Maja Rudolph | Research Statement

My research addresses machine-learning questions derived from real-world engineering problems: for example, how to model driving behavior, how to forecast the operating conditions of a device, or how to find anomalies in the sensor data of an assembly line.

Engineering devices typically rely on sensors to measure necessary inputs. Machine-learning methods trained on sensor data can increase the utility of such devices by improving their efficiency or enabling new features. However, modeling sensor data presents the following challenges: real-world data often exhibits *multimodal dynamics*; different sensors are usually *asynchronous*; and most industrial datasets are *unlabeled*. Learning from sensor data requires methods that are flexible enough to capture the underlying complexity (deep learning) while handling various types of noise and uncertainty (probabilistic modeling).

For this reason, the main technical theme in my research is how to best combine deep learning with probabilistic modeling. I describe how it enables multi-modal forecasting and modeling asynchronous data. Then, I present the theoretical and practical advantages of *neural transformation learning*, a tool we developed for learning from unlabeled data.

Forecasting multimodal dynamics. Many real-world time-series data sets are highly multimodal: at any given part of an observed sequence, there are multiple plausible continuations (e.g. a car could turn right or left at an intersection) but the average of these behaviors is unlikely or even physically impossible. Most deep probabilistic sequence models struggle with multimodality since their modeling assumptions and training objectives encourage averaging. To address this, we have developed a new approximate inference method for sequential latent variables that is multimodal by construction **[Preprint 2020]**. The key idea is to marginalize over past states instead of conditioning on them, as in previous work. Our approach improves forecasting in a number of real-world data sets, including taxi trajectories, movement patterns of basketball players, and pollution levels over time.

Modeling asynchronous sensors. Another common issue in dynamics modeling for industrial applications is that the sensors are usually not synchronized. The sampling rates are irregular and often the various sensors are only partially observed. A common approach is to interpolate the inputs before applying a standard deep learning method, such as a recurrent neural network (RNN). The *continuous-discrete recurrent Kalman network* (CDRKN) **[ML4ITS 2021]** does not require this heuristic pre-processing step. Like an RNN, it maintains a latent state which is recursively updated based on observations. After each observation is encoded by a neural network, the CDRKN uses the continuous formulation of a Kalman filter to optimally integrate the new information (including exactly how much time has passed between observations, which is also informative) into its latent state. The result is a practical neural architecture that can naturally handle asynchronous data.

Unlabeled industrial data. Similar modeling challenges (and opportunities) can be encountered in manufacturing. Production facilities generate a huge amount of sensor data and machine learning can help find parts that seem to be falling short of quality standards. While there has been extensive work in anomaly detection on images, many industrial applications require us to detect abnormalities in other types of data, such as time series (e.g. abnormal process curves of a drill), graphs (e.g. for detecting abnormal molecules in a chemical process), or events within a time series (e.g. for intrusion detection in a power plant). A major challenge for anomaly detection is that the data is not usually labeled. *Self-supervised learning* uses auxiliary tasks instead of labels as the prediction targets.

Self-supervised anomaly detection beyond images. Many self-supervised methods, especially in computer vision, rely on image transformations (such as rotations, cropping, and blurring) to construct auxiliary prediction tasks. Models trained on these tasks can then be used for anomaly detection or feature extraction. Their applicability to industrial data has been limited, since it is difficult to manually design appropriate data transformations for specialized domains. Our work overcomes this challenge by treating the transformations like model parameters and *learning* appropriate transformations from the data **[ICML 2021]**. We call this *neural transformation learning*. The main contribution is a novel training objective that encourages the neural transformations to generate alternative views of each data point that are diverse, yet characteristic of the original input. We prove that the objective avoids the trivial solutions (identity or constant) that one would encounter with existing approaches. Neural transformations achieve state-of-the-art results for self-supervised anomaly detection of entire time series and tabular data **[ICML 2021]**, for detecting anomalies *within* time series **[SSL 2021]**, and for anomaly detection on graphs **[Preprint 2021]**.

