Interactive user intent modeling for eliciting priors of a normal linear model

Iiris Sundin[1] **Luana Micallef**[1] **Pekka Marttinen**[1] **Muhammad Ammad-ud-din**[1]

Tomi Peltola[1] **Marta Soare**[1] **Giulio Jacucci**[2] **Samuel Kaski**[1]

[1] Helsinki Institute for Information Technology HIIT
Department of Computer Science
Aalto University
[2]Helsinki Institute for Information Technology HIIT
Department of Computer Science
University of Helsinki
`first.last@hiit.fi`

Abstract

In this extended abstract, we present a novel interactive method to elicit the tacit prior knowledge of an expert user for improving the accuracy of prediction models. The main component of our method is an interactive user intent model that models the domain expert's knowledge on the relevance of different features for a prediction task. The user intent model selects the features on which to elicit user's knowledge sequentially, based on the earlier user input. The results of a feasibility study show that our method improves prediction accuracy, when predicting the relative citation counts of scientific documents in a specific domain.

1 Introduction

Interactive machine learning has the potential to improve predictive performance of models when the number of samples is low compared to the number of features in the data. Prediction models can be improved by prior knowledge that indicates relevant variables and parameter values. Yet, this prior knowledge is often tacit, only available from domain experts, and eliciting such tacit knowledge is difficult and excessively laborious when the number of features is large. On the other hand, fitting regression models in the "small n large p" problem requires regularizing the regression coefficients of the model [1, 2, 3]. Typically, the level of the regularization is tuned by estimating a hyperparameter from the data, but this neglects prior information that could be available on the problem. Indeed, knowledge of how the features affect the predicted target variable can significantly improve predictions [4]. Knowing the weights of the features will effectively aid relevance determination in sparse data.

Our contribution is a novel method that interactively models the domain expert user's tacit knowledge and uses this knowledge as prior information for improved predictions. An interactive user intent model (or a user model in short) selects features for which the user then indicates the relevance. Here, a relevant feature is a feature that is positively correlated with the target value in general, even if not necessarily in the training data. The user intent model iteratively elicits this information to build a model of the user's tacit knowledge, and uses sequential decision making to select other features that would benefit from the user's input. The user input is then encoded into prior knowledge for

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

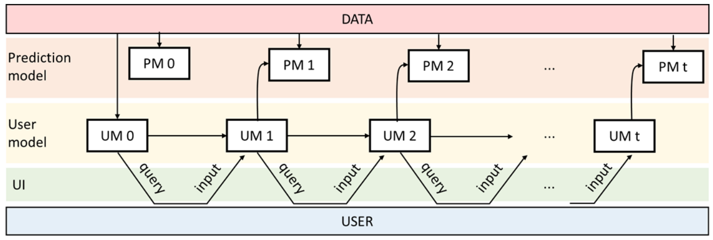

Figure 1: Illustration of the proposed method for eliciting tacit user knowledge for improving predictions. The user intent model (UM) queries the user to give input to a few features. The user indicates the relevances of the shown features, after which the user intent model and prediction model (PM) are updated accordingly, resulting in improved predictions in successive iterations t.

the prediction model to improve its accuracy. We present here results from a feasibility study on real-world data.

2 Related work

Expert knowledge can be integrated to prediction models by defining prior distributions for model parameters. Typically prior elicitation requires the experts to define full prior distributions [5, 6], which is time consuming and not feasible in high-dimensional problems even if interactive tools are provided. A simpler way to integrate expert knowledge was studied for Bayesian Networks in [7], where the expert only gives input on the presence or absence of the most uncertain causal relationships. The problem of choosing which features to elicit the information from can be framed as an information retrieval problem. In information retrieval, relevant resources can be found based on user's previous input, using interactive intent modeling [8]. Deciding which features to ask user input on is done iteratively, by balancing the *exploitation* of the currently most promising features and the *exploration* of uncertain, possibly interesting ones. The balancing is done with linear bandit algorithms [9]. However, to date this approach has only been used for information retrieval and not for prior elicitation.

In [10], possibly important features were shown to the user and included interactively to a classifier of webpage topics. A visual approach in [11] was used to address the quality of a classifier by showing interpretable features that best explain predictions locally. As a result, the user could also reject features in order to improve the classifier. However, simply including or excluding a feature is sensitive to errors and not sufficient for the "small n large p" problems. The method in [12] tackles this problem with the simplifying assumption that the expert may give noisy input directly on the regression coefficients. In addition, [13] performed a direct elicitation of logistic regression coefficients non-interactively. However, none of the current methods use user input on interactively selected features to improve the accuracy of prediction models.

