

WORKSHOP 2

Fehler- und Gefahrenanalysen in landwirtschaftlichen Betrieben

Katharina Dahlhoff (Universität Bonn)

„Beratung von Milchvieh haltenden Betrieben auf der Grundlage von Verhaltens- und
Erscheinungsparametern ihrer Milchkühe“

Tilman Wilke (GIQS e.V., Universität Bonn)

„fmea3d – Implementierung einer alternativen RPZ Berechnungsmethode in R“

Jan Behmann (Universität Bonn, Institut für Geodäsie und Geoinformation)

„Recognition of Activity States in Dairy Cows with SVMs and Graphical Models“

Daniel Martini (Kuratorium für Technik und Bauwesen in der Landwirtschaft KTBL)

„Planungsdaten schnell finden und einfach nutzen: Linked Open Data und semantische
Suche im Einsatz für das KTBL-Datenangebot“

Prof. Wolfgang Büscher (Universität Bonn, Institut für Landtechnik)

Moderation und Diskussionsleitung

Zusammenfassung: Im Management und der Beratung landwirtschaftlicher Betriebe, spielen Fehler- und Gefahrenanalysen eine wichtige Rolle. Die besondere Herausforderung liegt darin, im landwirtschaftlichen Bereich vorhandenes Erfahrungswissen und Wissensbestände in praxistaugliche, vernetzte, mobile, selbstlernende, intelligente Management-Support-Tools zu verwandeln. Andererseits geht es auch darum, bewährte Management-Werkzeuge (wie zum Beispiel die FMEA) für neue Anwendungsfälle zu erschließen oder leichter verfügbar zu machen.

Beratung von Milchvieh haltenden Betrieben auf der Grundlage der Verhaltens- und Erscheinungsparameter ihrer Milchkühe

Katharina Dahlhoff, Wolfgang Büscher, Andreas Pelzer

Institut für Landtechnik, Universität Bonn
Verfahrenstechnik in der tierischen Erzeugung
Nußallee 5
53115 Bonn
katharina.dahlhoff@gmx.de
buescher@uni-bonn.de

Abstract: Zur standardisierten Beurteilung der Haltungsbedingungen in Liegeboxenlaufställen für Milchkühe wurde eine Schwachstellenanalyse („On-Farm Welfare Assessment“) entwickelt, die mit Hilfe tierbezogener Indikatoren Schwachstellen in den Bereichen Haltung und Management aufdeckt. Dabei basiert die Schwachstellenanalyse auf einem Expertensystem, das den Vergleich des Einzelbetriebes mit definierten Referenzwerten, die Identifizierung der haltungsbedingten Risikofaktoren sowie die Ableitung von standardisierten Beratungsempfehlungen zur Verbesserung der Tiergerechtheit vorsieht. Die Auswertung der erfassten Daten erfolgt digital auf dem Betrieb und liefert somit schnelle und anschauliche Beratungsergebnisse.

1 Einführung

Die produktionstechnische Beratung von Milchviehbetrieben hat zum Ziel, die in vielen Ställen häufig noch ungenutzten Reserven in den Bereichen Haltung und Management zu mobilisieren und Defizite aufzudecken. Dabei steht eine gezielte Ausrichtung der Haltungsbedingungen auf die natürlichen Bedürfnisse der Kühe zur Sicherstellung der Tiergerechtheit im Vordergrund.

Um eine standardisierte Bewertung der Haltungsbedingungen zu ermöglichen, wurde eine Schwachstellenanalyse („On-Farm Welfare Assessment“) für Liegeboxenlaufställe entwickelt, die auf der objektiven und systematischen Erfassung ausgewählter tierbezogener Parameter basiert [PE07].

2 Inhalte und Ablauf der Schwachstellenanalyse

Als Grundlage für das Expertensystem wurde in einer Felderhebung ein umfangreicher Datenpool angelegt. Dieser enthielt neben den in Abbildung 1 aufgeführten

tierbezogenen Parametern zum Verhalten der Kühe sowie deren Erscheinung auch umfassende bautechnische und managementspezifische Kriterien von 66 nordrhein-westfälischen Milchviehbetrieben. Durch die Integration verschiedener direkter und indirekter Kriterienbereiche soll eine vollständige und multifaktorielle Betrachtungsweise sichergestellt werden [Su98].

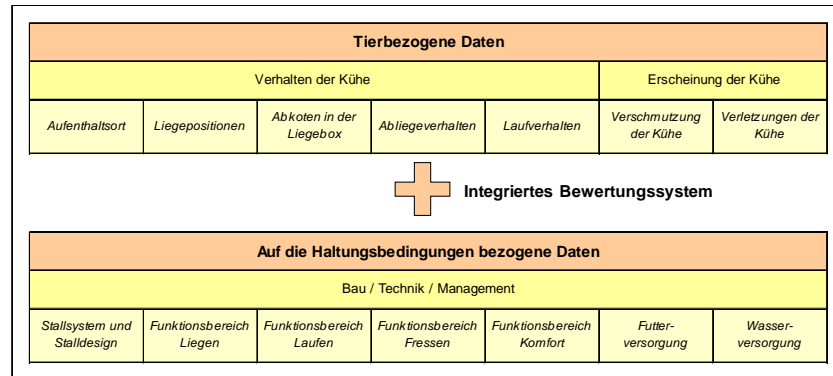


Abbildung 1: In der Schwachstellenanalyse verwendete Kriterienbereiche

Wie in Abbildung 2 dargestellt, wird die einzelbetriebliche Schwachstellenanalyse als dreistufiger Prozess durchgeführt [Da09; Rü10].

