Feature Selection for Wheat Yield Prediction

Georg Ruß, Rudolf Kruse

Abstract Carrying out effective and sustainable agriculture has become an important issue in recent years. Agricultural production has to keep up with an ever-increasing population by taking advantage of a field's heterogeneity. Nowadays, modern technology such as the global positioning system (GPS) and a multitude of developed sensors enable farmers to better measure their fields' heterogeneities. For this small-scale, precise treatment the term *precision agriculture* has been coined. However, the large amounts of data that are (literally) harvested during the growing season have to be analysed. In particular, the farmer is interested in knowing whether a newly developed heterogeneity sensor is potentially advantageous or not. Since the sensor data are readily available, this issue should be seen from an artificial intelligence perspective. There it can be treated as a *feature selection* problem. The additional task of yield prediction can be treated as a multi-dimensional regression problem. This article aims to present an approach towards solving these two practically important problems using artificial intelligence and data mining ideas and methodologies.

Key words: Precision Agriculture, Data Mining, Feature Selection, Regression

1 Introduction

In the recent past, it has become obvious that agriculture is playing a crucial role for sustaining the economy and population growth. In industrialised as well as in developing countries, improvements can be made by introducing modern GPS and advanced sensor technology to make use of a field's heterogeneity. Since this heterogeneity implies a small-scale, precise crop treatment, the term *precision agriculture* has been introduced. According to [34], precision agriculture is the sampling,

Georg Ruß, Rudolf Kruse

Otto-von-Guericke-Univ. Magdeburg, Germany e-mail: {russ,kruse}@iws.cs.uni-magdeburg.de

mapping, analysis and management of production areas that recognises the spatial variability of the cropland.

In artificial intelligence terms, the area of precision agriculture (PA) is quite an interesting one as it involves methods and algorithms from numerous areas that the artificial intelligence community has been dealing with extensively. When analysing the data flow that results from using PA, one is quickly reminded of *data mining*: an agriculturist collects data from his cropland (e.g., when fertilising or harvesting) and would like to extract information from those data and use this information to his (economic) advantage. Obviously, he is also interested in knowing whether a particular sensor which has been introduced will eventually be of use to him in terms of predicting current year's yield precisely.

Two major parts emerge from the above issue: deciding whether a sensor is useful for yield prediction is actually a feature selection task; and the necessary yield prediction turns out to be a multi-dimensional regression problem. Both parts have been studied extensively in AI and numerous approaches exist.

1.1 Research Target and Article Structure

With this contribution we aim at developing a suitable approach to evaluate sensor data. We are interested in identifying those sensors which are most applicable for measuring a field's heterogeneity. Here, a *good* sensor is supposed to improve the precision of yield prediction. Eventually, a feature selection approach shall be developed. Since existing work has mainly been devoted to using feature selection with a classification task, but not a regression task, this work is aimed at evaluating regression approaches. Furthermore, some more research will be devoted to finding a generally applicable regression model which is to be used in the feature selection approach.

In two ways, the feature selection approach taken here is novel: first, the application area of precision agriculture data is certainly a new one. Second, feature selection has mainly been used for classification tasks instead of regression tasks.

Our previous work in this area has been focused on either evaluating regression models ([26], [27], [29]) or visualising the existing agriculture data [28]. Results from these articles will be incorporated into our feature selection approach where appropriate.

After a brief overview of the available data in Section 2, our feature selection approach will be presented in Section 3. The experimental layout and the results are presented in Section 4. In the end, a conclusion is given and future work is pointed out.

2 Data Description

The data available in this work have been obtained in the years 2003–2006 on three fields near Köthen, north of Halle, Germany[1]. All information available for these 65-, 72- and 32-hectare fields[2] was interpolated using kriging [32] to a grid with 10 by 10 meters grid cell sizes. Each grid cell represents a record with all available information. During the growing season of 2006, the latter field was subdivided into different strips, where various fertilization strategies were carried out. For an example of various managing strategies, see e.g. [30], which also shows the economic potential of PA technologies quite clearly. The field grew winter wheat, where nitrogen fertilizer was distributed over three application times during the growing season.

Overall, for each field there are seven input attributes – accompanied by the respective current year's yield (2004 or 2006) as the target attribute. Those attributes will be described in the following. In total, for the F04 field there are 5241 records, for F131 there are 2278 records, for F330 there are 4578 records, thereof none with missing values and none with outliers.

