Self Study

Based on my Understanding of the Subject

Graph Neural Network

Compilation of my study materials

Author: Dr. Md Arafat Hossain Khan

Last updated : November 29, 2022

Contents

Chapter 1

Fundamentals

Artificial intelligence is a tool, not a threat.Artificial intelligence is a tool, not a threat.Artificial intelligence is a tool, not a threat.

Rodney Brooks

1.1 Graph Neural Network Model Mathematics

1.1.1 Labels of a graph

A graph G is a pair (N, E), where $N = \{n_1, n_2, \ldots, n_{|N|-1}, n_{|N|}\}$ is the set of nodes and $E = \{e_1, e_2, \ldots, e_{|E|-1}, e_{|E|}\}$ is the set of edges. $\mathcal{N}(n)$ stands for the neighbors of $n \in N$. If an *undirected* edge e connects node x and y, we can also denote e as (x, y) or (y, x). All nodes are embedded into some euclidean space of dimention $d_N \in \mathbb{N}$ and all edges are embedded into some euclidean space of dimention $d_E \in \mathbb{N}$, e.g.

$$\text{label (embedding) for node } n \text{ is defined as } \ell_n : n \mapsto \mathbb{R}^{d_N}, \quad \text{where } n \in N$$
$$\text{label (embedding) for edge } e \text{ is defined as } \ell_e : e \mapsto \mathbb{R}^{d_E}, \quad \text{where } e \in E$$

The notion of labels can be extended to multiple nodes and edges. Suppose, we take a node set, $\mathfrak{N} = \{\mathfrak{n}_1, \mathfrak{n}_2, \ldots, \mathfrak{n}_{|\mathfrak{N}|-1}, \mathfrak{n}_{|\mathfrak{N}|}\} \subseteq N$. Then we define,

$$\ell_{\mathfrak{N}} : \mathbb{R}^{d_N \times |\mathfrak{N}|} \mapsto \mathbb{R}^{d_N} \quad \text{as} \quad \left(\ell_{\mathfrak{n}_1}, \ell_{\mathfrak{n}_2}, \ldots, \ell_{\mathfrak{n}_{|\mathfrak{N}|-1}}, \ell_{\mathfrak{n}_{|\mathfrak{N}|}}\right) \mapsto \mathbb{R}^{d_N}$$

Here, we want $\ell_{\mathfrak{N}}$ to be permutation invariant. Similarly, we take an edge set, $\mathfrak{E} = \{\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_{|\mathfrak{E}|-1}, \mathfrak{e}_{|\mathfrak{E}|}\} \subseteq E$. Then we define,

$$\ell_{\mathfrak{E}} : \mathbb{R}^{d_E \times |\mathfrak{E}|} \mapsto \mathbb{R}^{d_E} \text{ as } \left(\ell_{\mathfrak{e}_1}, \ell_{\mathfrak{e}_2}, \ldots, \ell_{\mathfrak{e}_{|\mathfrak{E}|-1}}, \ell_{\mathfrak{e}_{|\mathfrak{E}|}}\right) \mapsto \mathbb{R}^{d_E}$$

Here, we want $\ell_{\mathfrak{E}}$ to be permutation invariant. We extend the notion of the labels even further for the entire graph by the following permutation invariant map,

$$\ell : \mathbb{R}^{d_N \times |N|} \times \mathbb{R}^{d_E \times |E|} \mapsto \mathbb{R}^D \text{ as } \left(\ell_{n_1}, \ell_{n_2}, \ldots, \ell_{|N|-1}, \ell_{|N|}, \ell_{e_1}, \ell_{e_2}, \ldots, \ell_{|E|-1}, \ell_{|E|}\right) \mapsto \mathbb{R}^D$$

1.1.2 Learning set

Let us sonsider an arbitrary set of graphs $\mathcal{G}$. We take a finite set of graphs from $\mathcal{G}$ as,

$$\{G_1, G_2, \ldots, G_i, \ldots, G_{p-1}, G_p\} \subseteq \mathcal{G} \implies p \leq |\mathcal{G}|.$$

Let us take a subset of the node set of $G_i = (N_i, E_i)$ as,

$$\{n_{i,1}, n_{i,2}, \ldots, n_{i,j}, \ldots, n_{i,q_i-1}, n_{i,q}\} \subseteq N_i \implies q \leq |N_i|.$$

We consider a desired target association $t_{i,j} \in \mathbb{R}^m$ of the node $n_{i,j}$, which is a vector in some euclidean space of dimension m.

Definition 1: Learning Set

We assume a supervised learning framework with the following set,

$$\mathcal{L} = \{(G_i, n_{i,j}, t_{i,j}) | G_i = (N_i, E_i) \in \mathcal{G}, n_{i,j} \in N_i, t_{i,j} \in \mathbb{R}^m, 1 \leq i \leq p, 1 \leq j \leq q_i\}$$

$\mathcal{L}$ is called the learning set.

