

Recent Advances in Kernel-Based Graph Classification

Nils M. Kriege^(✉) and Christopher Morris

Department of Computer Science, TU Dortmund University, Dortmund, Germany
{nils.kriege, christopher.morris}@tu-dortmund.de

Abstract. We review our recent progress in the development of graph kernels. We discuss the hash graph kernel framework, which makes the computation of kernels for graphs with vertices and edges annotated with real-valued information feasible for large data sets. Moreover, we summarize our general investigation of the benefits of explicit graph feature maps in comparison to using the kernel trick. Our experimental studies on real-world data sets suggest that explicit feature maps often provide sufficient classification accuracy while being computed more efficiently. Finally, we describe how to construct valid kernels from optimal assignments to obtain new expressive graph kernels. These make use of the kernel trick to establish one-to-one correspondences. We conclude by a discussion of our results and their implication for the future development of graph kernels.

1 Introduction

In various domains such as chemo- and bioinformatics, or social network analysis large amounts of graph structured data is becoming increasingly prevalent. Classification of these graphs remains a challenge as most graph kernels either do not scale to large data sets or are not applicable to all types of graphs. In the following we briefly summarize related work before discussing our recent progress in the development of efficient and expressive graphs kernels.

1.1 Related Work

In recent years, various graph kernels have been proposed. Gärtner et al. [5] and Kashima et al. [8] simultaneously developed graph kernels based on random walks, which count the number of walks two graphs have in common. Since then, random walk kernels have been studied intensively, see, e.g., [7, 10, 13, 19, 21]. Kernels based on shortest paths were introduced by Borgwardt et al. [1] and are computed by performing 1-step walks on the transformed input graphs, where edges are annotated with shortest-path lengths. A drawback of the approaches mentioned above is their high computational cost. Therefore, a different line of research focuses particularly on scalable graph kernels. These kernels are typically computed by explicit feature maps, see, e.g., [17, 18]. This allows to bypass

the computation of a gram matrix of quadratic size by applying fast linear classifiers [2]. Moreover, graph kernels using assignments have been proposed [4], and were recently applied to geometric embeddings of graphs [6].

2 Recent Progress in the Design of Graph Kernels

We give an overview of our recent progress in the development of scalable and expressive graph kernels.

2.1 Hash Graph Kernels

In areas such as chemo- or bioinformatics edges and vertices of graphs are often annotated with real-valued information, e.g., physical measurements. It has been shown that these attributes can boost classification accuracies [1, 3, 9]. Previous graph kernels that can take these attributes into account are relatively slow and employ the kernel trick [1, 3, 9, 15]. Therefore, these approaches do not scale to large graphs and data sets. In order to overcome this, we introduced the *hash graph kernel framework* in [14]. The idea is to iteratively turn the continuous attributes of a graph into discrete labels using randomized hash functions. This allows to apply fast explicit graph feature maps, e.g., [17], which are limited to discrete annotations. In each iteration we sample new hash functions and compute the feature map. Finally, the feature maps of all iterations are combined into one feature map. In order to obtain a meaningful similarity between attributes in \mathbb{R}^d , we require that the probability of collision $\Pr[h_1(x) = h_2(y)]$ of two independently chosen random hash functions $h_1, h_2: \mathbb{R}^d \rightarrow \mathbb{N}$ equals an adequate kernel on \mathbb{R}^d . Equipped with such a hash function, we derived approximation results for several state-of-the-art kernels which can handle continuous information. Moreover, we derived a variant of the Weisfeiler-Lehman subtree kernel which can handle continuous attributes.

Our extensive experimental study showed that instances of the hash graph kernel framework achieve state-of-the-art classification accuracies while being orders of magnitudes faster than kernels that were specifically designed to handle continuous information.

2.2 Explicit Graph Feature Maps

Explicit feature maps of kernels for continuous vectorial data are known for many popular kernels like the Gaussian kernel [16] and are heavily applied in practice. These techniques cannot be used to obtain approximation guarantees in the hash graph kernel framework. Therefore, in a different line of work, we developed explicit feature maps with the goal to lift the known approximation results for kernels on continuous data to kernels for graphs annotated with continuous data [11]. More specifically, we investigated how general convolution kernels are composed from base kernels and how to construct corresponding feature maps. We applied our results to widely used graph kernels and analyzed

for which kernels and graph properties computation by explicit feature maps is feasible and actually more efficient. We derived approximative, explicit feature maps for state-of-the-art kernels supporting real-valued attributes. Empirically we observed that for graph kernels like GraphHopper [3] and Graph Invariant [15] approximative explicit feature maps achieve a classification accuracy close to the exact methods based on the kernel trick, but required only a fraction of their running time. For the shortest-path kernel [1] on the other hand the approach fails in accordance to our theoretical analysis.

Moreover, we investigated the benefits of employing the kernel trick when the number of features used by a kernel is very large [10, 11]. We derived feature maps for random walk and subgraph kernels, and applied them to real-world graphs with discrete labels. Experimentally we observed a phase transition when comparing running time with respect to label diversity, walk lengths and subgraph size, respectively, confirming our theoretical analysis.