2.1 Nitrogen Fertilizer – N1 , N2 , N3

The amount of fertilizer applied to each subfield can be easily measured. It is applied at three points in time into the vegetation period, which is the standard strategy for most of Northwest Europe [22].

2.2 Vegetation – REIP32 , REIP49

The *red edge inflection point* (REIP) is a second derivative value calculated along the red edge region of the spectrum, which is situated from 680 to 750nm. Dedicated REIP sensors are used in-season to measure the plants' reflection in this spectral band. Since the plants' chlorophyll content is assumed to highly correlate with the nitrogen availability (see, e.g. [20]), the REIP value allows for deducing the plants' state of nutrition and thus, the previous crop growth. For further information on certain types of sensors and a more detailed introduction, see [16] or [33]. Plants that have less chlorophyll will show a lower REIP value as the red edge moves toward the blue part of the spectrum. On the other hand, plants with more chlorophyll will have higher REIP values as the red edge moves toward the higher wavelengths. For the range of REIP values encountered in the available data, see Tables 1(b) and 1(c). The numbers in the REIP32 and REIP49 names refer to the growing stage of winter wheat, as defined in [18].

---

[1] GPS: Latitude N 51 40.430, Longitude E 11 58.110

[2] We will call them *F04*, *F330* and *F131*, respectively

2.3 Electric Conductivity – EM38

A non-invasive method to discover and map a field's heterogeneity is to measure the soil's conductivity. Commercial sensors such as the EM-38[3] are designed for agricultural use and can measure small-scale conductivity to a depth of about 1.5 metres. There is no possibility of interpreting these sensor data directly in terms of its meaningfulness as yield-influencing factor. But in connection with other site-specific data, as explained in the rest of this section, there could be coherences. For a more detailed analysis of this particular sensor, see, e.g. [6]. For the range of EM values encountered in the available data, see Tables 1(a) to 1(c).

2.4 YIELD

Here, yield is measured in metric tons per hectare ($\frac{t}{ha}$) For the yield ranges for the respective years and sites, see Tables 1(b) and 1(c). It should be noted that for the F131 and F330 data sets the yield was reduced significantly due to bad weather conditions (lack of rain) during the growing season 2006.

2.5 TRACFORCE

The tractive force (or tractive effort) sensor measures the force that has to be exerted when a plough or similar (modern) devices are pulled along the field. Hence, it is assumed that the upper soil layer and its condition may have an effect on the final yield. This sensor is only available in the F04 data set.

2.6 Data Overview

In this work, we evaluate data sets from three different fields. A brief summary of the available data attributes for both data sets is given in Tables 1(a) to 1(c).

2.7 Points of Interest

It would be interesting to see how much the influencable factor "fertilization" really influences the yield in the current site-year. Furthermore, at the core of this article, we are interested in finding out which sensor data are actually useful for the purpose

---

[3] trademark of Geonics Ltd, Ontario, Canada

Table 1: Overview of the F04, F131 and F330 data sets

(a) Data overview, F04

F04	*min*	*max*	*mean*	*std*
YIELD03	1.19	12.38	6.27	1.48
EM38	17.97	86.45	33.82	5.27
N1	0	100	57.7	13.5
N2	0	100	39.9	16.4
N3	0	100	38.5	15.3
REIP32	721.1	727.2	725.7	0.64
REIP49	722.4	729.6	728.1	0.65
YIELD04	6.42	11.37	9.14	0.73

(b) Data overview, F131

F131	*min*	*max*	*mean*	*std*
YIELD05	1.69	10.68	5.69	0.93
EM38	51.58	84.08	62.21	8.60
N1	47.70	70	64.32	6.02
N2	14.80	100	51.71	15.67
N3	0	70	39.65	13.73
REIP32	719.6	724.4	722.6	0.69
REIP49	722.3	727.9	725.8	0.95
YIELD06	1.54	8.83	5.21	0.88

(c) Data overview, F330

F330	*min*	*max*	*mean*	*std*
YIELD05	4.64	14.12	10.62	0.97
EM38	25.08	49.48	33.69	2.94
N1	24.0	70	59.48	14.42
N2	3.0	100	56.38	13.35
N3	0.3	91.6	50.05	12.12
REIP32	719.2	724.4	721.5	1.03
REIP49	723.0	728.5	726.9	0.82
YIELD06	1.84	8.27	5.90	0.54

of yield prediction. There may be data attributes which are irrelevant and others which may be highly relevant. This *feature selection* problem will be described in the following section.