Theoretical advantages of neural transformations. The neural transformation learning objective we have developed has another theoretical advantage: it can be used as a regularizer for deep one-class classification (OCC). In deep OCC, a neural network maps each data point into a hyper-sphere in an embedding space. The distance to the center of the sphere represents how "normal" a sample is and serves as the anomaly score. While deep OCC works well on images, it suffers from *mode collapse*: the optimal solution to the OCC objective is a constant feature extractor that maps everything to the center of the hypersphere. This trivial solution can not be used for anomaly detection, but can be mitigated by our approach. We prove in **[Preprint 2021]** that combining OCC with transformation learning completely resolves the mode collapse issue. The theoretical findings are supported by our empirical results where our combined method, *one-class graph transformation learning*, achieves the highest accuracy in detecting abnormal graphs.

Practical advantages of neural transformations. In addition to the theoretical and empirical advantages of neural transformations over other deep-learning approaches for self-supervised anomaly detection, it also has an important practical advantage: since the data transformations are not crafted by hand but are instead learned by neural networks, we can apply data augmentations to data types that we have less intuition about than natural images. In fact, we can even apply the neural transformations in an abstract feature space **[NeurIPS 2017, SSL 2021]**, sometimes called the *embedding space*. The embedding vector of each data point captures characteristics of the data that are useful for neural networks but not necessarily accessible to human interpretation. Their abstract nature makes it hard to design transformations by hand but lends itself to neural transformations.

Representation learning for industrial applications. Embeddings are a powerful tool for working with high-dimensional data. Word embeddings, for example, capture semantic similarity among terms in a vocabulary and are useful input features for downstream tasks **[PNAS 2020]**. To make them applicable to other types of data, we have developed *exponential family embeddings* **[NeurIPS 2016]**, which can be combined with a dynamic prior to study how similarities change over time **[WWW 2018]**. There is a great potential for employing embeddings in data-driven manufacturing. Dense representations that capture the characteristics of parts, stations, and/or production steps could simplify tasks such as root cause analysis and reduce the amount of required training data for quality prediction. In future work, I am particularly excited to study how neural transformations **[ICML 2021]** can improve self-supervised representation learning, both in theory and in practice.

My future research will continue to explore novel ways of combining deep learning and probabilistic methods for forecasting, anomaly detection, and representation learning and I will continue to focus on research questions that are derived from industrial applications.

Selected Bibliography

[ICML 2021]

- C. Qiu, T. Pfrommer, M. Kloft, S. Mandt, and **M. Rudolph**. Neural transformation learning for deep anomaly detection beyond images. In *International Conference on Machine Learning*, 2021.

[Preprint 2021]

- C. Qiu, M. Kloft, S. Mandt, and **M. Rudolph**. Raising the bar in graph-level anomaly detection. *under review*.

[ML4ITS 2021]

- M. Schirmer, M. Eltayeb, and **M. Rudolph**. Countinuous-discrete recurrent Kalman networks for irregular time-series. In *ML4ITS Workshop*, 2021.

[SSL 2021]

- T. Schneider, C. Qiu, M. Kloft, D. Latif, S. Staab, S. Mandt, and **M. Rudolph**. Detecting anomalies in time series with local neural transformations. In *ICML Workshop: Self-Supervised Learning for Reasoning and Perception*, 2021.

[PNAS 2020]

- J. L. McClelland, F. Hill, **M. Rudolph**, J. Baldridge, and H. Schütze. Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models. In *Proceedings of the National Academy of Sciences*, 2020.

[Preprint 2020]

- C. Qiu, S. Mandt, and **M. Rudolph**. History marginalization improves forecasting in variational RNNs. *under review*.

[WWW 2018]

- **M. Rudolph** and D. Blei. Dynamic embeddings for language evolution. In *World Wide Web Conference*, 2018.

[NeurIPS 2017]

- **M. Rudolph**, F. Ruiz, S. Athey, and D. Blei. Structured embedding models for grouped data. In *Neural Information Processing Systems*, 2017.

[NeurIPS 2016]

- **M. Rudolph**, F. Ruiz, S. Mandt, and D. Blei. Exponential family embeddings. In *Neural Information Processing Systems*, 2016.