SUPPORT OF MANAGERIAL DECISION MAKING BY TRANSDUCTIVE LEARNING

Hubertus Brandner
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
D-20146 Hamburg, Germany

Stefan Lessmann
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
D-20146 Hamburg, Germany

Stefan Voß
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
D-20146 Hamburg, Germany

ABSTRACT

Transductive inference has been introduced as a novel paradigm towards building predictive classification models from empirical data. Such models are routinely employed to support decision making in, e.g., marketing, risk management and manufacturing. To that end, the characteristics of the new philosophy are reviewed and their implications for typical decision problems are examined. The paper's objective is to explore the potential of transductive learning for corporate planning. The analysis reveals two main factors that govern the applicability of transduction in business settings, decision scope and urgency. In a similar fashion, two major drivers for its effectiveness are identified and empirical experiments are undertaken to confirm their influence. The results evidence that transductive classifiers are well superior to their inductive counterparts if their specific application requirements are fulfilled.

1. INTRODUCTION

Methods and models for information and data processing are the main focus of *information systems* (IS) (see, e.g., [21, 56]). Within the discipline, the support of managerial decision making has a long tradition and dates back to - at least - *Gorry and Morton's* well known framework for structuring different types of information systems [26]. Roughly speaking, computer-based tools for decision aiding provide access to heterogeneous sources of information and the functionality required for filtering, relating and aggregating such information to gain insight into business processes, identify moderators of process efficiency and effectiveness, and eventually form adept business decisions. Respective systems are commonly referred to as analytical information systems (AIS) or business intelligence systems (see, e.g., [12, 22, 23, 38]).

Data Mining is part of the AIS-family and stipulates a machine-centric approach towards decision support [5]. Specifically, the core of Data Mining consists of a set of methods, each of which is designed for a particular analytical task. For example, association rule mining identifies co-occurrences of frequent items in transactions, such as products commonly purchased together in shopping transactions, whereas cluster analysis facilitates an autonomous categorization of objects into subgroups, a common task in marketing to segment a heterogeneous market into more homogeneous segments [7, 8, 20]. A somewhat more guided approach is taken in classification analysis. Here, a functional relationship between a discrete variable of interest (i.e., a class label) and a set measurement is inferred from past data with the objective to employ the resulting function for prediction.

Applications of classification in corporate settings are manifold and include, e.g., process and quality control in manufacturing, the screening of loan applications in the financial service industry, detecting fraudulent transactions in the telecommunication or insurance business as well as marketing decision support in response modeling or customer attrition analysis (see, e.g., [42, 51]). Several authors have argued that the predictive accuracy of classification models is imperative in such applications (see, e.g., [41, 44, 53]). This view can be understood by noting that the number of predictions in large-scale corporate applications is massive. Consequently, even marginal gains in accuracy may translate into substantial financial returns [36].

Therefore, the role of IS in the quest for improved decision quality comprises monitoring methodological advancements in computational learning and statistical inference. In fact, it has been argued that a key responsibility of IS as scientific discipline consists of bridging the gap between method-centric domains such as computer sciences or statistics and application-supplying fields like business administration [37, 39]. A similar argument may be put forward from the perspective of innovation management (see, e.g., [29]). According to the *technology-push* hypothesis, advancements in basic research are a key contributor towards innovation and economic growth (see, e.g., [21, 40]), which emphasizes the importance of IS as an integrative discipline that matches business requirements with technological opportunities.

In hunting for ever more accurate prediction models, a vast number of different classification methods have been considered and evaluated in corporate applications (see [35] for a survey). The present study is basically in line with these endeavors and examines the potential of transductive learning [54] to aid managerial decision making. However, as will become clear in the remainder of the paper, transductive learning is more than a new *method.* Transductive learning is a different philosophy towards constructing predictive classification models from data. Compared to the classi-

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

cal inductive paradigm, transductive inference (TI) pays a price in terms of universality of application, but holds the promise of increased robustness and higher predictive accuracy in particular settings. Specifically, TI stipulates a direct estimation of class memberships to simplify the overall modeling task and circumvents the detour of building a general model to predict specific cases, characteristic to classical statistical inference.

The efficacy of TI has been evidenced in several studies (see, e.g., [6, 31, 46]). However, experiences in corporate environments are yet lacking. The objective of this paper is thus to introduce the novel approach to a business audience and disclose planning tasks that could benefit from its application. The potential of TI to increase decision quality in the settings identified must be sought in its ability to deliver more accurate class predictions. Consequently, the influence of TI's specific features on forecasting performance is appraised to clarify how accuracy gains may be attained and under which circumstances they should be particularly exposed. To complement the conceptual analysis of TI's potential, an empirical study is undertaken. Using data from the field of risk-management in consumer lending, the performance of transductive techniques is compared to their inductive counterparts in different scenarios. The results observed confirm the general superiority of the former and provide valuable insight into the prerequisites for its success.

The remainder of the paper is organized as follows: Section 2 provides a brief introduction into the theory of transductive learning and its background in statistical learning. Subsequently, specific algorithms embodying this learning paradigm are introduced. Section 3 examines the implication of TI's characteristics to discern factors governing its applicability and effectiveness in business applications. To verify the importance of the identified concepts, an empirical case study is undertaken whose results are presented in Section 4. Section 5 concludes the paper with a summary of key findings and an outlook to future research.

2. BACKGROUND

2.1 Classification

Given a set of u objects $\{x_j^\star\}_{j=1}^u$, the aim of classification is to predict corresponding group memberships $\{y_j^\star\}_{j=1}^u$. The objects are characterized by a set of n attributes, which are believed to determine an object's class. Thus, all $x_j^\star$ are vectors in $\mathcal{R}^n$. For example, the objects could represent assembled products, which are to be categorized into the groups functional or defective; similarly, a marketing objective could be to distinguish customers who are responsive to direct mail from those who aren't, on the basis of demographical and transactional customer attributes. Without loss of generality [1], we focus on such binary classification problems, whereby the two possible classes are encoded as -1 and $+1$ in the following.

To perform the categorization, a classification model or classifier is derived from an example set $D = \{(x_i, y_i)\}_{i=1}^l$ of l observations with known class memberships. D is referred to as learning or training set. Keeping in mind the goal of classification, the overall model building process and the use of D in particular need to be organized in a way so as to maximize accuracy when classifying novel objects not contained in the training set. Different learning paradigms can be distinguished according to their philosophy to pursue this objective.