3 Feature Selection Approach

This section deals with the developed feature selection approach. First, according to the literature, some decisions have to be made regarding the type and general structure of feature selection. The ensuing section presents a suitable algorithm, of which the details will be presented. The feature selection algorithm incorporates a regression task, therefore regression models will be presented. Additionally, the details of error measurement are shown.

3.1 Approach Justification

In the data encountered here, the main reason for applying data mining techniques is that the interesting features are hidden in a search space of high dimensionality.

Reducing the dimensionality of the data could be done via sampling or otherwise discarding data records, but in our case this would render the precision agriculture approach useless. Therefore, the actual features of the data should be condensed to those that are most promising for a yield prediction task. Hence, feature selection eventually is a means of removing irrelevant and/or redundant features.

According to [15], the selection of features can be achieved in two ways. The first one is to evaluate and rank features according to some criterion. The other one is to select a minimal subset of features without deteriorating learning performance. The latter approach can usually be run automatically, while the first one provides reasonable guidance to an expert user. Since the data analysis task in this work is not aimed at the average user, but should preferably contribute insights into the data for professionals, the "evaluate-and-rank"-approach will be pursued in the following.

Once we have decided for evaluating and subsequently ranking features, there are still a multitude of options to select the most promising features. There are quite a few surveys which aim to categorise existing approaches, such as [1], [2], [8], [14]. One of the categories is usually how the search for features is conducted: *forward selection* starts with an empty set and keeps adding features until a stopping criterion is met. *Backward elimination* starts with a set including all features and subsequently keeps removing features. Obviously, both cases could also yield a ranked list of features. Since both approaches have their advantages and drawbacks, but backward elimination is computationally heavier, we decided to employ forward selection and leave backward elimination for future work.

A second category is how the feature space is traversed in the search for features to exclude or include. Doing a complete search is the straightforward option, but should usually be avoided due to computational constraints. Heuristic search may miss optimal subsets, but generates good solutions quickly with a certain probability depending on the heuristic used. Non-deterministic search randomly explores the feature space, obviously limited by the available computational resources. Since this work is the first to evaluate whether feature selection may be a successful approach to decide which agricultural sensor data are useful, we simply employ a complete search. This is possible since the number of features in the available data sets is small. In future work, different search strategies will be evaluated.

A third decision to be made is whether to use a *filter* or *wrapper* approach. The earlier one does not depend on the actual induction (regression) algorithm for evaluating the generated subsets. On the contrary, the wrapper approach explicitly uses the induction method (e.g. a regression tree) for evaluating the subset. For a more detailed explanation, we refer to e.g. [13]. In the following work, we consider the wrapper approach because it also enables a comparison of the used regression techniques. A similar approach is taken by, e.g. [12].

3.2 Selection and Regression Approach

As mentioned in the previous section, we decided to use forward selection. A possible implementation is given in Algorithm 1. The algorithm works in a straightforward way: starting with an empty list of features *S*, and a list of all features *F*, it repeatedly moves the *best* features for the regression task from *F* to *S*. This series of steps is repeated until a regression error goal is met or *F* is empty.

Algorithm 1 *forward selection* of features

1: $S = [\,]$, $F \leftarrow$ features
2: **repeat**
3:    $E \leftarrow [\,]$
4:    **for** $j = 1\ldots length(F)$ **do**
5:       $f \leftarrow F[j]$ {select j-th feature}
6:       $S_j \leftarrow S \cup f$ {add current feature to S_j}
7:       $M_j \leftarrow model(S_j)$ {generate regression model from data}
8:       $E_j \leftarrow evaluate(M_j)$ {calculate modeling error}
9:       $E \leftarrow E || E_j$ {store error}
10:   **end for**
11:   $S \leftarrow S || F[min(E)]$ {add best feature to S}
12:   $F \leftarrow F - F[min(E)]$ {remove best feature from F}
13: **until** min(E) $\leq$ threshold OR $F = [\,]$
14: **return** S {return list of features, best one first}

3.3 Regression Modeling

This subsection serves as an overview of different regression techniques which may be used on the agriculture data. A more thorough description can be found in, e.g., [26]. Since the focus here is on feature selection, two of the four models presented in [26] are chosen, namely regression trees and support vector regression. On the one hand, the latter technique turned out to be the one that performed best on different data sets. On the other hand, regression trees convey a lot of understandability which might yield further insights into the data, hence they have been chosen here.

