Geophysical Research Abstracts
Vol. 13, EGU2011-5932, 2011
EGU General Assembly 2011
© Author(s) 2011

Learning Ground Motion Models from Data using Bayesian Networks

Kristin Vogel, Carsten Riggelsen, and Frank Scherbaum
University of Potsdam, Institute of Earth and Environmental Science, Germany (kvog@geo.uni-potsdam.de)

In probabilistic seismic hazard analysis (PSHA) it is essential to work with good ground motion models describing the (in)dependencies between ground motion, earthquake, path and site-related parameters with a minimal amount of epistemic uncertainty. To understand which predictor variables have the strongest influences, how they affect the ground motion parameter and each other, we take a machine learning approach: Based on the work of Kuehn et al. (BSSA 2011) we aim to learn from data the (in)dependency structure of ground motion parameter such as peak ground acceleration, peak ground velocity or spectral acceleration conditioned on predictor variables (e.g. magnitude, distance, local shear wave velocity, fault mechanism). To characterize the uncertainty on these parameters, they are usually treated as random variables.

Bayesian Networks (BN's) are a powerful probabilistic framework allowing to detect and describe conditional (in)dependencies between random variables. They combine methods from Graph Theory and Probability Theory. The nodes of a graph represent random variables and parameters, while the edges between nodes symbolize conditional (in)dependencies. The visualization of the graph allows experts and non-experts to quickly get an intuition about the dependency structure of all parameters at a glance. BN's do not only model the conditional probability of the target variable (i.e. the ground-motion parameter), but the joint probability of all random variables. There is no distinction between target and predictor variables. Using the joint probability we can compute all conditional probabilities of interest, in particular the conditional distribution of ground motion given earthquake, path and site-related parameters. However, for fast computation (inference) within a BN, the network has to fulfill certain criteria. Difficulties arise when we work with domains with both continuous and discrete variables, as we do in PSHA. How to counteract these problems is part of our work. One option is to assume distributions for the continuous variables, which allow fast and exact inference, like mixtures of Gaussians or mixtures of truncated exponentials. For the ground-motion parameter it is for instance reasonable to assume a log-normal distribution.

Another way is to discretize the continuous variables avoiding distributional assumptions a priori. In BN learning the way of discretizing can have significant impact on the network structure. We want to show the sensitivity of learned network structures on the used discretization of datasets and propose a multivariate Bayesian discretization scheme, where the discretization depends on the network structure, the recorded data and the chosen prior. We combine the multivariate Bayesian discretization method with a Hill-Climber Monte Carlo algorithm to learn a network structure. For each learned network structure a new optimal discretization is chosen. "Optimal" in this context, means under limitations of computational power. The process of structure learning and discretization is repeated iteratively. We investigate our method using synthetic datasets, generated with the NGA model of Boore and Atkinson (2008) and finally apply our method to the NGA dataset to learn a BN modeling the Ground-motion caused by earthquakes.