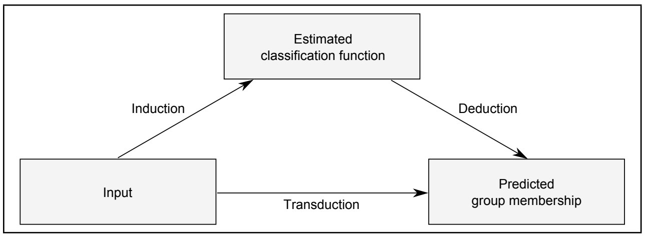

Figure 1: Inductive-deductive vs. transductive classification [54]

2.2 Inductive vs. Transductive Inference

Inductive inference and TI are different means towards building classification models from empirical data. The classical approach consists of two steps, induction and deduction. First, a classification function $f : x \mapsto y \; \forall x \in \mathcal{R}^n$ is derived from D, e.g., by minimizing the mismatch between model-estimated $(f(y_i))$ and actual class labels (y_i) over D. This model is subsequently employed for predicting class memberships for (arbitrary) novel objects. Note that the construction of a general model from a particular sample of data is a challenging undertaking, which holds several pitfalls (see, e.g., [16, 50]). Roughly speaking, representativeness of D for the whole problem space is essential, but may not always be taken for granted.

TI grounds on the observation that the complexity associated with building a global model from a limited sample can and should be avoided in settings where the objects to be classified are known in advance. That is, only the group memberships of some given objects are unknown and need to be predicted [54]. Whenever this requirement is met, the intermediate step of building a global model is unnecessarily complex and dispensable. Class labels should better be estimated in a direct manner [14, 43]. The conceptual difference between the two learning philosophies of transductive and inductive learning is illustrated in Figure 1.

In the transductive setting, class predictions are only sought for a clearly defined, a priori known set of objects, the *working set* $\{x_j^\star\}_{j=1}^u$. In this case, TI holds the promise to increase the accuracy and robustness of class predictions. To achieve this, the working set is considered alongside the ordinary training data during classifier construction. In other words, a classification rule is inferred from data comprising both, labeled and unlabeled examples. This way, a transductive classifier has access to the working set and the additional information contained therein. In other words, it is designed to handle precisely (and only) these objects. This differs notably from an inductive setting where the objects to be classified remain unknown until the (global) model is built. Therefore, a transductive classifier solves a much simpler estimation task and is thus less vulnerable to distributional discrepancies between training data and the working set. Consequently, it can be expected to be more accurate in classifying the points of interest (i.e., the working set) [14, 15, 17].

2.3 Support Vector Machines

Support Vector Machines (SVMs) are a popular method to construct inductive classification models. However, they can be extended to perform TI in a straightforward manner [31]. Therefore, SVMs facilitate an unbiased comparison between the two principles of inductive and transductive classification and will serve as representatives of both learning paradigms within the empirical evaluation. The respective procedures are sketched in the following, whereas a comprehensive introduction into the theory of SVMs can be found in textbooks like [17] or [28].

2.3.1 Inductive Support Vector Machines

The concept of inductive SVMs (iSVMs) is illustrated in Figure 2. Given a training set of examples together with their class membership, a hyperplane H is constructed so as to separate objects of adjacent classes with large margin. The concept of maximal margin is key to the SVM approach and has been shown to increase the accuracy of class predictions for novel objects [6, 54]. The hyperplane represents an inductive classification model, which classifies novel instances according to their relative position (above or below) to H. The model's parameters (normal and intercept) are estimated by solving a convex quadratic program for the training set (see, e.g., [28]).

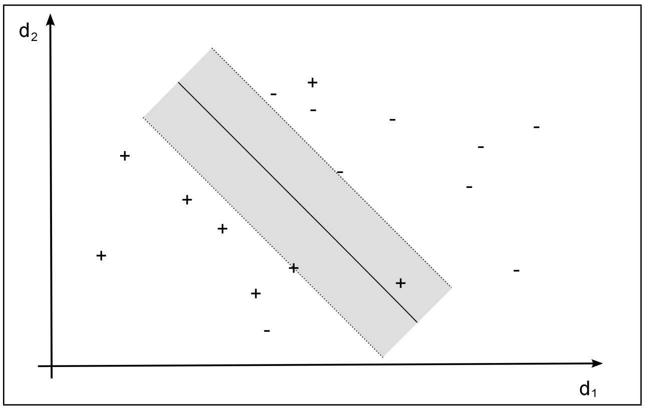

Figure 2: Inductive SVM in a linearly non-separable case in $\mathcal{R}^2$. $+, -$ represent training set objects $\{x_i\}_{i=1}^{l}$ and their classes ($y_i \in \pm 1$), respectively. The gray rectangle depicts the margin between the two classes. New data points $x_j^\star$ are classified according to their position, below ($f(x_j^\star) = y_j^\star = -1$) or above ($f(x_j^\star) = y_j^\star = +1$) the plane H.

2.3.2 Transductive SVM

A transductive SVM (tSVM) implements a similar strategy and differs only in the approach to determine the location of the separating hyperplane. Having access to the working set and thus information about the location of the points to classify, additional constraints concerning the orientation of H can be imposed: Objects with known class (i.e., the training set) should again be separated with large margin to increase the classifier's ability to generalize (see above). However, this principle cannot be applied to working set objects directly, since their class is, by definition, unknown. In order to approximate margin maximization over the working set, the hyperplane is pushed into a region with low data density. This approach is motivated by the *cluster-assumption*, which claims that data points are likely to be of the same class, if they are *close* to each other, i.e., in the same cluster of the space [13]. Consequently, pushing the hyperplane away from the unlabeled points can be expected to maximize the margin of separation over all data, the training set and the working set [17]. More specifically, it is reasonable that tSVM achieves larger margin, and thus higher accuracy, on working set examples, because these are considered during model fitting. To see this, consider Figure 3 that continues the previous example, but also depicts a possible set of points to classify.