The regression task can be formalized as follows: the training set

$$T = \{\{x_1,\ldots,x_n\}, y_i\}_{i=1}^{N} \tag{1}$$

is considered for the training process, where $x_i, i = 1,\ldots,n$ are continuous input values and $y_i, i = 1\ldots,N$ are continuous output values. Given this training set, the task of the regression techniques is to approximate the underlying function sufficiently well. The quality of the approximated function can be measured by error values.

Regression trees have seen some usage in agriculture [7, 11, 17]. Essentially, they are a special case of decision trees where the outcome (in the tree leaves) is a continuous function instead of a discrete classification. Further details can be found in section 3.3.1.

A technique that has, to the best of our knowledge, not been used on similar yield data, but for similar regression tasks, is a derivative of support vector machines (SVMs). Similar to decision trees, if the target attribute is discrete, SVMs would solve a classification task, whereas in the case of a continuous attribute, a regression task would be solved. Hence, support vector regression (SVR) will be explained in section 3.3.2.

3.3.1 Regression Tree

Regression trees approximate learning instances by sorting them down the tree from the root to some leaf node, which provides the value of the target attribute. Each node in the tree represents a split of some attribute of the instance and each branch descending from that node corresponds to one part left or right of the split. The value of the target attribute for an instance is determined by starting at the root node of the tree and testing the attribute specified by this node. This determines whether to proceed left or right of the split. Then we move down the tree and repeat the procedure with the respective subtree. In principle, there could be more than one split in a tree node, which would result in more than two subtrees per node. However, in this application scenario, we do not consider regression trees with more than two subtrees per split node.

Regression as well as decision trees are usually constructed in a top-down, greedy search approach through the space of possible trees [21]. The basic algorithms for constructing such trees are CART [4], ID3 [23] and its successor C4.5 [24]. The idea here is to ask the question "which attribute should be tested at the top of the tree?" To answer this question, each attribute is evaluated to determine how well it is suited to split the data. The best attribute is selected and used as the test node. This procedure is repeated for the subtrees. An attribute selection criterion that is employed by ID3 and C4.5 is the entropy and, resulting from it, the information gain. Entropy is a measure from information theory that describes the variety in a collection of data points: the higher the entropy, the higher the variety. In an attribute split we would like to lower the entropy of the two resulting split data sets. This reduction in entropy is called the information gain. For further information we refer to [21].

However, if the addition of nodes is continued without a specific stopping criterion, the depth of the tree continues to grow until each tree leaf covers one instance of the training data set. This is certainly a perfect tree for the training data but is likely to be too specific – the problem of overlearning occurs. For new, unseen data, such a specific tree will probably have a high prediction error. Therefore, regression trees are usually pruned to a specific depth which is a trade-off between high accuracy and high generality. This can easily be achieved by setting a lower bound

for the number of instances covered by a single node below which no split should occur. For our work we used the R implementation of `rpart`.

3.3.2 Support Vector Regression

Support Vector Machines (SVMs) are a supervised learning method discovered by [3]. However, the task here is regression, so we focus on support vector regression (SVR) in the following. A more in-depth discussion can be found in [9]. Given the training set, the goal of SVR is to approximate a linear function $f(x) = \langle w, x \rangle + b$ with $w \in \mathbb{R}^N$ and $b \in \mathbb{R}$. This function minimizes an empirical risk function defined as

$$R_{emp} = \frac{1}{N} \sum_{i=1}^{N} L_{\varepsilon}(\hat{y} - f(x)), \tag{2}$$

where $L_{\varepsilon}(\hat{y} - f(x)) = \max((|\xi| - \varepsilon), 0)$. $|\xi|$ is the so-called slack variable, which has mainly been introduced to deal with otherwise infeasible constraints of the optimization problem, as has been mentioned in [31]. By using this variable, errors are basically ignored as long as they are smaller than a properly selected ε. The function here is called ε-insensitive loss function. Other kinds of functions can be used, some of which are presented in chapter 5 of [9].

