Karlsruher Institut für Technologie

Institut für Programmstrukturen und Datenorganisation (IPD)
Lehrstuhl für Systeme der Informationsverwaltung, Prof. Böhm

Master Thesis

Deep Materials: Understanding and Explaining Material Defects with Machine Learning

Machine learning supports scientists to discover and to understand relationships in real systems. In a current research thread, we – a team of data scientists and materials scientists – strive for insights on material behavior under physical stress situations. Our goal is to better understand how materials defects at the atomic level behave, and how they influence the properties of materials.

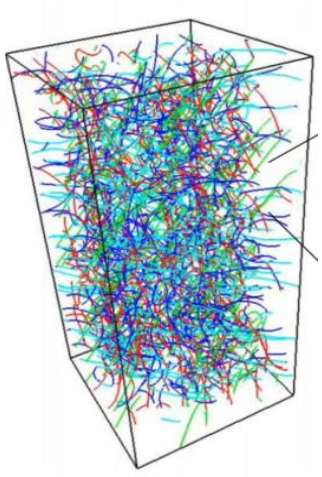

Figure 1 Dislocation Graphs.

Most materials have defects in their crystalline structure. To study these defects, materials scientists rely on computer simulations in which defects move through the crystalline structure, trigger new defects or interact with each other. Although a proper modelling of such complex dynamics is crucial to make the simulation realistic, the physics behind them is not yet well understood. The objective of our research is to help materials scientists in understanding and in explaining dynamics of crystalline defects with machine learning.

From a computer science perspective, this is an interesting task for several reasons. First, the simulation data is in a non-standard format. The defects in materials are represented as spatio-temporal graphs of so-called dislocations, see Figure 1. In other words, the simulation output are graphs where the nodes and edges move through the material, and where sub-graphs merge or split over time. This is challenging, because one must transform the input, e.g., to be static and in tabular form, so that it fits the machine learning algorithm.

Second, materials scientists are mainly interested in dislocations on a macro scale. This is, they are interested where large groups of dislocations move, e.g., towards the border of the material, rather than in the movement of single dislocations. To this end, one can think of many ways to aggregate the simulation data to a macro scale, for instance by slicing the material into a grid of small cubes. A hypothesis is that the size of these cubes plays an important role in capturing the dislocation dynamics. However, it is yet unclear how to select the grid size and an aggregation method.

All this results in the following tasks for this thesis:

- Formulation of materials science questions as a machine learning problem
- Identification, implementation and comparison of ways to preprocess spatio-temporal graphs (e.g., by learning Graph Representations [1]) so that one can apply machine learning
- Development and comparison of methods to aggregate dislocations on a macro scale
- Experimental evaluation on existing simulation data and interpretation of the results

In this thesis, you will be working on latest research questions and acquire practical knowledge on large-scale data analytics. You train highly demanded skills in machine learning and the design of data-driven applications. Elementary statistical knowledge, programming skills and the ability to accomplish conceptual work are desired.

[1] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation Learning on Graphs: Methods and Applications," arXiv [cs.SI], 17-Sep-2017.

Contact: Holger Trittenbach holger.trittenbach@kit.edu Building: 50.34 Room 338