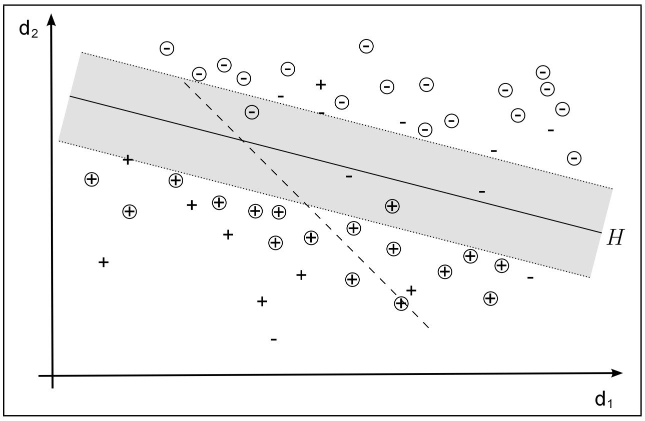

Figure 3: Transductive SVM in a linearly non-separable case in $\mathcal{R}^2$. $\oplus, \ominus$ represent the working set $\{x_j^\star\}_{j=1}^{u}$ and their class labels $\{y_j^\star\}_{j=1}^{u}$, as estimated by the transductive classifier (solid line). The dashed line equals the inductive classifier of Figure 2.

In appraising Figure 3, it is important to remember that the true class membership of working set objects is unknown. However, by visual inspection, one would assume that the class boundary of tSVM mimics the true relationship between object attributes and class membership more closely than those of iSVM since less points of the working set fall into the margin, i.e., the region around the estimated class boundary. This can be expected to provide more accurate class predictions over working set instances.

In order to take account of the working set during model fitting, the mathematical program underlying SVMs needs to be extended. The tSVM formulation used in this work has been proposed by [31]. It involves solving a mixed-integer program. Therefore, standard SVM training algorithms are no longer applicable. In order to eliminate factors influencing classification accuracy other than the embodied learning principle, we develop a novel metaheuristic that facilitates solving different variants of the SVM learning problem in a unified way. The procedure can be characterized as a *memetic* algorithm (see, e.g., [27]). It should be noted that the heuristic computes the equation of the optimal hyperplane (i.e., the global model) as byproduct of the learning phase. Although not the aim of transduction, H could be employed to classify all other $x \in \mathcal{R}^n$.[1]

---

[1] This principle, learning from both labeled and unlabeled examples to build global models, is called Semi-Supervised

3. APPLICABILITY AND EFFECTIVENESS OF TI IN BUSINESS CLASSIFICATION

In order to appraise the managerial utility of TI, it is essential to fully understand how its specific features create value (i.e., increase predictive accuracy), in which circumstances they are most effective, and how they affect TI's applicability in corporate contexts. In particular, a key characteristic and requirement of TI concerns the availability of the working set; in addition to a training dataset - always needed in predictive modeling - all objects which are to be classified need to be known in advance. The following sections clarify upon the implication of this particularity with respect to applicability and effectiveness of TI.

3.1 Moderators of TI Applicability

Considering typical business applications of classification (see, e.g., [35, 51]), it is easy to find examples where the requirement of an a priori given working is met. Consider, e.g., the task of targeting customers in direct marketing (see, e.g., [9, 32]). A mail-order company is well aware of the clients it could possibly solicit. In particular, the actual decision task is to select from a set of present customers those who are most likely to respond. Consequently, the objects to be classified are known in advance. A similar situation occurs in churn prediction (see, e.g., [41]): From the set of all current customers, a marketer wishes to identify those with highest attrition risk. Again, the objects to be classified are known in advance. On the other hand, decision support in credit scoring (see, e.g., [48]), also a popular application for classification in business, can be considered a counterexample. Here, loan approval models are employed to assess incoming credit applications, which have not been observed at the time the classification model was built.

The previous examples differ in terms of the number of objects that are classified. In credit scoring, decisions are sought for individual loan applications, whereas the marketing examples are associated with classifying a set of customers. We refer to this construct as decision scope and propose it as moderator of TI's applicability in business contexts. Specifically, TI benefits from wider decision scope, which involves handling a larger number of objects.

The previous assertion follows directly from the observation that if a decision concerns a (large) set objects, a working set containing these objects must be available. However, does this imply that singular decisions such as those in credit scoring prohibit use of TI? It seems that this is not necessarily the case. Consider for example the task of fraud detection in the insurance industry. Classification models are employed to screen incoming claims, e.g., to decide whether a claim can be settled immediately or whether it requires a closer examination by a human expert (i.e., appears suspicious). This task shares similarities with the credit approval example in that decision objects (insurance claims and loan applications, respectively) arrive periodically. However, as opposed to credit decisions which loan applicants will want to obtain almost instantly, insurance holders may not expect claims to be decided immediately, but accept some processing time. This delay will inevitably result in a queue of claims to be processed. Especially if the number of incoming claims is large, the amount of objects to classify (i.e., the working set size) may well be sufficient to justify a transductive approach. Therefore, less urgency in making decisions may facilitate use of TI, even if decision scope is narrow. Consequently, it seems well justified to consider decisions' urgency as an additional moderator of TI's applicability.

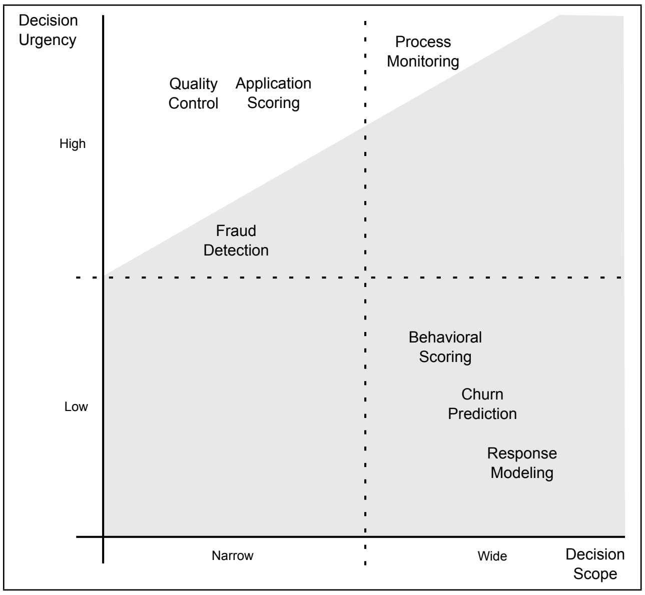

Figure 4: Decision problem characteristics and corresponding applications; the grey area depicts the region of TI-feasibility

The previous arguments are summarized in Figure 4. The grey zone within the decision urgency/decision scope portfolio delineates the general application area of TI. An ordinary scale has been employed to distinguish between two states for the identified moderators. In addition, some specific business classification tasks are given to exemplify applications suitable and unsuitable for TI.