Im ersten Schritt erfolgt ein Vergleich der einzelbetrieblichen, tierbezogenen Untersuchungsergebnisse mit definierten Vergleichs- und Referenzwerten (Ziel-, Richt- und Grenzwerte) aus dem Expertensystem. Dieses Benchmarking zielt darauf ab, konkret auftretende (Verhaltens-) Beeinträchtigungen der Kühe zu erkennen.

Im zweiten Schritt werden dann die potentiellen Risikofaktoren, die zu diesen haltungsbedingten Beeinträchtigungen der Kühe führen können, identifiziert. Dazu wurden auf Grundlage des Datenpools Zusammenhänge zwischen den Tieren, der Haltungsumwelt sowie dem Management ermittelt. Aufgrund des hohen Stellenwertes des Liegebereichs für die Kühe konnten vor allem die negativen Auswirkungen einer suboptimalen Liegeboxengestaltung auf das (Liege-) Verhalten und die Verletzungshäufigkeiten der Kühe herausgestellt werden, die häufig durch unzureichende Boxen- und Nackenriegelabmessungen, ungünstige Seitenabtrennungen sowie schlechte Liegeflächenqualitäten hervorgerufen werden.

Der dritte Schritt der Schwachstellenanalyse sieht schließlich eine Korrektur der bemängelten Kriterien der Haltungsbedingungen im Sinne der Tiergerechtigkeit vor [Wh03]. In einer stichprobenartigen Nachuntersuchung wurden in den beiden geprüften Betrieben nach Umsetzung der getätigten Beratungsempfehlungen wesentliche Verbesserungen bei den erfassten tierbezogenen Indikatoren festgestellt.

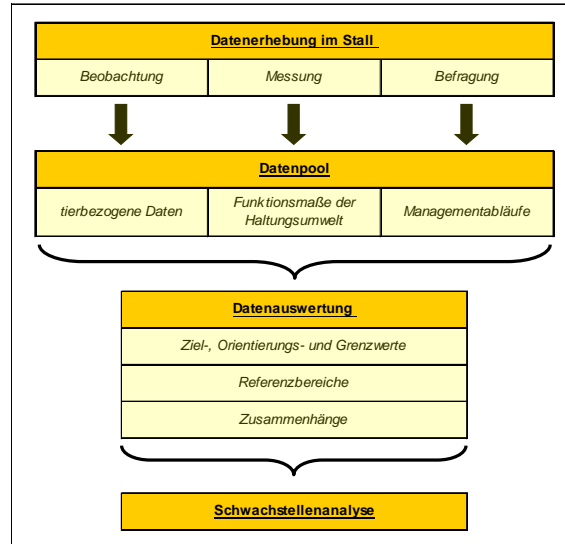


Abbildung 2: Arbeitsschritte der Schwachstellenanalyse [Da09]

Die Erhebung der Indikatoren erfolgt im Stallbereich mittels eines Eingabegeräts digital. Die eingegebenen Daten werden von der Auswertungssoftware direkt ausgewertet. Die Ergebnisse des Benchmarkings werden dabei in Bezug auf die Referenzwerte des Expertensystems grafisch dargestellt und sind somit leicht nachvollziehbar. Zudem werden die Zusammenhänge zwischen den Verhaltens- sowie Erscheinungsmerkmalen und den Haltungsbedingungen aufgezeigt und standardisierte Beratungsaussagen vorgeschlagen.

3 Evaluierung der Schwachstellenanalyse

Zur Evaluierung der Schwachstellenanalyse wurden die tierbezogenen Parameter hinsichtlich der Gütekriterien der Validität, Objektivierbarkeit, Reliabilität sowie Praktikabilität überprüft. Des Weiteren wurden zur Qualitätssicherung in einer Voruntersuchung statistische Auswertungen zu den Reliabilitäten der verwendeten Bewertungsschemata (Hygienescore, Integumentschäden) durchgeführt. Nach abschließender Bewertung stellten sich vor allem die verschiedenen Indikatoren des Liegeverhaltens sowie die Erfassung der Verschmutzung und der Verletzungen der Kühe als besonders geeignet heraus, um haltungs- und managementbezogene Schwachstellen in Liegeboxenlaufställen aufzudecken. Die erfolgreiche Umsetzung der Schwachstellenanalyse in der Nachuntersuchung lässt aufgrund der deutlichen Verbesserungen im Bereich der Tiergerechtigkeit insgesamt auf einen hohen Nutzen des Beratungssystems schließen.

4 Fazit

Im Rahmen der produktionstechnischen Beratung ermöglicht die Schwachstellenanalyse somit einen umfassenden Überblick über die Ist-Situation der Herde. Zudem wird dem Anwender durch die digitale Erfassung und Auswertung der Daten vor Ort sowie die eingängigen grafischen Darstellungen ein schneller Vergleich der einzelbetrieblichen Werte mit den definierten Referenzwerten aus dem Expertensystem ermöglicht. Anhand der Auflistung konkreter Risikofaktoren in der Haltungsumwelt lassen sich standardisierte Handlungsempfehlungen ableiten, die eine Verminderung der Beeinträchtigungen für die Tiere und somit eine Verbesserung der Tiergerechtigkeit der Haltungsbedingungen zum Ziel haben.

