Learning Label Initialization for *Time-Dependent* Harmonic Extension

Presented As Long Oral At IJCAI-22

Amitoz AZAD
amitoz.sudo@gmail.com
Domain: Semi-supervised, GraphML, PDE

Background. Node classification on graphs can be formulated as the following *time-dependent* Dirichlet problem:

$$\frac{\partial f(u,t)}{\partial t} = \Delta_w(f)(u,t) \qquad \forall u \in V \setminus V_0 \tag{1a}$$

$$f(u,t) = g(u) \qquad \forall u \in V_0 \tag{1b}$$

$$f(u,0) = \psi_0(u) \qquad \forall u \in V. \tag{1c}$$

Here, Δ_w is the normalized weighted graph Laplacian [3], u is a graph node, f is the signal, V is set of all vertices, V_0 is the set of all vertices for which the label is given, $g : V_0 \rightarrow R^k$ is the label function on the nodes, such that $g(u) \in \{\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_k}\}$, where $\mathbf{e_i} \in R^k$ is the i^{th} class one-hot vector. A common choice of setting the initialization vector $(\psi_0(u))$ is to keep it as the one-hot vector on the labeled nodes and the zero vector on the unlabeled nodes. The above PDE can also be seen as heat diffusion with in-homogeneous boundary condition on graphs.

Dataset	**Eq1**	**GCN**
MNIST	93.2 ± 0.0	91.3 ± 0.2
FMNIST	76.0 ± 0.0	77.4 ± 0.2
Cora	72.5 ± 0.0	81.5 ± —†
Citeseer	49.7 ± 0.0	70.0 ± —†
Pubmed	72.5 ± 0.0	79.0 ± —†

Table 1: Comparison over handcrafted *knn* (MNIST, FMNIST) graphs and the citation networks. †Values taken from the GCN paper.

Motivation. When comparing the results of (1) and the GCN [1] on the handcrafted (knn) graphs and the citation networks [2], one observes significant differences in the performances over the citation networks (Table 1). The plausible reason for the similar performances on handcrafted (knn) graphs would be that edges connect the similar nodes, as euclidean distances between the images happens to be a good metric for the graph construction. The poor performance of the PDE (1) on the citation graphs could be attributed to the fact edges don't explicitly translate to similarity between the nodes on these graphs. In fact, they corresponds to the citation link between the documents. A GCN on the other hand is able to perform well on the citation networks as it uses the graph structure and the node features (bag-of-words vector for citation graphs) where as the PDE only uses the graph structure.

Research Question. Knowing that the GCN beats the solution of (1) by utilizing the node features and the graph structure, could we use the node features and make the solution of the PDE at par with the GCN on the citation graphs?

Main Contributions.

- We show that the solution of the PDE (1) is improved significantly and made at par with several state-of-the-art methods for node classification by utilizing the node features to learn the initial condition (1c) (Figure 1).
- We show that learning the graph weights along with (1c) adds further marginal gains to the solution.
- Unlike GNNs, which have to be retrained every time to incorporate new labels, we show that our approach can easily include the new labels.

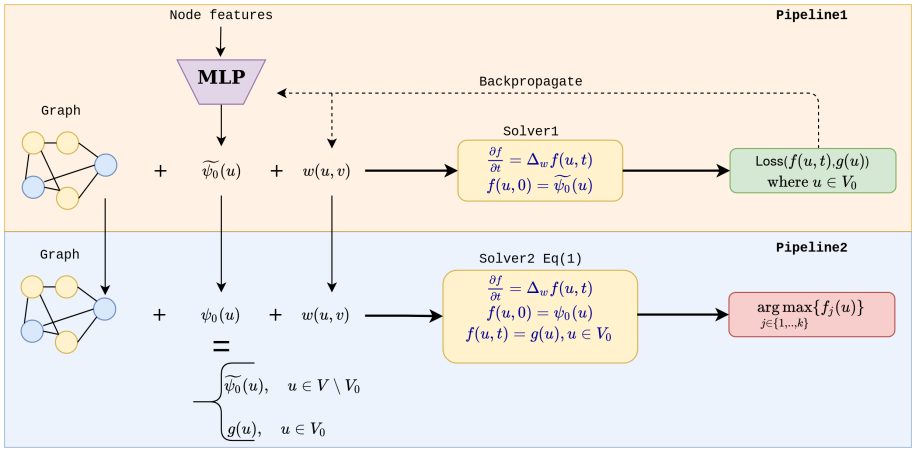

Figure 1: Proposed architecture to learn the initial condition (1c).

References

[1] Kipf, T. N., and Welling, M. Semi-supervised classification with graph convolutional networks. In *ICLR* (2017).

[2] Sen, P., Namata, G., Bilgic, M., Getoor, L., Galligher, B., and Eliassi-Rad, T. Collective classification in network data. *AI magazine* (2008).

[3] Zhou, D., and Schölkopf, B. Regularization on discrete spaces. In *Joint Pattern Recognition Symposium* (2005).