Clearly, the location of classification tasks in Figure 4 is debatable and should be taken as a proposition. This is especially true for decision urgency, so that the actual applicability of TI for, e.g., insurance fraud detection depends on several factors such as company size, type of claim, etc. Therefore, it needs to be scrutinized on a case by case basis. In that respect, the ultimate objective of Figure 4 is to provide a tool for structuring and systematizing such an appraisal.

We refrain from discussing every single positioning decision in detail. However, the following considerations may help to motivate our proposition: Planning decisions in response modeling, churn prediction and behavioral scoring[2] involve a campaign approach where a model is built at a particular point in time and employed to obtain scores for a collection of customers. Differences in terms of decision scope among these three may exist because of variations in the frequency with which campaigns are launched. However, in comparison to other tasks such as credit scoring, the com-

---

Learning (SSL) and resides somewhat between inductive and transductive classification [13].

[2] Behavioral scoring is related to application (credit) scoring but employed at a different stage within the customer lifecycle (see, e.g., [47, 49]). In particular, once credit has been granted, financial institutions have an interest in estimating the likelihood of default to take preemptive actions for high risk borrowers.

pany faces fewer constraints in timing the task since customer expectations in terms of process time do not exist. Consequently, these three tasks' urgency is considered as low. Quality control involves classifying assembled products according to their compliance with predefined quality indicators. This task shares many similarities with credit scoring in the sense that objects are classified on a one by one basis and that these decisions are relatively time critical. For instance, products of a particular lot may not be sold until quality has been verified. Contrary to quality control, process monitoring refers to the surveillance of a whole (manufacturing) system. Classification aids this endeavor by processing large amounts of data gathered by various process monitors to determine individual components' reliability, which, in turn, allows conclusions regarding the whole system's soundness to be drawn. This suggest wider decision scope. The positioning of fraud detection has been elaborated above.

3.2 Moderators of TI Effectiveness

In order to provide more accurate class predictions than an inductive model, a transductive classifier needs to extract additional information from the working set, over and above those already contained in the training set. This becomes possible if the latter is not well representative for the objects to be classified, i.e., if the distribution between the two sets differs in some fashion. Consequently, factors governing distributional discrepancy between training and testing data can be expected to influence the effectiveness of TI.

In previous work, TI has been found to be most useful in settings where unlabeled data is easily available, whereas obtaining labeled data is associated with high cost (see, e.g., [13, 31]). Such scenario appears well representative for many business applications. For example, mail-order companies have (or can easily gain) access to enormous amounts of customer data. However, whether a particular customer has previously responded to a catalog mailing (i.e., the class label) is known only if the customer has ever been solicited. In other words, obtaining the customer's label requires sending a catalog and is thus costly.

The fact that TI does work well in the case of an imbalance between the amounts of labeled and unlabeled data can be explained by distributional discrepancy. In particular, for an imbalance to exists, the size of the training dataset has to be small, relative to the working set, and a small sample may not represent well the overall distribution. This is especially so in corporate settings where the training data is commonly not a random sample. In particular, the objects whose class labels are known (i.e., the ones that can be employed for training) have themselves been selected by a previous classification. For example, loans are exclusively granted to applicants with low default probability, which, in turn, has been estimated by a credit scoring model. Since the information whether a debtor eventually defaults becomes available only for such preselected applicants, the data that can be employed for model building is restricted to low-risk applicants and thus a biased sample of the distribution of all applicants.[3] Therefore, representativeness of training data must not be taken for granted and may well be limited in some corporate applications.

Besides statistical consideration, mathematical arguments suggest that the effectiveness of TI depends upon an imbalance between labeled and unlabeled data (or the lack thereof). As explained in the case of tSVMs, training examples and unlabeled cases are employed jointly when constructing a transductive classifier. Consequently, the effect of the (possibly less representative) labeled instances on the classifier will be excessive unless sufficient unlabeled objects are available.

In view of the previous reasoning, we propose that the effectiveness of TI increases if the imbalance between labeled and unlabeled data (i.e., the ratio l/u) decreases.

In addition to being of insufficient size or sampling issues, another reason for discrepancy between training data and the cases to classify is change in the processes that provide these datasets. Mail-order companies, for example, commonly solicit customers with catalog mailings on a quarterly basis. The needs and preferences of customers, and also their affinity towards direct-mail, are not fixed but undergo a constant change. Therefore, the rules once inferred by a classifier to identify responsive customers may loose sharpness over time and require updating to take account of more recent developments in, e.g., customer behavior. In the case of credit scoring or fraud detection, this issue is even more severe since applicants/fraudsters may deliberately attempt to alter their characteristics so as to circumvent decision support models and obtain favorable classifications [10]. Clearly, such action - if successful - will also diminish the appropriateness of a previously derived classifier.

In summary, the relevance of model updating depends upon the stationarity of the data generation process, whereby regular changes (less stationarity) dictate more frequent model updates. By classifying working set objects directly and thus circumventing the construction of a general classification model, TI can be interpreted as an extreme approach in terms of model updating: A new model is constructed for every decision. Consequently, process stationarity can be identified as another moderator of TI's effectiveness, which leads to the portfolio shown in Figure 5.

The grey area illustrates scenarios, which appear particularly suited for TI, i.e., can be expected to benefit from its ability to classify working set instances directly. On the contrary, employing unlabeled data during model building seems less effective if the mechanisms that govern the relationship between class labels and object characteristics remain stable over time or the ratio of l/u is large.

Concerning the positioning of example applications, we motivate our choices as follows: It is reasonable to distinguish business applications that involve classifying customers from decision tasks referring to machines or manufactured products. The former change their behavior far more rapidly than, e.g., an assembly line its operation. Consequently, the stationarity of data generation processes is decreased in human-centric settings. Assuming that the size of available training data depends mainly upon company size, working set size becomes the main discriminator between example applications in terms of imbalance. The horizontal position of applications is then obtained by employing decision scope (see Figure 4) as proxy for amount of objects to classify.

---

[3] This problem is known as sample-selection-bias and elaborated in the context of reject inference (see, e.g., [4]).