Literaturverzeichnis

- [Da09] Dahlhoff, K. et al.: Ergebnisse zur Schwachstellenanalyse für die Bereiche Haltung und Management in Liegeboxenlaufställen im Rahmen des Projektes „Cows and more“.; In: 9. Tagung: Bau, Technik und Umwelt in der landwirtschaftlichen Nutztierhaltung, Berlin 2009, S. 374-380
- [Pe07] Pelzer, A. et al.: Cows and more – Was die Kühe uns sagen. Bonitieren – Bewerten – Beraten mit System.; In: 8. Tagung: Bau, Technik und Umwelt in der landwirtschaftlichen Nutztierhaltung, Bonn 2007, S. 97-102
- [Rü10] Untersuchung verschiedener Parameter auf ihre Eignung zur Bewertung der Tiergerechtigkeit von Laufställen für Milchkühe im Rahmen eines On-farm welfare assessment.; Ludwig-Maximilians-Universität München; München 2010; Diss. vet.-med.
- [Su98] Zur Beurteilung der Tiergerechtigkeit von Haltungsbedingungen landwirtschaftlicher Nutztiere.; In: Dtsch. Tierärztl. Wschr. 105, S. 65 – 72
- [Wh03] Whay, H.R. et al.: Assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records.; In: Vet. Record 153, S. 197-202

fmea3d – Implementierung einer alternativen RPZ-Berechnungsmethode in R

Tilman Wilke, Juliane O'Hagan, Sophia Schulze-Geisthövel, Brigitte Petersen

Universität Bonn
Abteilung Präventives Gesundheitsmanagement
Katzenburgweg 7-9
53115 Bonn
tilman.wilke@uni-bonn.de
b-petersen@uni-bonn.de

Abstract: Im Rahmen der präventiven Qualitätsmanagement-Methode FMEA (Failure Mode and Effects Analysis) wird jedes Risikoelement durch die Parameter Bedeutung (B), Auftretenswahrscheinlichkeit (A) und Entdeckungswahrscheinlichkeit (E) charakterisiert. Um die Risikoelemente rangieren zu können, wird für jedes Risikoelement eine Risikoprioritätszahl (RPZ) berechnet. Für diese Berechnung existieren verschiedene Methoden. Eine vielversprechende alternative Berechnungsmethode ist der von Werdich [Wer12] vorgeschlagene ‚3D-Ampelfaktor‘. Bisher gab es jedoch keine einfach zugängliche Implementierung des Ampelfaktors. Um diese Lücke zu schließen, wurde die Methode des Ampelfaktors auf der technischen Basis der Statistiksoftware R und Microsoft Excel™ implementiert. Das entwickelte R-Skript liest diese Daten ein, und erzeugt daraus eine interaktive 3D-Darstellung, 2D-Projektionen sowie eine tabellarische Ausgabe der Bewertungsergebnisse.

1 Hintergrund

Bei der FMEA (englisch: Failure Mode and Effects Analysis, deutsch: Fehlermöglichkeits- und Einflussanalyse) handelt es sich um eine präventive Qualitätsmanagement-Methode. Die Methode FMEA wird traditionell in der Entwicklung technischer Produkte oder Prozesse eingesetzt. Die FMEA wird aber auch in Bereichen der Tierproduktion und Lebensmittelsicherheit erfolgreich eingesetzt, z.B. im Qualitätsmanagement der Schweinefleischproduktion [GMP06] und in der Beratung landwirtschaftlicher Betriebe im Tiergesundheitsmanagement [Sch06].

Im Rahmen der FMEA wird das betrachtete System eingegrenzt, in seine Funktionen zerlegt, und Fehlerfolgen und Fehlerursachen dieser Funktionen gesucht (Systemanalyse). Anschließend werden die mit den Fehlerfolgen verbundenen Risiken bewertet, um Vermeidungsmaßnahmen vornehmen, priorisieren und kontrollieren zu können (Risikobewertung). Im Rahmen der Risikobewertung werden Fehlerfolgen hinsichtlich ihrer Bedeutung (B), ihrer Auftretenswahrscheinlichkeit (A) sowie ihrer Entdeckungswahrscheinlichkeit (E) bewertet. Um die Risikoelemente für das Risikomanagement rangieren zu können, wird auf Grundlage von B, A und E für jedes

Risikoelement eine Risikoprioritätszahl (RPZ) berechnet. Die am weitesten verbreitete und einfachste Form der RPZ ergibt sich als Produkt aus $B \cdot A \cdot E$.

Es bestehen jedoch erhebliche Zweifel an der Aussagekraft dieser einfachen RPZ-Berechnungsmethode und sie wird daher von FMEA-Spezialisten, Fachgremien und Normgebern mehrheitlich abgelehnt [Wer12]. Eine vielversprechende Alternative ist der von Werdich [Wer12] vorgeschlagene ‚3D-Ampelfaktor‘: Werdichs Methode basiert auf dem Prinzip der Risikomatrix, bei dem paarweise Kombinationen von B, A und E bewertet werden – üblicherweise mit einer 3-stufigen Bewertungsskala. In der Praxis werden die 3 Bewertungsstufen als Ampelfarben dargestellt (grün = kein Handlungsbedarf, gelb = kein zwingender Handlungsbedarf, rot = dringender Handlungsbedarf). Die Risikomatrizes geben also die Einstellung des Risikomanagers wieder. Werdichs Ampelfaktor berücksichtigt bei der Berechnung der RPZ die Kombination dieser 3 Risikomatrizen. Die drei Risikomatrizes bilden dabei die Seitenflächen eines Würfelraumes. Im ersten Schritt wird jedes Risikoelement anhand der Zahlenwerte aus der Risikobewertung in diesem B-A-E-Koordinatensystem platziert. Dann wird für jedes Risikoelement und für alle drei Seitenflächen ermittelt, welche Farbe das Feld der Risikomatrix besitzt, über dem das Risikoelement ‚schwebt‘. Indem die Ampelfarben der Risikomatrizen mit Zahlenwerten assoziiert sind (grün=0, gelb=1, rot=2), ergibt sich für jedes Risikoelement ein Zahlentripel zwischen (0, 0, 0) und (2, 2, 2). Um die finale RPZ zu errechnen wird die Summe dieses Tripels gebildet, die zwischen 0 und 6 liegen kann. Bei Untersuchungen des Fraunhofer IPA ergab diese RPZ eine sehr viel bessere Übereinstimmung mit einer von Experten erarbeiteten Referenzrangierung, als mit der klassischen RPZ-Berechnungsmethode [Sch11].