3 System and model description

Our method is illustrated in Fig. 1. First, the prediction model (PM) and the user model (UM) are initialized. The user model uses sequential decision making to select a set of features to show next to the user. The user then indicates the relevances of the shown features for the given prediction task, based on her prior tacit knowledge. After that, the user model and the prediction model are updated using the relevances provided by the user. The prediction accuracy improves iteratively as the user gives more input. Below we briefly describe the prediction model and the user model.

3.1 Prediction model

As input, the prediction model takes the training data points $(\mathbf{x}_i, y_i),\ i = 1, \ldots, N$, where $\mathbf{x}_i \in \mathbb{R}^K$ are the features and $y_i \in \mathbb{R}$ the value of the target variable for sample i. In addition, a vector of relevances $\mathbf{r} \in \{0,1\}^K$ is provided, where $r_j = 1$ if the feature is relevant, i.e., has received positive

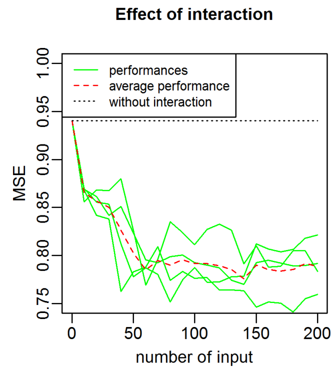

Figure 2: Mean squared error and the average performance w.r.t. the number of inputs provided by the participants of a feasibility study, compared with the MSE of a non-interactive model.

user input, and $r_j = 0$ otherwise. We assume a linear prediction model

$$y_i \sim N(\mathbf{x}_i^T \mathbf{w}, \sigma^2), \ i = 1, \dots, N,$$

where $\mathbf{w} \in \mathbb{R}^K$ is a vector of regression coefficients and σ^2 the noise variance. The relevances $\mathbf{r}$ enter the prediction model by modifying the prior distribution of the elements of $\mathbf{w}$ as follows:

$$w_j \sim N(0, \sigma_0^2), \text{ if } r_j = 0,$$
$$w_j \sim \text{half-}N(0, a\sigma_0^2), \text{ if } r_j = 1.$$

Here, half-N denotes the half-normal distribution. The intuition is that if a feature is relevant, the corresponding regression weight is assumed to have a wider prior distribution ($a \geq 1$) and is constrained to be positive.

3.2 Interactive user intent model

Efficient interaction balances between querying additional input on either the most promising relevant features (*exploitation*), or on the most uncertain ones (*exploration*). This is achieved by using the upper confidence bound criterion (UCB) to select features to show to the user, as in the algorithm LINREL [9]. At each iteration t, a user is shown n_t features with highest UCBs from the previous iteration. The user then specifies a binary relevance $r_j \in \{0, 1\}$ value to each feature j. We denote the inputs collected from the user before or at iteration t by $\mathbf{r}_t \in \mathbb{R}^{\sum_{i=1}^t n_i}$. At each iteration, the user intent model updates the estimated feature relevances $\hat{r}_{j,t}$ using a linear model:

$$\hat{r}_{j,t} = Z_j \hat{\mathbf{v}}_t + b \quad \forall \quad j \in 1, \dots, K$$
$$\hat{\mathbf{v}}_t = (Z_t^\top Z_t + \lambda \mathbf{I})^{-1} Z_t^\top (\mathbf{r}_t - b),$$

where $Z \in \mathbb{R}^{K \times N_Z}$ is a feature descriptor matrix, and its sub-matrix Z_t contains the descriptors $Z_j \in \mathbb{R}^{1 \times N_Z}$ corresponding to features that have received user input thus far, λ is a regularizer, and b determines the default relevance.

The UCBs are defined as $r_{j,t}^{UCB} = \hat{r}_{j,t} + c_{j,t}$, where $c_{j,t}$ is a high probability bound for relevance uncertainty, computed using SupLinUCB in [14].

4 Evaluation and results

We conducted a feasibility study with 4 participants as a preliminary evaluation of our approach in a real-world scenario. We compare the performance of a non-interactive prediction model, that is the prediction model in Section 3.1 without any user input, and the interactive method presented in Section 3. The task was to predict the relative citation count a scientific document will get in the domain of Artificial Intelligence (target variable) given that it has certain words (features) in the title, abstract or keywords. The participants had to indicate whether each of the 10 suggested features were relevant or not to the target, for 20 iterations. The data we used was a subset of Tang et al.'s citation data set [15] containing 162 scientific documents, for which we retrieved keywords manually and using Python Rake [16] and KP-Miner [17] keyword extractors, resulting in 457 unique keywords. The data collection was evenly split into training and test sets.