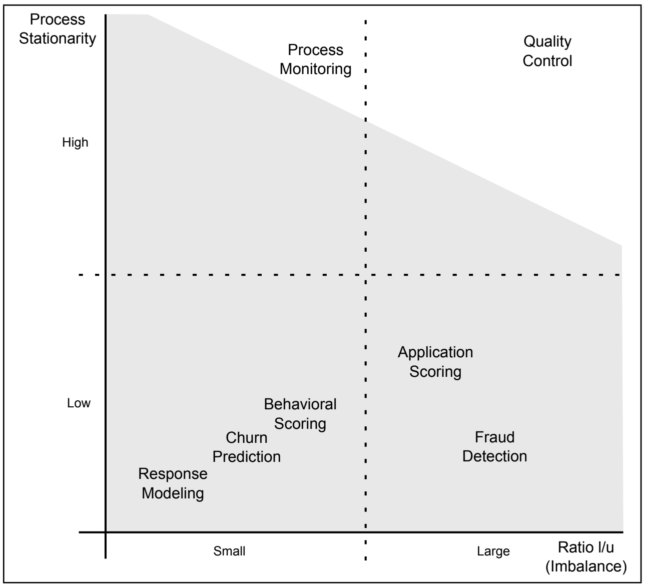

Figure 5: The marked area shows combinations of data imbalance and process stationarity promising an effective utilization of TI.

Overall, the managerial insight gained from the conceptual analysis of TI can be summarized as follows: When facing a classification problem, decision makers are well advised to examine the task's compliance with TI's requirements concerning decision characteristics. Problems that do exhibit wide scope and/or less urgency are candidate applications of TI and should be evaluated according to the imbalance between labeled and unlabeled objects and how likely changes in underlying data generation processes appear.[4] The procedures leading to such an appraisal have been exemplified in conjunction with seven business classification tasks and illustrate how promising candidate application can be identified. For the examples considered above, TI can be expected to be particularly effective in behavioral scoring, churn prediction and response modeling. It could also be considered in fraud detection applications, whereas application scoring, process monitoring and quality control remain an area for inductive classification.

4. EMPIRICAL EVALUATION

As pointed out by Hevner et al. [30], a careful and rigorous evaluation is crucial to ensure that an IT artifact is compatible with corporate requirements and successfully solves the problem it is meant to solve. The managerial utility of TI must be sought in its potential to give more accurate class predictions. As explained above, accuracy is often imperative in corporate settings and may offer substantial monetary rewards [41, 44, 53]. Whereas theoretical arguments in favor of TI's superiority over classical inductive learning have been put forward in the previous sections[5], additional empirical evidence is desirable to complement the evaluation of TI and confirm its effectiveness to deliver more accurate predictions.

4.1 Data

In order to verify the potential of TI for managerial decision support, inductive and transductive SVMs are evaluated on two real-life datasets: Australian (AC) and German Credit (GC). Both datasets originate from the domain of credit scoring and are publicly available in the UCI Machine Learning Repository [2]. They have been employed in numerous studies on classification (see, e.g., [3, 36, 52, 55]) and serve as examples for business classification problems in this study. Specifically, the binary target variable indicates whether a debtor x_i fulfills his obligation of repaying a loan ($y_i = -1$) or defaults ($y_i = +1$). Although this classification task belongs to the field of application scoring, which, due to narrow decision scope and high urgency, is less suitable for TI (see above), the variables provided to perform the classification are also relevant for behavioral credit scoring (see, e.g., [47]), i.e., a potential application area. Therefore, it seems appropriate to consider these datasets for the evaluation. Their characteristics are summarized in Table 1.

	$\lvert dataset \rvert$	n	A priori probability of class +1
AC	690	14	0.449
GC	1000	24	0.300

Table 1: Characteristics of credit scoring datasets.

4.2 Scenarios

A key objective of the comparison of inductive and transductive classifiers is to confirm the impact of imbalance and process stationarity on TI's effectiveness. To that end, two experimental scenarios are developed.

The impact of imbalance is examined by means of varying the ratio l/u, the cardinalities of the training and working set, respectively. Specifically, iSVM and tSVM models are built and compared in terms of their predictive accuracy for five settings of $l/u = 2, 4, 6, 8,$ and 10%. These ratios may be considered conservative; even more dramatic inequalities can be found in the literature [6, 15, 31].

To appraise the influence of distributional change, five discrete points in time are defined and the similarity between training data and working set data is decreased between consecutive time points. Specifically, distributional discrepancy is artificially introduced through systematical manipulation of the test set. Roughly speaking, the attributes of the test set examples are modified by adding a constant direction vector in general and a normal distributed variable individually. This is necessary since both sets are random samples drawn from the original dataset (i.e., AC or GC) and therefore equivalent in terms of their underlying data distribution.

---

[4] It should be noted that the problem of distributional change in data generation processes is examined in detail in the field of concept drift learning (see, e.g., [34]) and that dedicated procedures have been developed to cope with this modeling challenge (see, e.g., [33, 45]). However, these endeavors commonly concentrate on applications where new data arrives continuously (e.g., in streams) and is processed instantly. This differs notably from candidate applications of TI, where wide decision scope and low decision urgency cause/enable a batch-processing of the (working-set) instances to be classified.

[5] A formal mathematical discussion in terms of transductive and inductive techniques' ability to minimize bounds of generalization error can be found in [18].

4.3 Performance Measurement

Numerous approaches exist to measure a classifier's predictive power. We decided to assess performance by an indicator which grounds on the basis of a discrete categorization of predictions into four groups: false positive, false negative, true positive and true negative. The following contingency table depicts this principle.

		estimated class	
		-1	+1
real class	-1	true negative (tn)	false positive (fp)
	-1	false negative (fn)	true positive (tp)

The motivation for considering discrete class predictions as opposed to probabilistic or confidence based measures such as AUC (see, e.g., [19]) is that TI is designed to generate crisp classifications.

To measure the accuracy of classifications, we calculate the F_1-Score (FSC), a widely used metric in Information Retrieval, which is defined as the harmonic mean between precision $\frac{tp}{tp+fp}$ and recall $\frac{tp}{tp+fn}$ (see, e.g., [11])[6]:

$$F_1 := \frac{2 \ * \ \text{precision} \ * \ \text{recall}}{\text{precision} \ + \ \text{recall}}.$$

For each of the settings (imbalances or time points), a tenfold split-sample setup is employed. That is, the original dataset (AC or GC) is randomly partitioned into training and working set (i.e., labeled and unlabeled data) to construct and assess classification models. The resulting F_1-scores are then averaged to obtain a final performance estimate.

