Publications of Marcus Pappik

This document lists all peer-reviewed publications of Marcus Pappik, Chair for Algorithm Engineering, Hasso Plattner Institute, Potsdam, Germany. This listing was automatically generated on September 24, 2021. An up-to-date version is available online at hpi.de/friedrich/docs/publist/pappik.pdf.

Conference papers

[1] Friedrich, T., Göbel, A., Krejca, M., Pappik, M., A spectral independence view on hard spheres via block dynamics. In: *International Colloquium on Automata, Languages and Programming (ICALP)*, pp. 66:1–66:15, 2021.

The hard-sphere model is one of the most extensively studied models in statistical physics. It describes the continuous distribution of spherical particles, governed by hard-core interactions. An important quantity of this model is the normalizing factor of this distribution, called the partition function. We propose a Markov chain Monte Carlo algorithm for approximating the grand-canonical partition function of the hard-sphere model in d dimensions. Up to a fugacity of $\lambda < \mathrm{e}/2^d$, the runtime of our algorithm is polynomial in the volume of the system. This covers the entire known real-valued regime for the uniqueness of the Gibbs measure. Key to our approach is to define a discretization that closely approximates the partition function of the continuous model. This results in a discrete hard-core instance that is exponential in the size of the initial hard-sphere model. Our approximation bound follows directly from the correlation decay threshold of an infinite regular tree with degree equal to the maximum degree of our discretization. To cope with the exponential blow-up of the discrete instance we use clique dynamics, a Markov chain that was recently introduced in the setting of abstract polymer models. We prove rapid mixing of clique dynamics up to the tree threshold of the univariate hard-core model. This is achieved by relating clique dynamics to block dynamics and adapting the spectral expansion method, which was recently used to bound the mixing time of Glauber dynamics within the same parameter regime.

[2] Echzell, H., Friedrich, T., Lenzner, P., Molitor, L., Pappik, M., Schöne, F., Sommer, F., Stangl, D., Convergence and Hardness of Strategic Schelling Segregation. In: *Web and Internet Economics (WINE)*, pp. 156–170, 2019.

The phenomenon of residential segregation was captured by Schelling's famous segregation model where two types of agents are placed on a grid and an agent is content with her location if the fraction of her neighbors which have the same type as her is at least τ, for some $0 < \tau < 1$. Discontent agents simply swap their location with a randomly chosen other discontent agent or jump to a random empty cell. We analyze a generalized game-theoretic model of Schelling segregation which allows more than two agent types and more general underlying graphs modeling the residential area. For this we show that both aspects heavily influence the dynamic properties and the tractability of finding an optimal placement. We map the boundary of when improving response dynamics (IRD), i.e., the natural approach for finding equilibrium states, are guaranteed to converge. For this we prove several sharp threshold results where guaranteed IRD convergence suddenly turns into the strongest possible non-convergence result: a violation of weak acyclicity. In particular, we show such threshold results also for Schelling's original model, which is in contrast to the standard assumption in many empirical papers. Furthermore, we show that in case of convergence, IRD find an equilibrium in $O(m)$ steps, where m is the number of edges in the underlying graph and show that this bound is met in empirical simulations starting from random initial agent placements.

[3] Kumar Shekar, A., Pappik, M., Iglesias Sánchez, P., Müller, E., Selection of Relevant and Non-Redundant Multivariate Ordinal Patterns for Time Series Classification. In: *Discovery Science (DS)*, pp. 224–240, 2018.

Transformation of multivariate time series into feature spaces are common for data mining tasks like classification. Ordinality is one important property in time series that provides a qualitative representation of the underlying dynamic regime. In a multivariate time series, ordinalities from multiple dimensions combine together to be discriminative for the classification problem. However, existing works on ordinality do not address the multivariate nature of the time series. For multivariate ordinal patterns, there is a computational challenge with an explosion of pattern combinations, while not all patterns are relevant and provide novel information for the classification. In this work, we propose a technique for the extraction and selection of relevant and non-redundant multivariate ordinal patterns from the high-dimensional combinatorial search space. Our proposed approach Ordinal feature extraction (ordex), simultaneously extracts and scores the relevance and redundancy of ordinal patterns without training a classifier. As a filter-based approach, ordex aims to select a set of relevant patterns with complementary information. Hence, using our scoring function based on the principles of Chebyshev's inequality, we maximize the relevance of the patterns and minimize the correlation between them. Our experiments on real world datasets show that ordinality in time series contains valuable information for classification in several applications.

[4] Kirsch, L., Riekenbrauck, N., Thevessen, D., Pappik, M., Stebner, A., Kunze, J., Meissner, A., Kumar Shekar, A., Müller, E., Framework for Exploring and Understanding Multivariate Correlations. In: *Machine Learning and Knowledge Discovery in Databases (ECML/PKDD)*, pp. 404–408, 2017.

Feature selection is an essential step to identify relevant and non-redundant features for target class prediction. In this context, the number of feature combinations grows exponentially with the dimension of the feature space. This hinders the user's understanding of the feature-target relevance and feature-feature redundancy. We propose an interactive Framework for Exploring and Understanding Multivariate Correlations (FEXUM), which embeds these correlations using a force-directed graph. In contrast to existing work, our framework allows the user to explore the correlated feature space and guides in understanding multivariate correlations through interactive visualizations.