2 Problemstellung

Bisher gibt es keine frei zugängliche Implementierung des Ampelfaktors. Dies erschwert aus Sicht der Autoren die Anwendbarkeit dieser Methode in der FMEA-Praxis sowie in der Ausbildung im Bereich Qualitäts- und Risikomanagement. Denn hier kommen erfahrungsgemäß solche Methoden zum Einsatz, für die leicht zugängliche Werkzeuge verfügbar sind. In der Zusammenarbeit mit kleinen und mittleren Unternehmen der Agrar- und Ernährungswirtschaft machten die Autoren die Erfahrung, dass selten Bereitschaft besteht, für die Durchführung einer FMEA besondere Software anzuschaffen. Gleichzeitig ist es im Bereich der Ausbildung notwendig, dass den Studierenden Werkzeuge an die Hand gegeben werden können, die einerseits den Stand von Wissenschaft und Praxis widerspiegeln, und andererseits nicht als ‚Black-Box‘ arbeiten, sondern nachvollziehbar bleiben.

3 Lösungsansatz

Um dieses Problem zu lösen, sollte eine Software-Lösung entwickelt werden, die die Logik des Ampelfaktors implementiert, und außerdem für den Einsatz in der Lehre und in der Zusammenarbeit mit Unternehmen geeignet ist. Für die Implementierung des Ampelfaktors fiel die Wahl auf eine Kombination aus R und Microsoft Excel™ (oder ein zu Open Office XML kompatibles Tabellenkalkulationsprogramme). Solche Tabellenkalkulationsprogramme sind sowohl bei Unternehmen als auch bei Studierenden sehr weit verbreitet, und funktionieren mit Standard-Benutzerrechten. R wurde ausgewählt, weil es als frei verfügbares Statistikprogramm gerade bei Studenten wachsende Verbreitung findet. Falls R nicht bereits installiert ist, kann es ohne spezielle Kenntnisse und Benutzerrechte installiert werden. Weil R eine Skriptsprache ist, bleiben alle Arbeitsschritte des Programms nachvollziehbar und erweiterbar.

4 Ergebnis und Diskussion

Das Ergebnis ist eine Kombination aus einem R-Skript und einer Excel-Arbeitsmappe. Die Excel-Arbeitsmappe im Open Office XML Format mit vier Arbeitsblättern dient als Quelle für die drei Risikomatrizes und die B-A-E-Risikobewertungen. Das entwickelte R-Skript liest diese Datei ein und erzeugt eine interaktive 3D-Darstellung, 2D-Projektionen, sowie eine tabellarische Ausgabe der Bewertungsergebnisse. Das Programm ist auf individuelle Anforderungen einer FMEA anpassbar: Die Bewertungsskalen der FMEA (1-10) und der Risikomatrizes (1-3) können frei gewählt werden, ebenso wie die Farben, die Ausrichtungen und die Bezeichnungen der Risikodimensionen. Abbildung 1 zeigt ein Bildschirmfoto der interaktiven 3D-Grafik sowie die 2D-Darstellung in Form der drei Risikomatrizes. Die Zahlenwerte in der 2D-Darstellung geben an, wie viele Elemente auf einem Feld der Risikomatrix liegen und welche Bewertung diese Elemente haben.

Für die eigentliche FMEA-Arbeit ist die tabellarische Ausgabe am wichtigsten. Sie enthält neben dem Ampelfaktor auch die klassische RPZ und die Rangierung nach beiden Methoden. Die Tabelle kann als separate Datei gespeichert werden, oder als neues Arbeitsblatt in die Eingabe-Datei geschrieben werden. Im Rahmen von Abschlussarbeiten und in der Lehre soll die vorliegende Implementierung in Zukunft dazu genutzt werden, um den Einsatz zeitgemäßer FMEA-Methodik zu fördern, und die Bedeutung der Auswahl der RPZ-Berechnungsmethode zu veranschaulichen. Eine Erweiterung des Skripts um weitere RPZ-Berechnungsmethoden ist leicht möglich.