4.4 Empirical Results

4.4.1 Imbalance

Figure 6 summarizes the results obtained from experiments with increasing imbalance between the amount of labeled and unlabeled data. Individual settings are represented by stems, whose height measures ΔF_1-values, the difference between the F_1-Score of tSVM minus the F_1-Score of iSVM. Thus, a positive value implies superior performance of the transductive approach.

Overall, the observed results are in line with theory: tSVM consistently achieves higher performance, with the least imbalanced setting on GC being the only exception. Moreover, there is a clear trend of tSVM outperforming iSVM with increasing margin as imbalance increases. In other words, the more extreme the ratio between labeled and unlabeled data, the more accurate are the class predictions of tSVM compared to those of iSVM. This dependency is confirmed by a linear regression of imbalance setting (1,2,...,5) on ΔF_1, which gives an R^2 of 0.528 and is significant at the two-percent level. Therefore, one may speculate that the superiority of transductive methods over their inductive counterparts will be even larger, if more extreme ratios of $\frac{l}{u}$ are present (see also [31]).

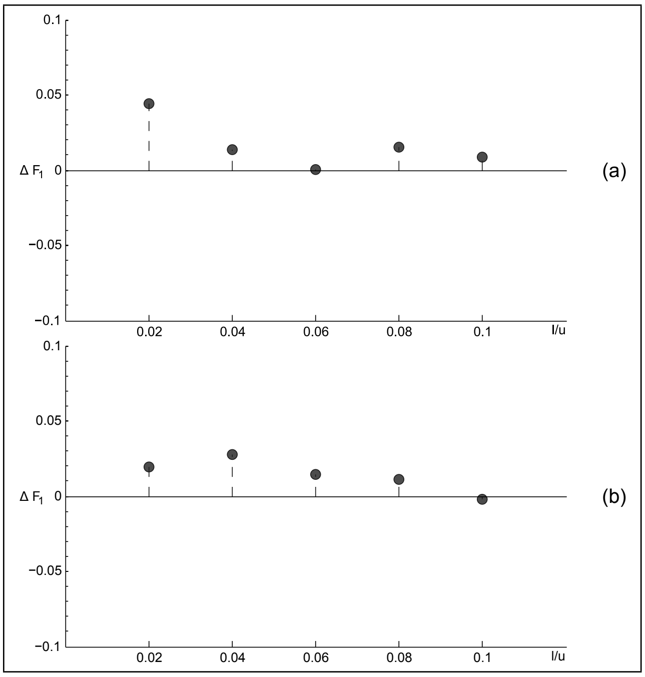

Figure 6: Results of the scenario "Imbalance" on AC (a) and GC (b)

In view of the fact that the availability of labeled data will often be closely related to the cost incurred by querying class labels (e.g., sending a catalog to a potential customer), the overall conclusion emerging from this experiment is that TI is most effective in settings where labeling costs are high. This will commonly be the case in corporate settings, especially if obtaining actual class memberships involves human experts. Exemplary classification tasks include, e.g., an assessor examining insurance claims or a quality manager appraising the functionality of manufactured products. In that respect, TI appears to be a promising alternative to established approaches for business classification.

4.4.2 Distributional Change

Results of the comparison of tSVM and iSVM under distributional change are presented in Figure 7. Here, the advantage of tSVM is even more exposed than in the previous experiment. The magnitude of the difference increases substantially with training data becoming less representative for the cases to classify. The results of a linear regression, $R^2 = 0.902$ and a p-value of the F statistic $< 0.01\%$, verify this finding. Therefore, the results confirm and reemphasize the view that the use of unlabeled data is most beneficial, if - for whatever reason - labeled examples collected in the past no longer reflect the present drivers of class membership. To further clarify upon the dominance of tSVM over iSVM in this experiment, a more detailed view on the empirical results is given in Figure 8. It depicts the raw F_1-Scores of the competing classifiers for AC and their development over the five time points with increasing discrepancy between the training and working set data distributions. Higher values once more indicate better performance.

In the first period, training and working set data are both random samples drawn from AC. Although the training data is thus well representative for the working set, a minor ad-

---

[6] Alternative performance metrics have also been considered, but their effect on the results was found to be negligible, they showed similar tendecies as the used measure. Therefore, the presentation is restricted to the F_1-Score.

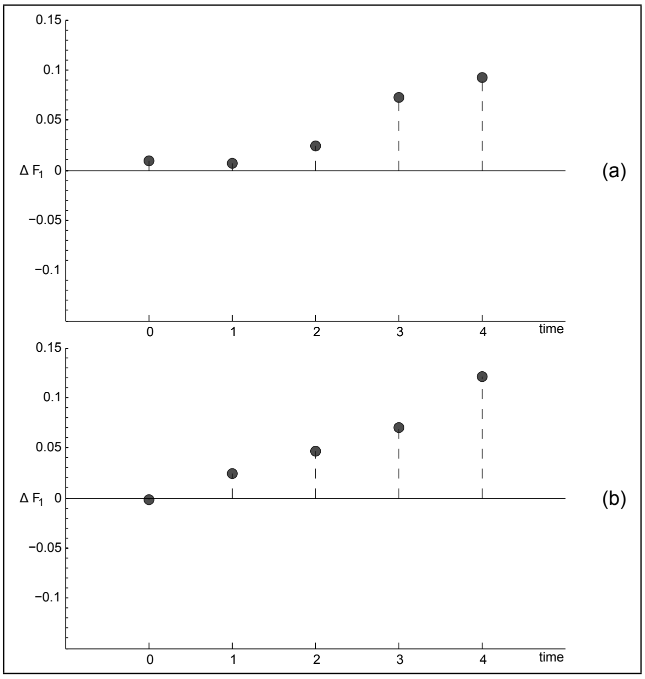

Figure 7: Results of the scenario "Change of distribution" on AC (a) and GC (b)

vantage of tSVM can be observed. This may be taken as empirical confirmation for the view that TI involves solving a simpler problem; i.e., class labels can be estimated in a direct manner if the instances to be classified are known. In the focal case, this approach indeed yields higher predictive accuracy.

Furthermore, Figure 8 illustrates the dramatic decline of iSVM's performance when training data becomes less representative. Although the transductive classifier's access to labeled data is also restricted to outdated training examples, it succeeds in distilling additional information from the unlabeled working set and, thereby, maintains its level of accuracy. For example, in the most extreme setting, the predictive ability of iSVM has declined by 10.85% , whereas tSVM is only 0.88% below its performance peak. Therefore, the results provide strong evidence for TI being indeed highly robust towards changes in data distributions.