To estimate $f(x)$, a quadratic problem must be solved, of which the dual form, according to [19] is as follows:

$$max_{\alpha, \alpha^*} - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \varepsilon \sum_{i=j}^{N} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \tag{3}$$

with the constraint that $\sum_{j=1}^{N}(\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C]$. The regularization parameter $C > 0$ determines the tradeoff between the flatness of $f(x)$ and the allowed number of points with deviations larger than ε. As mentioned in [9], the value of ε is inversely proportional to the number of support vectors. An adequate setting of C and ε is necessary for a suitable solution to the regression problem.

Furthermore, $K(x_i, x_j)$ is known as a kernel function which allows to project the original data into a higher-dimensional feature space where it is much more likely to be linearly separable. Some of the most popular kernels are radial basis functions (equation 4) and a polynomial kernel (equation 5):

$$K(x, x_i) = e^{-\frac{||x - x_i||^2}{2\sigma^2}} \tag{4}$$

$$K(x, x_i) = (\langle x, x_i \rangle + 1)^{\rho} \tag{5}$$

The parameters σ and ρ have to be determined appropriately for the SVM to generalize well. This is usually done experimentally. Once the solution for the above

optimization problem in equation 3 is obtained, the support vectors can be used to construct the regression function:

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x, x_i) + b \qquad (6)$$

In our experiments, we used the R SVM interface to libsvm [5] with the `e1071` R package.

3.4 Performance Measurement

The performance of the models will be determined using the root mean squared error (RMSE). For the RMSE, first the difference between an actual target value y_a and the model output value y is computed. This difference is squared and averaged over all training examples before the root of the mean value is taken, see equation 7.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=j}^{n} (y_i - y_{a,i})^2} \qquad (7)$$

In [26] it has been established that model parameters set for one data set may be carried over to a different data set. Hence, it is assumed that the RegTree and SVR parameters may be fixed. Nevertheless, although computationally heavy, a *complete search* feature selection strategy with an automatic fine-tuning of model parameters for the SVR and RegTree model has been used.

For training the models, a cross validation approach is taken. As mentioned in e.g. [10], the data will be split randomly into a training set and a test set. The model is trained using the training data and after each training iteration, the error on the test data is computed. During training, this error usually declines towards a minimum. Beyond this minimum, the error rises – overlearning (or overfitting) occurs: the model fits the training data perfectly but does not generalize well. Hence, the model training is stopped when the error on the test set starts rising. We use a size ratio of 9:1 for training and test set. The data sets are partitioned randomly 20 times, the models are trained and the error values are collected.

4 Experimental Results

The three data sets presented in Section 2 are evaluated here for the purpose of establishing whether a particular feature (sensor) is useful for yield prediction. The target attribute to predict via regression techniques is the respective current year's yield.

Two techniques have been selected: regression trees and support vector regression. 10-fold cross validation was applied for both techniques.

4.1 Feature Selection with RegTree

Regression trees were created using an auto-prune approach, which is similar to the grid search in the following section. A minimum split value of 30 is used, therefore leaves in the final tree will not contain more than 30 data observations for which the yield shall be predicted. An additional complexity parameter $cp = 0.001$ was set, which determines that a split must at least yield a prediction improvement of cp, otherwise it is not carried out.

As can be seen in Table 2, the predictive error of the regression tree decreases steadily as more predictors are added. This behaviour should be expected, since the regression tree receives more information to improve its predictive performance. However, after adding three or four features, the error seems almost to have leveled, which would imply that the remaining features are unnecessary for yield prediction. Generally, the REIP49 feature seems to be the one best suited for prediction. This should also be expected, since the measurement of REIP49 occurs late into the growing season when the crop is nearing harvest. In two out of the three data sets, the respective previous year's yield also seems to play an important role. Again, this should be expected, since areas with high/low yield are (without special treatment) bound to generate high/low yield in the following years as well. The EM38 measurement ranks third or fourth for the prediction task, even better than the REIP32 feature, which is a novel and important result. Even more useful might be the conclusion that N1 and N2 , the dressings of fertilizer early into the growing season, do not bear significant amounts of information in regard to yield prediction within this feature selection approach.

