TRANSDUCTIVE KERNELS FOR GAUSSIAN PROCESSES ON GRAPHS

Yin-Cong Zhi[⋆] *Felix L. Opolka*[†] *Yin Cheng Ng*[‡] *Pietro Liò*[†] *Xiaowen Dong*[⋆]

[⋆]University of Oxford [†]University of Cambridge [‡]Man AHL

Gaussian processes (GP) are flexible tools for Bayesian modelling, and recent studies into building GPs on graphs have proved to be competitive against state-of-the-art graph neural network (GNN) models, giving rise to various kernel-based approaches for semi-supervised classification in [1, 2]. In this work, we focus on semi-supervised problems where the set of nodes are only partially labelled, and we aim to predict the unlabelled nodes. Graphs data of this form often comes with node attributes, which we will use as feature data, and in designing a kernel for a graph GP, the challenge is in embedding the graph connectivity into the kernel along with the feature data.

The most representative approaches to building GP on graphs for semi-supervised classification are [1, 2], where the authors made use of a matrix transformation on a graph-less base kernel on the node feature data. This could be a limiting factor for the model as the graph information comes into the model through a linear transformation. Furthermore, in this work we found if the base kernel is not particularly suitable, adding the graph elements will not be effective, and these models do not have a way to adjust the influence of the graph against the feature data kernel.

To address these issues, we use the principles of kernel design through regularization to derive a kernel with the ability to naturally handle the graph and feature data. In this approach, kernels are obtained by choosing regularization functions and finding a reproducing kernel Hilbert space (RKHS) [3], typical continuous kernels such as RBF and Matérn kernels can all be derived in this manner.

The regularization approach has been translated to the graph domain in [4], but currently this kernel depends on the graph only, and cannot incorporate node data. However, this method can be extended to graphs with node feature data, and in this paper we show how this can be achieved. We start by introducing the regularization approach to derive kernels for feature data and for graphs separately. We then present our approach by combining the two regularizers to obtain a kernel for graphs with node data, leading to our proposed model. The resulting model has transductive properties, meaning it trains on all nodes data in the graph, but only the training node labels. In addition, our kernel provides a clear way to control the influence of the graph compared to the feature data kernel. We then show that our setup is general, and that many graph learning models are actually instances of our design. We demonstrate the advantages of our model on synthetic data that are highly non-Euclidean and the training set is small. Lastly, we test on various real world graph-data in semi-supervised classification, comparing against various graph GP and popular GNN models, the classification accuracies on the benchmark datasets are presented in Table 1. For technical details about our kernel derivation and design we refer readers to our paper in [5].

Method	**Texas**	**Cornell**	**Wisconsin**	**Chameleon**
# Nodes	183	183	251	2,277
Hom. Ratio	0.11	0.30	0.21	0.23
GCN	59.5 ±5.3	57.0 ±4.7	59.8 ±7.0	59.8 ±2.6
GAT	58.4 ±4.5	58.9 ±3.3	55.3 ±8.7	54.7 ±2.0
ChebNet	77.3 ±4.1	**74.3 ±7.5**	79.4 ±4.5	55.2 ±2.8
GP	78.4	73.0	78.4	46.1
GGP	78.4	62.1	60.8	**73.5**
ChebGP	**81.1**	64.9	**82.4**	**69.1**
WGGP	78.4	67.6	**84.3**	64.5
TGGP (ours)	**81.1**	**75.7**	**82.4**	63.2
GGP-X	78.4	56.8	60.8	**77.6**
WGGP-X	81.1	75.7	84.3	65.6
TGGP-X (ours)	**86.5**	**81.1**	**86.3**	63.4

Method	**Cora**	**Citeseer**	**Squirrel**	**Actor**
# Nodes	2,708	3,327	5,201	7,600
Hom. Ratio	0.81	0.74	0.22	0.22
GCN	80.5 ±0.8	68.1 ±1.3	36.9 ±1.3	30.3 ±0.8
GAT	**82.6 ±0.7**	**72.2 ±0.9**	30.6 ±2.1	26.3 ±1.7
ChebNet	78.0 ±1.2	70.1 ±0.8	43.9 ±1.6	34.1 ±1.1
GP	60.8	54.7	34.4	**34.9**
GGP	80.9	69.7	**64.8**	26.3
ChebGP	79.7	66.5	28.8	31.8
WGGP	**84.7**	**70.8**	**58.3**	32.6
TGGP (ours)	80.3	70.5	53.8	**34.9**
GGP-X	84.7	75.6	**71.9**	OOM
WGGP-X	**87.5**	**76.8**	61.3	OOM
TGGP-X (ours)	83.8	76.7	54.2	**36.9**

Table 1: Classification percentage accuracy on real world datasets. Entries with OOM are due to "Out Of Memory" error and thus we were unable to run the experiment on the dataset. The two best performing models and the highest **-X** version are highlighted.

1. REFERENCES

[1] Yin Cheng Ng, Nicolò Colombo, and Ricardo Silva, "Bayesian semi-supervised learning with graph gaussian processes," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[2] Felix Opolka, Yin-Cong Zhi, Pietro Liò, and Xiaowen Dong, "Adaptive gaussian processes on graphs via spectral graph wavelets," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 4818–4834.

[3] Alex J Smola, Bernhard Schölkopf, and Klaus-Robert Müller, "The connection between regularization operators and support vector kernels," *Neural networks*, vol. 11, no. 4, pp. 637–649, 1998.

[4] Alexander J Smola and Risi Kondor, "Kernels and regularization on graphs," in *Learning theory and kernel machines*, pp. 144–158. Springer, 2003.

[5] Yin-Cong Zhi, Felix L Opolka, Yin Cheng Ng, Pietro Liò, and Xiaowen Dong, "Transductive kernels for gaussian processes on graphs," *arXiv preprint arXiv:2211.15322*, 2022.