5. SUMMARY AND CONCLUSIONS

Classification is a well established approach to support various decision making tasks. Due to the large number of decisions and thus classifications, the accuracy of classification models is commonly considered pivotal in business applications. Therefore, we aimed at exploring the potential of TI, a novel approach towards building predictive classification models, for corporate planning. A key characteristic of the novel paradigm involves a direct estimation of class labels to increase predictions' accuracy and robustness. The theoretical background of this principle has been reviewed and an analysis of the implications of TI's requirement has led to the identification of two factors which govern its applicability in corporate planning: decision scope and decision urgency. In a similar way, the factors imbalance and process stationarity have been proposed as major determinants of TI's expected utility. These four concepts provide a framework for evaluating whether business decision tasks may benefit from an application of TI. An empirical case study has been carried out to complement the valuation of the novel approach. Overall, TI compares favorably to inductive classification and provides more accurate class predictions in most settings. More specifically, although TI's specific requirements in terms of working set availability constrain its applicability in general, TI has been shown to be well competitive if not superior to conventional techniques whenever these requirements are fulfilled. In particular, the performance of tSVM was at least comparable to those of iSVM throughout all experiments, and substantially better in most cases.

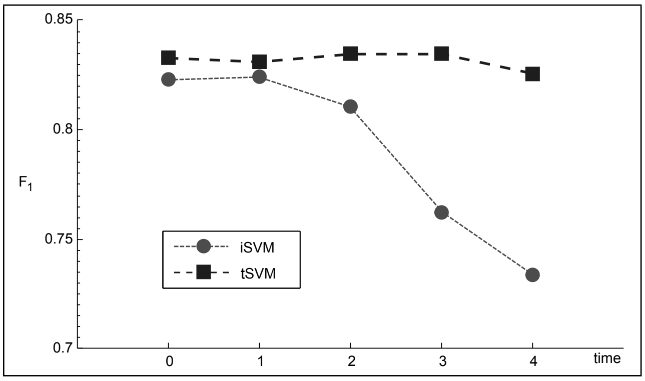

Figure 8: Change of distribution: Developing of F_1-scores

Clearly, empirical findings are restricted to the employed datasets and may not generalize to other applications. Although the data has been drawn from the domain of corporate planning, there is no guarantee that similar results can be observed in, e.g., marketing or manufacturing problems. For example, there is a reasonable risk that the relatively small size of our datasets have granted TI an advantage. This follows directly from the results of the imbalance experiment. Therefore, a careful evaluation of TI in other corporate planing domains is an important area for future research. Our work supports this undertaking by identifying the moderators of TI's applicability and effectiveness and proposing a methodology for appraising a decision task's fit with the novel paradigm. Respective procedures have been discussed in the context of typical corporate applications and concrete examples of promising decision problems have been provided. Moreover, the experimental scenarios developed for TI's evaluation may prove useful in subsequent studies.

On the other hand, there may be no need for being overly pessimistic when appraising the present results' external validity. Drawing inspiration from typical modeling challenges in business classification, all experiments have been carefully designed to assess particular features of TI, which theory would predict to be beneficial. In other words, the encouraging findings of the imbalance and distributional change experiment can well be explained with TI's statistical and mathematical underpinnings.

In view of the overall insights gained during the course of this evaluation, a general conclusion might be that TI represents a novel decision support tool which has the potential to complement or even replace established (i.e., inductive) techniques, if its particular requirements are fulfilled.

However, one may object that this view centers too drastically on technology, rather than decision makers' requirements. Alternatively, TI could be characterized as a planning tool offering higher task-technology fit [24, 25] in specific circumstances (i.e., when seeking class predictions for a known working set of objects) and is, in this sense, preferable - and supposedly superior - to standard methods like inductive classification. In that respect, TI may be considered an example, how advancements in data analytical techniques can and should be geared towards concrete business needs. That is, instead of being forced to match a given decision problem to some standardized procedures (e.g, a standard data mining algorithm for classification, clustering or association), dedicated planning tools should be devised that take account for application specific requirements. This is maybe the most promising avenue for future research in corporate decision support, with TI being a very first step towards the long term objective of a requirement-driven data mining.

6. REFERENCES

[1] E. L. Allwein, R. E. Schapire, and Y. Singer. Reducing multi-class to binary: a unifying approach for margin classifiers. *Journal of Machine Learning Research*, 1:113–141, 2000.

[2] A. Asuncion and D. Newman. UCI Machine Learning Repository, 2007.

[3] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6):627–635, 2003.

[4] J. Banasik and J. Crook. Reject inference, augmentation, and sample selection. *European Journal of Operational Research*, 183:1582–1594, 2007.

[5] U. Bankhofer. Data Mining und seine betriebswirtschaftliche Relevanz. *Betriebswirtschaftliche Forschung und Praxis*, 56:395–412, 2004.

[6] K. P. Bennett and A. Demiriz. Semi-supervised Support Vector Machines. In *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems*, volume 2, pages 368–374, 1999.

[7] M. J. A. Berry and G. Linoff. *Data Mining Techniques: For Marketing, Sales and Customer Relationship Management.* Wiley, New York, 2 edition, 2004.

[8] N. Bissantz and J. Hagedorn. Data Mining (Datenmustererkennung). *Wirtschaftsinformatik*, 51(1):139–144, 2009.

[9] A. Bodapati and S. Gupta. A direct approach to predicting discretized response in target marketing. *Journal of Marketing Research*, 41(1):73–85, 2004.

[10] F. Boylu, H. Aytug, and G. J. Köhler. Induction over strategic agents. *Information Systems Research*, 21(1):170–189, 2010.

[11] R. Caruana and A. Niculescu-Mizil. Data Mining in Metric Space: an Empirical Analysis of Supervised Learning Performance Criteria. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 69–78, New York, 2004.

[12] P. Chamoni and P. Gluchowski. *Analytische Informationssysteme*, pages 3–22. Springer, Berlin Heidelberg, 3 edition, 2006.

[13] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning.* The MIT Press, London, 2006.

[14] O. Chapelle, V., and J. Weston. Transductive inference for estimating values of functions. *Advances in Neural Information Processing Systems*, 12:421–427, 1999.

[15] Y. Chen, G. Wang, and S. Dong. Learning with Progressive Transductive Support Vector Machine. In *International Conference on Data Mining*, pages 67–74, 2002.

[16] V. Cherkassky and Y. Ma. Another look at statistical learning theory and regularization. *Neural Networks*, 22(7):958–969, 2009.