2.3 Optimal Assignment Kernels

For non-vectorial data, Fröhlich et al. [4] proposed kernels for graphs derived from an optimal assignment between their vertices, where vertex attributes are compared by a base kernel. However, it was shown that the resulting similarity measure is not necessarily a valid kernel [20, 21]. Hence, in [12], we studied optimal assignment kernels in more detail and investigated which base kernels lead to valid kernels. We characterized a specific class of kernels and showed that it is equivalent to the kernels obtained from a hierarchical partition of their domain. When such kernels are used as base kernel the optimal assignment (i) yields a valid kernel; and (ii) can be computed in linear time by histogram intersection given the hierarchy. We demonstrated the versatility of our results by deriving novel graph kernels based on optimal assignments, which are shown to improve over their convolution-based counterparts. In particular, we proposed the Weisfeiler-Lehman optimal assignment kernel, which performs favorable compared to state-of-the-art graph kernels on a wide range of data sets.

3 Conclusion

We gave an overview about our recent progress in kernel-based graph classification. Our results show that explicit graph feature maps can provide an efficient computational alternative for many known graph kernels and practical applications. This is the case for kernels supporting graphs with continuous attributes and for those limited to discrete labels, even when the number of features is very large. Assignment kernels, on the other hand, are computed by histogram intersection and thereby again employ the kernel trick. This suggests to study the application of non-linear kernels to explicit graph feature maps in more detail as future work.

Acknowledgement. We would like to thank the co-authors of our publications [10–12, 14]. This research was supported by the German Science Foundation (DFG) within the Collaborative Research Center SFB 876 “Providing Information by Resource-Constrained Data Analysis”, project A6.

References

1. Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. In: IEEE International Conference on Data Mining, pp. 74–81 (2005)
2. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: a library for large linear classification. *J. Mach. Learn. Res.* **9**, 1871–1874 (2008)
3. Feragen, A., Kasenburg, N., Petersen, J., Buijne, M.D., Borgwardt, K.: Scalable kernels for graphs with continuous attributes. In: *Advances in Neural Information Processing Systems*, pp. 216–224 (2013)
4. Fröhlich, H., Wegner, J.K., Sieker, F., Zell, A.: Optimal assignment kernels for attributed molecular graphs. In: *22nd International Conference on Machine Learning*, pp. 225–232 (2005)
5. Gärtner, T., Flach, P., Wrobel, S.: On graph kernels: hardness results and efficient alternatives. In: Schölkopf, B., Warmuth, M.K. (eds.) *COLT-Kernel 2003*. LNCS (LNAI), vol. 2777, pp. 129–143. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45167-9_11
6. Johansson, F.D., Dubhashi, D.: Learning with similarity functions on graphs using matchings of geometric embeddings. In: *21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 467–476 (2015)
7. Kang, U., Tong, H., Sun, J.: Fast random walk graph kernel. In: *SIAM International Conference on Data Mining*, pp. 828–838 (2012)
8. Kashima, H., Tsuda, K., Inokuchi, A.: Marginalized kernels between labeled graphs. In: *20th International Conference on Machine Learning*, pp. 321–328 (2003)
9. Kriege, N., Mutzel, P.: Subgraph matching kernels for attributed graphs. In: *29th International Conference on Machine Learning* (2012)
10. Kriege, N., Neumann, M., Kersting, K., Mutzel, M.: Explicit versus implicit graph feature maps: a computational phase transition for walk kernels. In: *IEEE International Conference on Data Mining*, pp. 881–886 (2014)
11. Kriege, N.M., Neumann, M., Morris, C., Kersting, K., Mutzel, P.: A unifying view of explicit and implicit feature maps for structured data: Systematic studies of graph kernels. *CoRR* abs/1703.00676 (2017). <http://arxiv.org/abs/1703.00676>
12. Kriege, N.M., Giscard, P.-L., Wilson, R.C.: On valid optimal assignment kernels and applications to graph classification. In: *Advances in Neural Information Processing Systems*, pp. 1615–1623 (2016)
13. Mahé, P., Ueda, N., Akutsu, T., Perret, J.L., Vert, J.P.: Extensions of marginalized graph kernels. In: *Twenty-First International Conference on Machine Learning*, pp. 552–559 (2004)
14. Morris, C., Kriege, N.M., Kersting, K., Mutzel, P.: Faster kernel for graphs with continuous attributes via hashing. In: *IEEE International Conference on Data Mining*, pp. 1095–1100 (2016)
15. Orsini, F., Frasconi, P., De Raedt, L.: Graph invariant kernels. In: *Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 3756–3762 (2015)
16. Rahimi, A., Recht, B.: Random features for large-scale kernel machines. In: *Advances in Neural Information Processing Systems*, pp. 1177–1184 (2008)

17. Shervashidze, N., Schweitzer, P., van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-Lehman graph kernels. *J. Mach. Learn. Res.* **12**, 2539–2561 (2011)
18. Shervashidze, N., Vishwanathan, S.V.N., Petri, T.H., Mehlhorn, K., Borgwardt, K.M.: Efficient graphlet kernels for large graph comparison. In: Twelfth International Conference on Artificial Intelligence and Statistics, pp. 488–495 (2009)
19. Sugiyama, M., Borgwardt, K.M.: Halting in random walk kernels. In: Advances in Neural Information Processing Systems, pp. 1639–1647 (2015)
20. Vert, J.P.: The optimal assignment kernel is not positive definite. CoRR abs/0801.4061 (2008). <http://arxiv.org/abs/0801.4061>
21. Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R., Borgwardt, K.M.: Graph kernels. *J. Mach. Learn. Res.* **11**, 1201–1242 (2010)