Table 2: Results for Feature Selection via Regression Tree

step	F04 error	feature	F131 error	feature	F330 error	feature
1	0.342	YIELD03	0.235	REIP49	0.279	N3
2	0.262	REIP49	0.136	YIELD05	0.246	REIP49
3	0.228	N3	0.104	EM38	0.223	EM38
4	0.215	EM38	0.104	REIP32	0.199	REIP32
5	0.210	REIP32	0.104	N1	0.189	N1
6	0.209	TRACFORCE	0.104	N2	0.187	N2
7	0.208	N1	0.104	N3	0.181	YIELD05
8	0.205	N2				

4.2 Feature Selection with SVR

For support vector regression, a grid search in the parameter space within reasonable bounds has been performed when searching for the optimal regression parameters, using the `svm.best()` R routine. The regularization parameter C was set to vary between 2^3 and 2^7, the ε parameter between 0.5 and 2, while a radial kernel was used.

Table 3 shows the results for SVR. In each feature selection step the best model from this particular step was chosen. Due to the *complete search* approach, the prediction accuracy may decline from one step to the next. Nevertheless, SVR agrees with the RegTree that the REIP49 feature is the one bearing most predictive power since it is chosen in the first or second step. It also agrees on the predictive power in the respective previous year's yield, which ranks high in F04 and F131, but low in F330. N1 and N2 do not improve the prediction significantly. In addition, the overall prediction error levels are higher than with the simple regression tree. This might be due to the grid search approach for selecting the best SVR model in each step. This grid search optimized the most important SVR parameters, while more may be chosen. Nevertheless, SVR tends to produce comparable results.

Table 3: Results for Feature Selection via Support Vector Regression

step	F04 error	feature	F131 error	feature	F330 error	feature
1	0.557	YIELD03	0.469	REIP49	0.491	N3
2	0.509	REIP49	0.356	YIELD05	0.493	REIP49
3	0.466	N2	0.337	N2	0.469	EM38
4	0.483	REIP32	0.335	N1	0.519	N2
5	0.449	N1	0.303	REIP32	0.454	N1
6	0.471	EM38	0.333	N3	0.508	YIELD05
7	0.444	N3	0.285	EM38	0.439	REIP32
8	0.469	TRACFORCE				

5 Conclusion

In this paper we presented a novel application of a feature selection approach by using it on agriculture data. We were interested in evaluating the data attributes with regard to their utility for yield prediction. The presented approach employs *forward feature selection* and a complete search strategy. With it, two regression models (SVR and RegTree) were used to compare the yield prediction results for the different data sets. Both the SVR and RegTree regression models produced slightly different, but yet comparable results. Nevertheless, on the one hand both models returned understandable and explicable feature rankings, while, on the other hand, provid-

ing novel knowledge about the data sets and their features. Hence, the presented feature selection approach presents an immediately useful application of artificial intelligence techniques and may be developed further.

5.1 Future Work

Future work will focus on evaluating additional feature selection approaches like the reverse process of *backward elimination*. Additional regression models are to be included. Furthermore, different approaches to judging the importance of certain features may be used, such as principal components analysis or clustering.

References

1. A. Arauzo-Azofra and J. M. Benitez. Empirical study of feature selection methods in classification. In *Hybrid Intelligent Systems, 2008. HIS '08. Eighth International Conference on*, pages 584–589, 2008.
2. Avrim L. Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97:245–271, 1997.
3. Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press, 1992.
4. L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.
5. Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.
6. D. L. Corwin and S. M. Lesch. Application of soil electrical conductivity to precision agriculture: Theory, principles, and guidelines. *Agron J*, 95(3):455–471, May 2003.
7. S. F. Crone, S. Lessmann, and S. Pietsch. Forecasting with computational intelligence - an evaluation of support vector regression and artificial neural networks for time series prediction. In *Neural Networks, 2006. IJCNN '06. International Joint Conference on*, pages 3159–3166, 2006.
8. M. Dash and H. Liu. Feature selection for classification. *Intelligent Data Analysis*, 1:131–156, 1997.
9. S.R. Gunn. Support vector machines for classification and regression. Technical Report, School of Electronics and Computer Science, University of Southampton, Southampton, U.K., 1998.
10. Robert Hecht-Nielsen. *Neurocomputing*. Addison-Wesley, September 1990.
11. Chengquan Huang, Limin Yang, Bruce Wylie, and Collin Homer. A strategy for estimating tree canopy density using landsat 7 etm+ and high resolution images over large areas. In *Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry*, 2001.
12. M. Karagiannopoulos, D. Anyfantis, S. B. Kotsiantis, and P. E. Pintelas. Feature selection for regression problems. In *Proceedings of HERCMA'07*. Athens University of Economics and Business, September 2007.