[17] V. Cherkassky and F. M. Mulier. *Learning from Data: Concepts, Theory, and Methods.* Wiley & Sons, New Jersey, 2 edition, 2007.

[18] S. Decherchi, P. Gastaldo, S. Ridella, and R. Zunino. Explicit overall risk minimization transductive bound. *Atlas Conferences*, 2008.

[19] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[20] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, pages 37–45, 1996.

[21] A. Fink, G. Schneidereit, and S. Voß. *Grundlagen der Wirtschaftsinformatik.* Physica-Verlag, Heidelberg, 2 edition, 2005.

[22] P. Gluchowski, R. Gabriel, and C. Dittmar. *Management Support Systeme und Business Intelligence: Computergestützte Informationssysteme für Fach- und Führungskräfte.* Springer, Berlin, 2 edition, 2008.

[23] P. Gluchowski and H.-G. Kemper. Quo vadis business intelligence? *BI-Spektrum*, 1:12–19, 2006.

[24] D. L. Goodhue. Understanding user evaluations of information systems. *Management Science*, 41(12):1827–1844, 1995.

[25] D. L. Goodhue and R. L. Thompson. Task-technology fit and individual performance. *MIS Quarterly*, 19(2):213–236, 1995.

[26] G. A. Gorry and S. Morton. A framework for management information systems. *Sloan Management Review*, 13(1):55–70, 1971.

[27] W. E. Hart, N. Krasnogor, and J. E. Smith. *Recent Advances in Memetic Algorithms.* Springer, Berlin Heidelberg, 1 edition, 2005.

[28] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* Springer, New York, 2009.

[29] J. Hauschild. *Innovationsmanagement.* Vahlen, München, 3 edition, 2004.

[30] A. R. Hevner, S. T. March, J. Park, and S. Ram. Design science in information systems research. *MIS Quarterly*, 28(1):75–105, 2004.

[31] T. Joachims. Transductive Inference for Text Classification Using Support Vector Machines. In *Proceedings of the 16th International Conference on*

Machine Learning, pages 200–209, 1999.

[32] Y. S. Kim, W. N. Street, G. J. Russell, and F. Menczer. Customer targeting: a neural network approach guided by genetic algorithms. *Management Science*, 51(2):264–276, 2005.

[33] J. Z. Kolter and M. A. Maloof. Dynamic weighted majority: A new ensemble method for tracking concept drift. *Journal of Machine Learning Research*, 8:2755–2790, 2007.

[34] L. I. Kuncheva. Classifier Ensembles for Chaning Environments. In F. Roli, J. Kittler, and T. Windeatt, editors, *Proceedings of the 5th International Workshop on Multiple Classifier Systems*, volume 3077 of *Lecture Notes in Computer Science*, pages 1–15, Cagliari, Italy, 2004. Springer.

[35] S. Lessmann and S. Voß. Supervised Classification for Decision Support in Customer Relationship Management. In A. Bortfeldt, J. Homberger, H. Kopfer, and R. G. Pankratz R. Strangmeier, editors, *Intelligent Decision Support*, pages 231–253. Gabler, Wiesbaden, 2008.

[36] S. Lessmann and S. Voß. Unterstützung kundenbezogener Entscheidungsprobleme - Eine Analyse zum Potenzial moderner Klassifikationsverfahren. *Wirtschaftsinformatik*, 52(2):79–93, 2010.

[37] P. Mertens. Geschichte und ausgewählte Gegenwartsprobleme der Wirtschaftsinformatik. *Wirtschaftswissenschaftliches Studium*, 27:170–175, 1998.

[38] P. Mertens. Business Intelligence - Ein Überblick. *Information Management & Consulting*, 22:65–73, 2002.

[39] H. Müller-Merbach. Die ungenutzte Synergie zwischen Operations Research und Wirtschaftsinformatik. *Wirtschaftsinformatik*, 34(3):334–339, 1992.

[40] G. F. Nemet. Demand-pull, technology-push, and government-led incentives for non-incremental technical change. *Research Policy*, 38(5):700–709, 2009.

[41] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason. Defection detection: measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, 43(2):204–211, 2006.

[42] E. W. T. Ngai, L. Xiu, and D. C. K. Chau. Application of data mining techniques in customer relationship management: a literature review and classification. *Expert Systems with Applications*, 36(2, Part 2):2592–2602, 2009.

[43] S. Pang and N. Kasabov. Inductive vs transductive inference, global vs local models: SVM, TSVM, and SVMT for gene expression classification problems. *Neural Networks*, 2:1197–1202, 2004.

[44] F. F. Reichheld and W. Sasser. Zero defections: quality comes to service. *Havard Business Review*, 68(5):105–111, 1990.

[45] M. Scholz and R. Klinkenberg. Boosting classifiers for drifting concepts. *Intelligent Data Analysis*, 11(1):3–28, 2007.

[46] M. M. Silva, T. T. Maia, and A. P. Braga. An Evolutionary Approach to Transduction in Support Vector Machines. In *Proceedings of the 5th International Conference on Hybrid Intelligent Systems*, pages 329–334, 2005.

[47] L. C. Thomas. A survey of credit and behavioral scoring; forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16:149–172, 2000.

[48] L. C. Thomas, J. Crook, and D. Edelman. *Credit Scoring and Its Applications.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.

[49] L. C. Thomas, R. Oliver, and D. J. Hand. A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society*, 56(9):1006–1015, 2005.

[50] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems.* Winston, Washington, D.C., 1977.

[51] K. L. Tsui, V. C. P. Chen, W. Jiang, and Y. A. Aslandogan. Data Mining Methods and Applications. In H. Pham, editor, *Springer Handbook of Engineering Statistics*, pages 651–669. Springer, London, 2006.

[52] B. Twala. Multiple classifier application to credit risk assessment. *Expert Systems with Applications*, 37(4):3326–3336, 2010.

[53] D. Van den Poel and B. Lariviere. Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157(1):196–217, 2004.

[54] V. Vapnik and S. Kotz. *Estimation of Dependences Based on Empirical Data.* Springer, New York, 2 edition, 2006.

[55] G. Wang, J. Hao, J. Ma, and H. Jiang. A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, 2010.

[56] Wissenschaftliche Kommission Wirtschaftsinformatik. Profil der Wirtschaftsinformatik. *Wirtschaftsinformatik*, 36(1):80–81, 1994.