The final predictions of the interactive prediction model were more accurate than those of the non-interactive prediction model for all 4 participants. The user input always increased prediction accuracy, and the Mean Squared Error (MSE) decreased as the participants provided more input (Fig. 2). MSE without user input was 0.94, and with user input at the end of interaction 0.79 (mean) $\pm$0.025 (sd). Thus, without user input the prediction model explained about 6% of the variance in the target variable, and 21% with the user input received during interaction. The amount of explained variance is small, which illustrates the difficulty of the "small n large p" problem in noisy data. In this case study, the number of samples (81) is much smaller than the number of features (457), and the data set is very sparse.

5 Discussion and future work

In this extended abstract, we have presented a novel method for improving the accuracy of a prediction model in "small n large p" problems using interactive user intent modeling to model the domain expert's understanding of features' relevance. A feasibility study indicates that this approach is promising for improving prediction accuracy in contrast to a non-interactive prediction model. In this first feasibility study, we compared a prediction model which does not use any user input with the full system that uses both interaction and the user input. However, for the complete evaluation of the proposed method, an empirical study which also compares the proposed interactive selection of the features with randomly selected features will be conducted.

The future work will also include conducting a user study with more users to test the effectiveness of the approach. In addition, although providing user input on positive effects was natural for the prediction task considered here, in other cases negative user input may be useful. Furthermore, if the user is assumed to know more of the problem, we might also consider a continuous input instead of binary. Another task is to extend the method to multiple output learning. We will consider these questions in future work.

Acknowledgments

We would like to thank the participants of our user study, and the reviewers for their insightful and helpful comments. We acknowledge the computational resources provided by the Aalto Science-IT project. This work was funded by the Academy of Finland (Centre of Excellence in Computational Inference Research COIN, grant no. 284642; and grants no. 286607 and 294015 to PM).

References

[1] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[2] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[3] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[4] Atish P Sinha and Huimin Zhao. Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Decision Support Systems*, 46(1):287–299, 2008.

[5] Paul H. Garthwaite, Shafeeqah A. Al-Awadhi, Fadlalla G. Elfadaly, and David J. Jenkinson. Prior distribution elicitation for generalized linear and piecewise-linear models. *Journal of Applied Statistics*, 40(1):59–75, 2013.

[6] Joseph B. Kadane, James M. Dickey, Robert L. Winkler, Wayne S. Smith, and Stephen C. Peters. Interactive Elicitation of Opinion for a Normal Linear Model. *Journal of the American Statistical Association*, 75(372):845–854, 1980.

[7] A. Cano, A. R. Masegosa, and S. Moral. A Method for Integrating Expert Knowledge When Learning Bayesian Networks From Data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(5):1382–1394, Oct 2011.

[8] Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. Interactive Intent Modeling: Information Discovery Beyond Search. *CACM*, 58(1):86–92, 2015.

[9] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *The Journal of Machine Learning Research*, 3:397–422, 2003.

[10] Michael Brooks, Saleema Amershi, Bongshin Lee, Steven M Drucker, Ashish Kapoor, and Patrice Simard. FeatureInsight: Visual support for error-driven feature ideation in text classification. In *IEEE VAST*, pages 105–112, 2015.

[11] Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *ACM SIGKDD*, 2016.

[12] M. Soare, M. Ammad ud din, and S. Kaski. Regression with $n \to 1$ by Expert Knowledge Elicitation. In *arXiv preprint arXiv:1605.06477*, 2016.

[13] Rebecca A. O'Leary, Samantha Low Choy, Justine V. Murray, Mary Kynn, Robert Denham, Tara G. Martin, and Kerrie Mengersen. Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby. *Environmetrics*, 20:379–398, 2009.

[14] Wei Chu, Lihong Li, Lev Reyzin, and Robert E. Schapire. Contextual Bandits with Linear Payoff Functions. In *Proc of AISTATS*, pages 208–214, 2011.

[15] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: extraction and mining of academic social networks. In *Proc of ACM SIGKDD*, pages 990–998, 2008.

[16] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. Automatic keyword extraction from individual documents. *Text Mining*, pages 1–20, 2010.

[17] Samhaa R El-Beltagy and Ahmed Rafea. KP-Miner: A keyphrase extraction system for English and Arabic documents. *Information Systems*, 34(1):132–144, 2009.