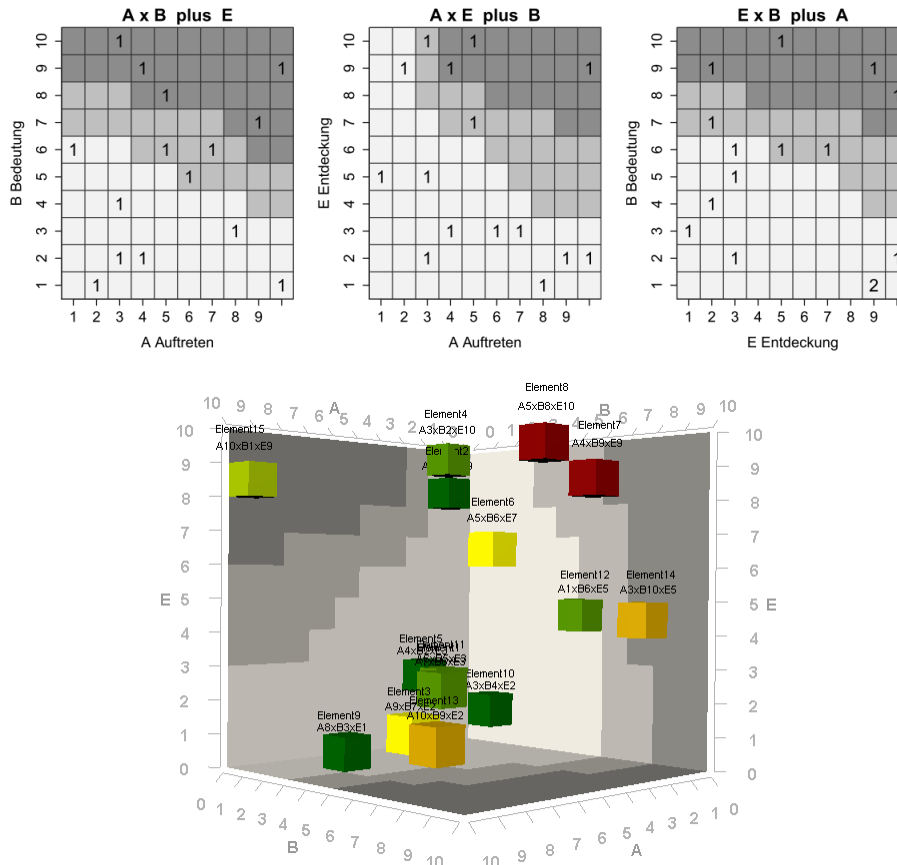


Abbildung 1: Vom R-Skript *fmea3d* erzeugte Darstellungen der FMEA-Risikobewertung mittels der RPZ-Berechnungsmethode ‚Ampelfaktor‘: 2D-Darstellung und interaktive 3D-Darstellung.

Literaturverzeichnis

- [GMP06] Gödderz, A. et al.: Failure Mode and Effect Analysis (FMEA) as a decision support tool within a quality information system in pork production chains. In (Theuvsen, L. et al. Hrsg.): Quality Management in Food Chains. Wageningen Academic Publishers, 2006; S. 139–148.
- [Sch06] Schmitz, T.: Entwicklung eines Modells zum Einsatz präventiver Qualitätsmanagementmethoden in der Beratung von Unternehmen der Agrar- und Ernährungswirtschaft. Dissertation, Universität, Bonn, 2005. Shaker, Aachen, 2006.
- [Sch11] Schloske, A.: Studie zur Risikobewertung in der FMEA. Fraunhofer Institut IPA, Osnabrück, 2011.
- [Wer12] Werdich, M. Hrsg.: FMEA - Einführung und Moderation. Durch systematische Entwicklung zur übersichtlichen Risikominimierung (inkl. Methoden im Umfeld). Vieweg & Teubner, 2012.

Recognition of Activity States in Dairy Cows with SVMs and Graphical Models

J. Behmann, K. Hendriksen, U. Müller, S. Walzog, W. Büscher, L. Plümer

University of Bonn
Institute of Geodesy and Geoinformation
Meckenheimer Allee 172
53115 Bonn
behmann@igg.uni-bonn.de
hendriksen@uni-bonn.de

Abstract: Activity patterns of dairy cattle have received increasing interest in recent years because they promise insights into health state and well-being. The fusion with data from additional sensor signals promises a comprehensive monitoring of activity patterns composed of sequences of single activity states. We used a combination of a Support Vector Machine (SVM), a state of the art classification method, and a Conditional Random Field (CRF). SVMs distinguish single states, whereas CRFs label state sequences under consideration of specified constraints. In a preliminary experiment, a Local Positioning System was combined with a heart rate sensor in order to estimate seven spatiotemporal activity states. The application of the CRF to the SVM result caused a slight increase in accuracy (5%) but a major improvement at the correct determination of long sequences (increasing length of the longest common subsequence from 3481 to 6207 periods). This robust detection of long lying sequences allowed for the unaffected extraction of the resting pulse.

1 Introduction

Optimal conditions and continuous monitoring are a major factor for achieving optimum levels of animal welfare with high performance dairy cows. Activity patterns are sensitive indicators for animal welfare - consequently many sensors have been applied for the automatic monitoring of cow activity patterns. The limiting factor for the monitoring of dairy cattle is the versatility of activity states, each with individual characteristics. These versatile distinguishing features may be covered by a high number of specialized sensors or more efficiently by a combined analysis of complementary sensors. For this data fusion we utilized the Machine Learning methods SVM and CRF. Our objective was to develop a model that is able to derive the most probable activity state for every observed point in time (the spatiotemporal activity pattern) for every individual cow. In a preliminary experiment, a combination of two sensor types was used in order to deduce reliable information about seven activity states. Two sensor systems were used: a local position measurement system and a heart rate sensor.