Note that $\mathcal{G}$ or $\{G_1, G_2, \ldots, G_i, \ldots, G_{p-1}, G_p\}$ can also be considered as one single graph which may have multiple connected components and modify the definition of learning set accordingly.

1.1.3 Local transition and output functions

Let us take $n \in N$ from the graph $G = (N, E)$ and define the following,

- $\ell_n \in \mathbb{R}^{d_N}$: node label of $n \in N$.
- $\ell_{\mathcal{N}(n)} \in \mathbb{R}^{d_N}$: label of the neighborhood $\mathcal{N}(n) = \{\nu_1, \nu_2, \ldots, \nu_{|\mathcal{N}(n)|-1}, \nu_{|\mathcal{N}(n)|}\} \subseteq N$.
- $\mathcal{E}(n)$: edges connected to n i.e $\mathcal{E}(n) = \{(\nu_1, n), (\nu_2, n), \ldots, (\nu_{|\mathcal{N}(n)|-1}, n), (\nu_{|\mathcal{N}(n)|}, n)\}$.
- $\ell_{\mathcal{E}(n)}$: edge label of $\mathcal{E}(n)$.

Now we can define something called the *state* of n, denoted as $x_n \in \mathbb{R}^s$, which captures the the local information of n using its own node label $\ell_n \in \mathbb{R}^{d_N}$, node label of the neighbors $\ell_{\mathcal{N}(n)} \in \mathbb{R}^{d_N}$, connected edges lalel $\ell_{\mathcal{E}(n)}$, and the *state* of the neighbors $x_{\mathcal{N}(n)} \in \mathbb{R}^s$. Note that, $x_{\mathcal{N}(n)}$ is a permutation invariant map, defined by

$$x_{\mathcal{N}(n)} : \mathbb{R}^{s \times |\mathcal{N}(n)|} \mapsto \mathbb{R}^s \text{ as } \left(x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_{|\mathcal{N}(n)|-1}}, x_{\nu_{|\mathcal{N}(n)|}}\right) \mapsto \mathbb{R}^s.$$

Then we define the the *local transition function*, that expresses the dependence of a node on its neighborhood as

$$f_w : \mathbb{R}^{d_N} \times \mathbb{R}^{d_E} \times \mathbb{R}^s \times \mathbb{R}^{d_N} \mapsto \mathbb{R}^s \quad \text{as} \quad \left(\ell_n, \ell_{\mathcal{E}(n)}, x_{\mathcal{N}(n)}, \ell_{\mathcal{N}(n)}\right) \mapsto \mathbb{R}^s.$$

$$\begin{aligned}
\text{where} \quad & \left(\ell_{(\nu_1,n)}, \ell_{(\nu_2,n)}, \ldots, \ell_{(\nu_{|\mathcal{N}(n)|-1},n)}, \ell_{(\nu_{|\mathcal{N}(n)|},n)}\right) := \ell_{\mathcal{E}(n)} \\
& \left(x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_{|\mathcal{N}(n)|-1}}, x_{\nu_{|\mathcal{N}(n)|}}\right) := x_{\mathcal{N}(n)} \\
& \left(\ell_{\nu_1}, \ell_{\nu_2}, \ldots, \ell_{\nu_{|\mathcal{N}(n)|-1}}, \ell_{\nu_{|\mathcal{N}(n)|}}\right) := \ell_{\mathcal{N}(n)}
\end{aligned}$$

And we also express *the local output function* that describes how the output is produced as follows,

$$g_w : \mathbb{R}^s \times \mathbb{R}^{d_N} \mapsto \mathbb{R}^r \quad \text{as} \quad (x_n, \ell_n) \mapsto \mathbb{R}^r$$

Hence, we have the following model,

$$\begin{aligned}
\textit{local transition function}, \quad & x_n = f_w\left(\ell_n, \ell_{\mathcal{E}(n)}, x_{\mathcal{N}(n)}, \ell_{\mathcal{N}(n)}\right) \\
\textit{local output function}, \quad & o_n = g_w(x_n, \ell_n)
\end{aligned}$$

1.1.4 Model Variation

Definition 2: Minimal Model

A model is said to be minimal if it has the smallest number of variables while retaining the same computational power.

Note that,

1. One may wish to remove the labels $\ell_{\mathcal{N}(n)}$, since they include information that is implicitly contained in $x_{\mathcal{N}(n)}$. Thus one can hunt for a minimal model. It can be shown that the minimal model exists but not unique.
2. The neighborhood could contain nodes that are two or more links away from n capturing a bigger local region.

