ПРОБЛЕМИ РОЗВИТКУ ФІНАНСОВО-ЕКОНОМІЧНОГО КОНТРОЛЮ, ЕКОНОМІЧНОГО АНАЛІЗУ ТА СТАТИСТИКИ В УПРАВЛІННІ БІЗНЕСОМ

JazaAbir,
Departement « Dauphine Recherches en Management »,
Université Paris Dauphine, France
Svitlana Galeshchuk,
Ternopil National EconomicUniversity, Université Paris Dauphine, France

INTERACTIVE PREFERENCE ELICITATION IN COST-UTILITY ANALYSIS

Cost-utility analysis is one of the predominant approaches in economic evaluation. Recently a suite of machine learning methods helped improve the analysis with the state-of-the-art developments, in particular supervised classification technics. Preference elicitation tools assist the decision maker.They do so by interactively learning the decision making's preferences through appropriately chosen features (e.g., queries) and suggesting high-quality outcomes based on the preference estimates (Dragone et al., 2018). Thus, obtaining classifiers based on a low-cost set of input features with acceptable classification accuracy using the preference elicitation theory is of interest to practitioners and researchers.

This paper aims at providing a theoretical foundation of the method for obtaining low cost classifiers that meet specified accuracy requirements under dynamically changing costs. Given a set of relevant input features and accuracy requirements, the goal is to identify all qualifying classifiers based on subsets of the feature set.Then, for anyarbitrary costs associated with the features, the cost of the classifiers may be computed and candidate classifiers selected based on cost-accuracy tradeoff. Since the number of relevant input features k tends to be large for some 2^{k-1} problems, training and testing classifiers based on all possible non-empty subsets of features is computationally prohibitive. Under the reasonable assumption that the accuracy of a classifier is no lower than that of any classifier based on a subset of its input features, developing an efficient method to identify all qualifying classifiersis required.

For a given cost function on the input features, qualifying classifiers may be ranked based on weighted averages of the accuracy measures and input feature cost and classifiers on the Pareto optimal frontier that are not dominated by other classifiers in the cost-accuracy space may be identified.The total cost of applying an active classifier may be estimated as the total cost of obtaining values for the features (attributes) required and the sum of penalty if theclassification outcome is wrong (Grenier et al., 2002). The benefits of the preference elicitation method may be demonstrated by obtaining costs and acceptable accuracy thresholds for the proposed classifier identification system.

Formal Definition of the classifier identification problem:

Let $D^{F,m}$ be a classifier trained on a set of input features F using supervised machine learning method m. Let $A_j(D^{F,m})$ be the expected accuracy of $D^{F,m}$ as determined by testing the trained model with respect to some accuracy measure j.Given minimum acceptable accuracy thresholds $l_j$for accuracy measures $j \in \mathcal{A}$, $D^{F,m}$ is said to be *acceptable*with respect to $\mathcal{A}$ if $A_j(D^{F,m}) \geq l_j \; \forall j \in \mathcal{A}$. Given an *acceptable* classifier $D^{F,m}$ with respect to minimum acceptable accuracy thresholds l_j for $j \in \mathcal{A}$, identify all subsets $F_s \subset F$ such that $D^{F_s,m}$ is also acceptable.

Since the size $|F|$ of the feature set may be large, it may be computationally infeasible to train and test classifiers using all $2^{|F|} - 1$ subsets of features. We make a reasonable assumption to make the problem tractable:If a classifier $D^{F_s,m}$ is not acceptable, then all classifiers trained on a proper subset of F_s are also not acceptable.Under this assumption, this approach will apply depth first tree search to identify all subsets of F that result in acceptable classifiers. Nodesare represented by the feature set used to train and test the model. The root node isrepresented by the full feature set *F*. For a node represented by feature set F_s, a set of $|F_s|$ successor nodes is obtained as $\{F_s - \{f\} | f \in F_s\}$ by removingone feature at a time. Nodes representing classifiers that are not acceptable are terminal nodes in our search tree; all non-terminal nodes are acceptable classifiers.

A judicious choice of the order in which features are removed can reduce the total number of classifiers to be trained and tested, thus reducing the total time needed to identify the set of acceptable classifiers.The relative importance of features in a classifier may be estimated. The total number of nodes in our search tree may be reduced by considering a successor node obtained by the removal of a relatively more important feature before a successor node obtained by the removal of a relatively less important feature.

Given acceptable accuracy thresholds, we shall identify all acceptable classifiers. For each acceptable classifier, its accuracy profile will be presented in terms of all accuracy measures $j \in \mathcal{A}$. Each classifier will be presented as a set of rules. Once the relative costs for obtaining the features have been elicitated, the cost associated with applying a trained classifier $D^{F,m}$ may be taken to be the sum of the cost of obtaining values for its input feature set F. Let c_i be the cost of obtaining a sample value for feature i. Then the cost of applying $D^{F,m}$ may be computed as $C(D^{F,m}) = \sum_{i \in F} c_i$. Using these costs, a non-dominated set of classifiers in the Pareto-optimal frontier of the cost-accuracy space will be obtained. Decision makers may make informed decisions regarding the results to obtain by selecting a subset of features used by some model in this frontier.

References

1. Greiner, R., Grove, A. J., & Roth, D. (2002). Learning cost-sensitive active classifiers. Artificial Intelligence, 139(2), 137-174.
2. Dragone, P., Teso, S., &Passerini, A. (2018). Constructive preference elicitation. Frontiers in Robotics and AI, 4, 71.