2 Methods for Data Analysis

The data analysis methods were applied in three steps within the derivation process of the activity patterns. In the first step, spatiotemporal features were extracted from filtered sensor data. The position and heart rate were synchronized and thereafter joint spatiotemporal features were extracted. The position features are based on spatial knowledge about the layout of the cattle barn and included the distances to important objects like cubicles and feeding stations. The heart rate sensor recorded the heart rate (HR) together with a time stamp at intervals of one second. In addition to the HR, the most common heart rate variability (HRV) parameter (RMSSD; in ms) and additional features which are gained by the Recurrence Quantitative Analysis [MLN02] was added to the feature set.

In the second step, preliminary probabilities for activity states were derived by a multiclass Support Vector Machine (SVM) [CV95]. SVMs are state of the art methods for supervised classification and provide linear and non-linear discrimination functions [CV95]. The SVM model consists of a subset of weighted data instances - so called support vectors. They define a hyperplane separating classes in the feature space. The distance of a feature vector to the hyperplane is transformed to class probabilities. The used SVM achieved a prediction accuracy of 79.01% for the classification of activity states. Consequently it outperformed the widespread Naive Bayes approach (56.33%) and is comparable to the Random Forest classifier (78.03%).

In the last step, these probabilities were linked with each other and combined with contextual knowledge by a Conditional Random Field (CRF) [LNP01]. The probabilities of the unobserved variables were computed given the evidence of observed variables in the inference step. When applied for activity state recognition, the observation nodes were realized by the class probabilities derived by the SVM model. The edges between the label nodes describe a semi-Markov process. This facilitates the duration dependent control of the model inertia. Two different functions for the transition probabilities for the different activity states were used: a linear model for fast state transitions ("feeding", "drinking" and "walking") and sigmoidal model for a longer persistence in a state ("lying" and "standing"). Specific transitions in the resulting sequence e.g. direct transition from "lying" to "drinking" are prohibited by these functions. Thereby, it was forced that the states "standing up" and "walking" has to occur intermediately. Hence, the generated time series of labels are always valid from the viewpoint of physical constraints of the cattle in the barn. The positive effects of the graphical model as refinement of the SVM result are based on the inclusion of three additional priors: (i) consecutive points in time have most probable the same state, (ii) some state transitions are impossible and (iii) the duration of different states follows different distributions.

3 Exemplary Application

An exemplary application was implemented at a herd of 65 German Holstein Friesian cows which were loose-housed in a two-row open free-stall barn with cubicles and concrete floor. The study regarded seven activity states "standing", "lying", "walking", "feeding", "drinking", "standing up", and "lying down" defined in a protocol and used as

labels for the training data (manual annotation). The observations were taken at three periods for at least 4 hours in the morning (from 07:00 am to 12:30 pm). 12 cows were selected to cover the herd range of lactation numbers, social status and day of pregnancy and overall, 43 complete and plausible time series were recorded.

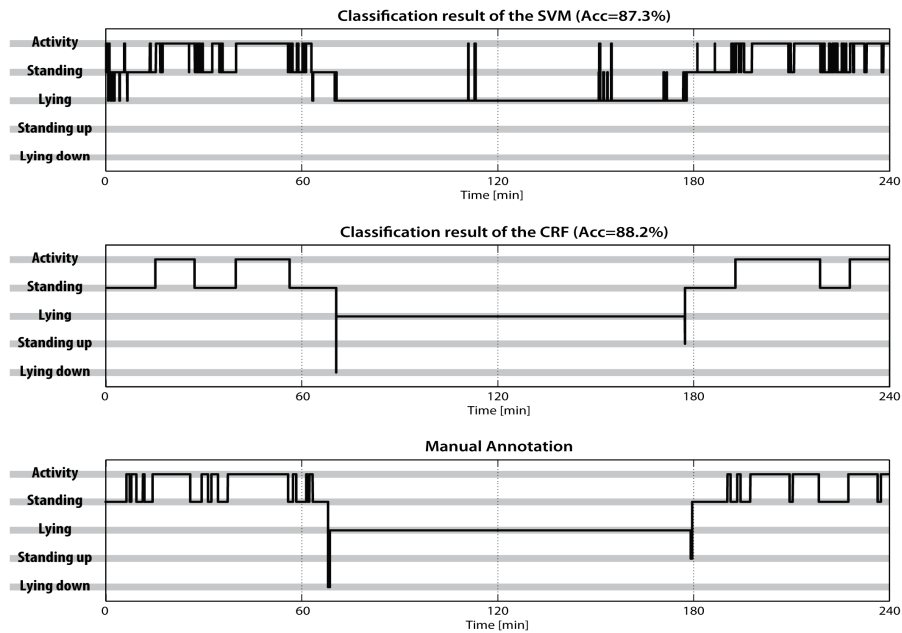


Figure 1: Prediction of activity states showing the different characteristics of the SVM classification result and the result of the graphical model. From left to right: (a) classification result of the SVM with prediction accuracy, (b) classification result of the CRF with prediction accuracy and (c) the manual annotation. The state “activity” comprises the states “walking”, “feeding” and “drinking”.

As position sensor the Local Position Measurement system (LPM, Abatec group AG, Austria) based on runtime measurement was applied [GNB07]. In the cow barn, shadings and reflections and the small distances between the installed antennas greatly impair sensor signals. The accuracy of x- and y-coordinate fluctuates between 10 cm and 4 m. The z-coordinate was neglected for an improved accuracy in the remaining dimensions [GNB07]. For the measurement of HR and the HRV, sensors from Polar Electro GmbH (Büttelborn, Germany) were used (Polar Equine RS800, Polar Equine RS800CX). The electrodes were integrated in a chest belt (Polar Equine WearLink® W.I.N.D. transmitter) and the signals were transmitted wireless to a data logger.