For directed or mixed graph we may find a variation of model definition by incorporating the following map into the model definition,

$$\mathfrak{d}_{\mathcal{E}(n)} : \mathcal{E}(n) \mapsto \{0, 1\} \quad \text{as}$$

$$\mathfrak{d}_{\mathcal{E}(n)}(e) = \mathfrak{d}_e = \begin{cases} 1 & \text{if } e \in \mathcal{E}(n) \text{ directs towards } n \\ 0 & \text{if } e \in \mathcal{E}(n) \text{ comes out of } n \end{cases}$$

Thus the definition of *local transition function* changes as follows,

$$f_w : \mathbb{R}^{d_N} \times \mathbb{R}^{d_E} \times \mathbb{R}^s \times \mathbb{R}^{d_N} \times \{0, 1\}^{|\mathcal{N}(n)|} \mapsto \mathbb{R}^s \quad \text{as} \quad \left(\ell_n, \ell_{\mathcal{E}(n)}, x_{\mathcal{N}(n)}, \ell_{\mathcal{N}(n)}, \mathfrak{d}_{\mathcal{E}(n)}\right) \mapsto \mathbb{R}^s.$$

where

$$\begin{aligned}
& \left(\ell_{(\nu_1,n)}, \ell_{(\nu_2,n)}, \ldots, \ell_{(\nu_{|\mathcal{N}(n)|-1},n)}, \ell_{(\nu_{|\mathcal{N}(n)|},n)}\right) := \ell_{\mathcal{E}(n)} \\
& \left(x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_{|\mathcal{N}(n)|-1}}, x_{\nu_{|\mathcal{N}(n)|}}\right) := x_{\mathcal{N}(n)} \\
& \left(\ell_{\nu_1}, \ell_{\nu_2}, \ldots, \ell_{\nu_{|\mathcal{N}(n)|-1}}, \ell_{\nu_{|\mathcal{N}(n)|}}\right) := \ell_{\mathcal{N}(n)} \\
& \left(\mathfrak{d}_{(\nu_1,n)}, \mathfrak{d}_{(\nu_2,n)}, \ldots, \mathfrak{d}_{(\nu_{|\mathcal{N}(n)|-1},n)}, \mathfrak{d}_{(\nu_{|\mathcal{N}(n)|},n)}\right) := \mathfrak{d}_{\mathcal{E}(n)}
\end{aligned}$$

However, unless explicitly stated, all the results proposed in this paper hold also for directed graphs and for graphs with mixed directed and undirected links.

Moreover, each node can have its own *local transition function* and *the local output function* definition,

$$\begin{aligned} \textit{local transition function}, \quad x_n &= f_{w_n}^{(n)}\left(\ell_n, \ell_{\mathcal{E}(n)}, x_{\mathcal{N}(n)}, \ell_{\mathcal{N}(n)}\right) \\ \textit{local output function}, \quad o_n &= g_{w_n}^{(n)}(x_n, \ell_n) \end{aligned}$$

However, for the sake of simplicity, our analysis will consider a particular model where all the nodes share the same implementation.

1.1.5 Global transition and output functions

Note that we have *local transition functions* for each node of the graph $G = (N, E)$, where $N = \{n_1, n_2, \ldots, n_{|N|-1}, n_{|N|}\}$ is the set of nodes and $E = \{e_1, e_2, \ldots, e_{|E|-1}, e_{|E|}\}$ is the set of edges.

$$\begin{aligned} x_{n_1} &= f_w\left(\ell_{n_1}, \ell_{\mathcal{E}(n_1)}, x_{\mathcal{N}(n_1)}, \ell_{\mathcal{N}(n_1)}\right) \\ x_{n_2} &= f_w\left(\ell_{n_2}, \ell_{\mathcal{E}(n_2)}, x_{\mathcal{N}(n_2)}, \ell_{\mathcal{N}(n_2)}\right) \\ &\vdots \\ x_{n_{|N|}} &= f_w\left(\ell_{n_{|N|}}, \ell_{\mathcal{E}(n_{|N|})}, x_{\mathcal{N}(n_{|N|})}, \ell_{\mathcal{N}(n_{|N|})}\right) \end{aligned}$$

We define a *global transition function*, F_w that takes the graph G as input and returns the state x_n for each $n \in N$.

Now, observe the *local output functions*,

$$\begin{aligned} o_{n_1} &= g_w(x_{n_1}, \ell_{n_1}) \\ o_{n_2} &= g_w(x_{n_2}, \ell_{n_2}) \\ &\vdots \\ o_{n_{|N|}} &= g_w(x_{n_{|N|}}, \ell_{n_{|N|}}) \end{aligned}$$

We define a *global output function*, G_w that takes the graph G as input and returns the output o_n for each $n \in N$.

At this point we have to make sure that such *global transition function* and *global output function* exists and are unique.

1.1.6 Existence and uniqueness of global transition and output functions

This section will contain the proof