

**Dissertation press release****3.12.2020**

# Probabilistic Modeling for Extracting and Combining Human Knowledge with Machine Learning Models

<b>Title of the dissertation</b>	Methods for probabilistic modeling of knowledge elicitation for improving machine learning predictions
<b>Contents of the dissertation</b>	<p>Datasets with a small number of samples and a large number of variables are common in many real-world applications. Statistical modeling in such applications is challenging and it is known that models have limits in how low in sample size they can go. An approach to alleviate these challenges is to rely on other sources of information to be combined with the modeling to improve the final model. One of these additional sources is the domain expert; a person who knows the application of interest well enough. However, extracting knowledge from human expert is difficult.</p> <p>This dissertation developed approaches to efficiently extract knowledge from human users and effectively combine them with machine learning models through Bayesian framework to improve the predictive performance of models. Among different types of knowledge, this thesis work focused on knowledge on variables, in particular (i) relevance of single variables to a task at hand, and (ii) pairwise similarities of variables. The conclusion from this study is that in cases where the training data contains small number of samples with large number variables, if modeled properly, expert knowledge on features can bring predictive gain for the model comparative to collecting more samples.</p> <p>Literature is replete with various approaches for extracting knowledge from expert and combining them with the model. However, the literature mostly focused on extracting knowledge on samples. The approaches developed in this dissertation are complementary to those approaches in the literature and the combination of these methods can be adopted in applications with small number of samples and large number of variables to improve modeling.</p>
<b>Field of the dissertation</b>	Machine Learning
<b>Doctoral candidate</b>	Homayun Afrabandpey, M.Sc.
<b>Time of the defence</b>	18.12.2020 time 15:00 (EET)
<b>Place of the defence</b>	Remote via Zoom: <a href="https://aalto.zoom.us/j/65442048371">https://aalto.zoom.us/j/65442048371</a>
<b>Opponent</b>	Professor Patrick Shafto, Department of Mathematics and Computer Science, Rutgers University, USA
<b>Custos</b>	Professor Samuel Kaski, Aalto University, School of Science, Department of Computer Science
<b>Electronic dissertation</b>	
<b>Doctoral candidate's contact information</b>	Homayun Afrabandpey, Nokia <a href="mailto:homayun.afrabandpey@aalto.fi">homayun.afrabandpey@aalto.fi</a>

---