AN ALGORITHM FOR THE AUTOMATIC CONSTRUCTION OF BAYESIAN NETWORKS WITH LIMITED DOMAIN KNOWLEDGE, AS APPLIED TO THE PREDICTION OF ECONOMIC AND DEVELOPMENT INDICATORS OF 248 COUNTRIES AND WORLD REGIONS

Fernando Javier Torre-Mora

Dr. Yi Shang, Thesis Supervisor

ABSTRACT

Humans have a natural tendency to express knowledge in terms of generalities, instead of individually measurable variables. However, to make any computational sense of a domain, information must be expressed in terms of measurable variables. Current computational methods either require the domain representation to be expressed in terms these variables (which can be hard for a human in many domains), or seek to discover these relationships by assuming the variables are not generalizable (ignoring human knowledge entirely). This project proposes a method to allow representational machine learning methods to be complemented by expert knowledge even when said knowledge is incomplete. The method is implemented as a Bayesian network construction algorithm using the standard error for a least-squares linear regression (STE) on the training data to evaluate each arc—as such, it is a variation of the method proposed by Friedman, Nachman, and Peér in 1999. A proof is included that this variation allows the computation time to be reduced, from factorial order, to quadratic order, with respect to the number of variables. The domain it is tested on is economic forecasting, using models from the literature. In particular, the method's capabilities in reproducing the eminent Smets-Wouters economic model, and in creating a new concrete model from the abstract ideas in the UNESCO world engineering report, are evaluated. The document concludes by studying the differences when comparing the generated models to each other, and to the Smets-Wouters economic model.