13. Pat Langley. Selection of relevant features in machine learning. In *In Proceedings of the AAAI Fall symposium on relevance*, pages 140–144. AAAI Press, 1994.
14. H. Liu and L. Yu. Feature selection for data mining. Technical report, Arizona State University, 2002.
15. Huan Liu and Hiroshi Motoda, editors. *Computational Methods of Feature Selection*. Data Mining and Knowledge Discovery. Chapman & Hall/CRC, October 2007.
16. J. Liu, J. R. Miller, D. Haboudane, and E. Pattey. Exploring the relationship between red edge parameters and crop variables for precision agriculture. In *2004 IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1276–1279, 2004.
17. David B. Lobell, J. Ivan Ortiz-Monasterio, Gregory P. Asner, Rosamond L. Naylor, and Walter P. Falcon. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. *Agronomy Journal*, 97:241–249, 2005.
18. U. Meier. *Entwicklungsstadien mono- und dikotyler Pflanzen*. Biologische Bundesanstalt fr Land- und Forstwirtschaft, Braunschweig, Germany, 2001.
19. Iván Mejía-Guevara and Ángel Kuri-Morales. Evolutionary feature and parameter selection in support vector regression. In *Lecture Notes in Computer Science*, volume 4827, pages 399–408. Springer, Berlin, Heidelberg, 2007.
20. E. M. Middleton, P. K. E. Campbell, J. E. Mcmurtrey, L. A. Corp, L. M. Butcher, and E. W. Chappelle. "Red edge" optical properties of corn leaves from different nitrogen regimes. In *2002 IEEE International Geoscience and Remote Sensing Symposium*, volume 4, pages 2208–2210, 2002.
21. Tom M. Mitchell. *Machine Learning*. McGraw-Hill Science/Engineering/Math, March 1997.
22. Jacques J. Neeteson. *Nitrogen Management for Intensively Grown Arable Crops and Field Vegetables*, chapter 7, pages 295–326. CRC Press, Haren, The Netherlands, 1995.
23. J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, March 1986.
24. Ross J. Quinlan. *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*. Morgan Kaufmann, January 1993.
25. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.
26. Georg Ruß. Data mining of agricultural yield data: A comparison of regression models. In Petra Perner, editor, *Advances in Data Mining*, Lecture Notes in Computer Science, pages –. Springer, July 2009. (to appear).
27. Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Optimizing wheat yield prediction using different topologies of neural networks. In José Luis Verdegay, Manuel Ojeda-Aciego, and Luis Magdalena, editors, *Proceedings of IPMU-08*, pages 576–582. University of Málaga, June 2008.
28. Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Visualization of agriculture data using self-organizing maps. In Tony Allen, Richard Ellis, and Miltos Petridis, editors, *Applications and Innovations in Intelligent Systems*, volume 16 of *Proceedings of AI-2008*, pages 47–60. BCS SGAI, Springer, January 2009.
29. Georg Ruß, Rudolf Kruse, Peter Wagner, and Martin Schneider. Data mining with neural networks for wheat yield prediction. In Petra Perner, editor, *Advances in Data Mining (Proc. ICDM 2008)*, pages 47–56, Berlin, Heidelberg, July 2008. Springer Verlag.
30. M. Schneider and P. Wagner. Prerequisites for the adoption of new technologies - the example of precision agriculture. In *Agricultural Engineering for a Better World*, Düsseldorf, 2006. VDI Verlag GmbH.
31. Alex J. Smola and Bernhard Sch Olkopf. A tutorial on support vector regression. Technical report, Statistics and Computing, 1998.
32. Michael L. Stein. *Interpolation of Spatial Data : Some Theory for Kriging (Springer Series in Statistics)*. Springer, June 1999.
33. Georg Weigert. *Data Mining und Wissensentdeckung im Precision Farming - Entwicklung von ökonomisch optimierten Entscheidungsregeln zur kleinräumigen Stickstoff-Ausbringung*. PhD thesis, TU München, 2006.
34. Michael D. Weiss. Precision farming and spatial economic analysis: Research challenges and opportunities. *American Journal of Agricultural Economics*, 78(5):1275–1280, 1996.