4 Results and Discussion

The presented method of analysis combining feature extraction, SVM classification, and graphical model derived the most probable sequence of activity states. One exemplary time series is presented in Figure 1 showing the accuracy of the classification by the SVM and its improvement by the additional use of the graphical model. Long sequences

like "lying" were classified without single errors, which enabled the robust extraction of parameters like "longest lying sequence" or "number of standing up events". This focus is considered by utilizing the additional quality measurements "Longest Common Subsequence" (LCS) and average "Common Subsequences" (CS) which demonstrates the improved result quality of the CRF (Table 1). The point in time of "standing up" and "lying down" was correctly determined.

Table 1: Comparison of classification results of the SVM and the CRF. Prediction quality is assessed by prediction accuracy in percent, the Longest Common Subsequence (LCS), and the Mean Length of all "Common Subsequences" (CS).

| Evaluation step | Count of state changes | Mean lengths of LCS | Mean length of all CS | Accuracy "Lying" | Accuracy "Standing" |
|-----------------|------------------------|---------------------|-----------------------|------------------|---------------------|
| SVM | 327.72 | 3481 | 107 | 84% | 81% |
| CRF | 6.12 | 6207 | 1271 | 86% | 89% |

The automatic detection of resting sequences by the analysis method in combination with the recorded signal of the heart rate sensor allowed for the automatic determination of the stable pulse rate while lying. The mean resting pulse rate (RPR) of individual animals varied from 64.9 to 81.0 bpm (SD = 4.86 bpm). The mean SD of the RPR of an individual cow was 1.96 bpm (within the range of 0.29 to 3.66 bpm). Furthermore it was noticed that the automatically derived RPR depended significantly on the respective stage of pregnancy ($p = 0.049$; by Kruskal-Wallis H Test). Increasing resting heart rate with duration of pregnancy of women is well documented.

The developed transition model supports the robust determination of "standing up" and "lying down" sequences (Figure 1), although this information is not accessible by a single sensor. HR and HRV help to identify these processes which are not covered by the LPM and enable the automated recognition of spatiotemporal activity pattern. The presented, integrated analysis uncovers this information because it combines the outstanding generalization ability of the SVM and the duration dependent sequence model of the CRF. The analysis method can be applied to data of additional sensors like rumination sensors or pedometers, whereby the prediction quality will continue to improve and further states like "rumination" will become accessible.

References

- [CV95] Cortes, C.; Vapnik V.: Support-vector networks. *Machine Learning* 20(3), 1995; S. 273-297.
- [GNB07] Gygas, L.; Neisen, G.; Bollhalder, H.: Accuracy and validation of a radar-based automatic local position measurement system for tracking dairy cows in free-stall barns. *Computers and Electronics in Agriculture*. 56(1), 2007; S. 23–33.
- [LNP01] Lafferty, J.D.; McCallum, A.; and Pereira, F.C.N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proc. Int. Conf. Machine Learning*, Morgan Kaufmann Publishers Inc., 2007; S. 282–289.
- [MLN02] Mohr, E.; Langbein, J.; Nürnberg, G.: Heart rate variability: a noninvasive approach to measure stress in calves and cows. *Physiology and Behavior*, 75(1), 2002; S. 251–259.

Planungsdaten schnell finden und einfach nutzen: Linked Open Data und semantische Suche im Einsatz für das KTBL-Datenangebot

Daniel Martini¹, Daniel Herzig², Martin Kunisch¹, Günter Ladwig²

- 1) Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (KTBL)
Bartningstraße 49, 64289 Darmstadt
- 2) Daniel Herzig & Günter Ladwig Softwarelösungen GbR (SearchHaus)
d.martini@ktbl.de
herzig@searchhaus.net

Abstract: Um Entscheidungen fundiert treffen zu können, sind Daten als Grundlage unerlässlich. Der Aufwand für die Recherche nach relevanten Daten sowie die anschließende Aufbereitung der Daten für die Verwendung in Kalkulationswerkzeugen stellen die wesentlichen Herausforderungen für datenbasierte Entscheidungen dar. In diesem Beitrag wird eine Lösung zur Unterstützung von datengestützten Entscheidungen in der Agrarwirtschaft vorgestellt, die die Recherche nach Planungsdaten aus dem KTBL-Datenangebot durch den Einsatz einer semantischen Suchmaschine vereinfacht und durch die Bereitstellung der KTBL-Daten gemäß der Linked Open Data Prinzipien eine leichte Weiterverarbeitung ermöglicht.

1 Einleitung

In der Landwirtschaft werden täglich Entscheidungen über Investitionen, Arbeitsplanungen und Produktionsabläufe anhand von Daten und darauf aufsetzenden Berechnungen gefällt. Wenn beispielsweise ein Landwirt überlegt, in den Anbau von Weizen einzusteigen und dafür die Investition für neue Maschinen abwägen möchte, konnte er in der Vergangenheit auf umfangreiche Datensammlungen des Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V. in gedruckter Form zurückgreifen, wie beispielsweise die Faustzahlen Landwirtschaft [K09] oder die Datensammlung Betriebsplanung [K12]. Bereits seit einigen Jahren spielt das Internet auch im Agrarbereich eine zunehmende Rolle bei der Veröffentlichung und Verbreitung von Daten. Bislang waren online angebotene Daten eng integriert mit ihren Anwendungen und dem Nutzer lediglich über vorgefertigte, von den jeweiligen Entwicklern in den Bedienoberflächen festgelegte Interaktionswege zugänglich. Eine Weiternutzung und Verarbeitung der Daten in anderen Systemen wie z. B. Farmmanagement-Informationssystemen oder Beratungswerkzeugen oder ein Zusammenführen von Datenbeständen verschiedener Organisationen zur Beantwortung neuer Fragestellungen war mithin nur erschwert möglich und erforderte aufwändige Arbeiten für Einpflegen oder Import.

