 (graph neural network) は、グラフ構造を持つデータに対する機械学習モデルです。ここでは、基本的なグラフニューラルネットワークである**KipfとWeillingの空間畳み込み層** (Kipf-Welling spatial convolution layer) を用いたグラフニューラルネットワークを紹介します。

以下では無向グラフで、特に頂点のみが特徴量を持つ機械学習課題を考えます。

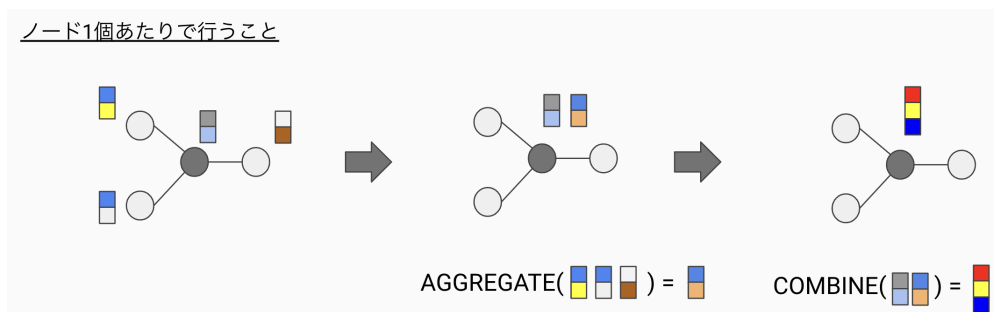
1.3.1 空間畳み込み層

一般に**グラフ畳み込み層** (graph convolution layer) とよばれる層には、**空間畳み込み層** (spatial convolution layer) と**スペクトラル畳み込み層** (spectral convolution layer) があり、KipfとWellingの空間畳み込み層はこの空間畳み込み層の一種です。以

下では空間畳み込み層とは何かを解説します。スペクトラル畳み込み層の理解にはグラフ理論の知識が必要なので次回以降で解説をしていきます。

空間畳み込み層は、近傍の頂点の特徴量を用いて、各頂点の特徴量をより予測に適した特徴量に更新するような層です。このように近傍の頂点からの情報を優先するバイアスを取り入れる性質を**relative inductive bias**といいます。具体的には、以下の2つのステップで構成されます。

1. **集約 (aggregation)** : 注目している頂点の近傍の頂点の特徴量を単純集計する。
2. **結合 (combine)** : 注目している頂点の特徴量と集約で作った特徴量を組み合わせる。



頂点 v の特徴量を h_v 、頂点 v の近傍を $\mathcal{N}(v)$ と表すことにしましょう。集約には合計や平均、最大値などが考えられます。例えば、合計の場合は次のようになります。

$$\text{agg}[v] = \sum_{v' \in \mathcal{N}(v)} h_{v'}$$

結合にはいろいろな方法がありますが、例えば注目している頂点 v の特徴量 h_v と集約によって得られた特徴量 $\text{agg}[v]$ を同じ重み Θ をつけた和にしたあと、バイアス項 β を加えて、非線形な活性化関数 a を適用する方法があります。重み Θ とバイアス項 β はデータから学習したい空間畳み込み層のパラメータになります。また活性化関数にはよくReLUが用いられます。

$$h_v \leftarrow a(\beta + \Theta h_v + \Theta \text{agg}[v])$$

1.3.2 KipfとWellingの空間畳み込み層

平均値を用いた集約は次のように定義されます。頂点 v をなにかしらの頂点と結ぶ辺の数を**次数 (degree)** といい、 $\text{deg}(v)$ と表します。

$$\text{agg}[v] \leftarrow \frac{1}{\text{deg}(v)} \sum_{v' \in \mathcal{N}(v)} h_{v'}$$

これを、多くの頂点と辺が結ばれている頂点ほどユニークな情報を提供しないという仮説に基づいて、重みをつけながら集約を行う方法に変更したものが**Kipf normalization**