Technologien im Umfeld des Semantic Web erlauben inzwischen für die Datennutzer elegantere und flexiblere Lösungen. So wird im Rahmen der Linked Open Data (LOD) Initiative des World Wide Web Consortium versucht, Daten von Auswertungslogik und Benutzerschnittstelle zu trennen und mit einfachen Internet-Protokollen (in erster Linie HTTP) zugänglich zu machen [Be06, HB11]. Insbesondere werden die Daten hier auch maschinenlesbar in standardisierten Formaten bereitgestellt, sodass diese über einfache URL-Aufrufe abgerufen und automatisiert in Anwendungen eingelesen werden können.

Für die Repräsentation der Daten setzt man auf das Resource Description Framework (RDF), sowie Vokabulare, die z. B. in RDF Schema [BG04] erstellt werden können. Kernaspekt ist dabei die Modellierung von Daten als gerichteter Graph. In diese allgemeine und generische Datenstruktur können Datensätze jeglicher Art meist einfach überführt werden. Außerdem werden durch diese Art der Repräsentation Beziehungen zwischen Daten explizit modelliert und spezifiziert, sodass Datensätze besonders zugänglich werden für Suche, Navigation und die Beantwortung komplexer Fragestellungen, die die Einbeziehung einer Reihe von Bezügen erfordern. Im Rahmen des iGreen-Projektes und einer langjährigen Kooperation mit der FAO konnten die Autoren grundlegende Entwicklungen in diesem Bereich vorantreiben. Um die Daten des KTBL gemäß dieser Prinzipien zugänglich zu machen, wurden das Vokabular agroXML¹ sowie Ergebnisse aus dem iGreen-Projekt² als Grundlage zur Beschreibung verwendet.

Aufbauend auf diesen Vorarbeiten hat das KTBL nun auf Basis der vorhandenen KTBL-Datenbank mit Informationen zu Arbeitsverfahren und Maschinenkosten einen Linked Open Data Service aufgesetzt, der den maschinellen Abruf dieser Daten erlaubt. Obenauf wurde von SearchHaus, einem Spin-Off des Karlsruher Institut für Technologie (KIT), eine intelligente semantische Suche eingerichtet, die ein gezieltes Auffinden von Datensätzen ermöglicht. Wird beispielsweise die Frage untersucht, wie hoch die Investitionskosten für Maschinen sind, die zur Aussaat von Weizen verwendet werden, so kann mit einfach Stichwortanfragen, z.B. „*anschaffungspreis maschinen weizen saat*“, der KTBL-Datenbestand durchsucht werden. Die Anfragen können frei formuliert werden und es sind keine speziellen Kenntnisse über Aufbau oder Abfrage der Datenbank notwendig. Für mögliche Treffer werden verschiedene Interpretationen, die anhand der in den Daten vorhandenen Beziehungen ermittelt werden, angezeigt, Ergebnisse können anschließend über dynamisch berechnete Facetten eingeschränkt werden.

2 Material und Methoden

Der Linked Open Data Service wurde mit Hilfe des Open Source Werkzeugs d2rq (<http://d2rq.org>) auf die bestehende Oracle 11 Datenbank aufgesetzt. Hierfür wurde mit der d2rq mapping language ein Mapping erstellt, das spezifiziert, wie relationale Datenbanktabellen in ein RDF Graphen-Modell zu überführen sind. Als Ergebnis steht

¹ <http://www.agroxml.de>, abgerufen am 14.10.2013

² <http://www.igreen-projekt.de>, abgerufen am 14.10.2013

sowohl ein http-Service, der Daten sowohl im HTML-Format für den Browser als auch im maschinenlesbaren Turtle-Format [PC13] für RDF ausliefern kann, sowie ein SPARQL-Endpoint [HS13] für formal spezifizierte Abfragen zur Verfügung. Neben den in der lokalen Datenbank vorhandenen Daten wurden in den Datensatz auch Verknüpfungen zu dem multilingualen AGROVOC Thesaurus der FAO³ eingebaut, sodass für einige Maschinenarten die entsprechenden Konzepte dort sowie die Übersetzungen der Bezeichnung in eine Vielzahl von Sprachen abgerufen werden können. Die semantisch aufbereiteten Daten bilden die Grundlage für die Suche.

Mit der semantischen Suche von SearchHaus können die Benutzer einfach in den KTBL-Daten mit Schlüsselwörtern suchen, so wie sie es von Websuchmaschinen gewohnt sind. Die Suchmaschine nimmt Schlüsselwortanfragen entgegen, interpretiert die Anfragen mit Hilfe des Datenbestands und liefert dem Benutzer die passenden Ergebnisse aus den KTBL-Daten. Intern aggregiert und gruppiert die Suchmaschine die komplexen, strukturierten Daten, so dass eine schnelle Beantwortung der Anfragen auch über große Datenbestände möglich ist. Auf dieser kompakten Datenrepräsentation werden dann mittels Explorationsalgorithmen und Rankingverfahren die Suchergebnisse ermittelt [LT10]. In einem zweiten Schritt kann der Benutzer das Suchergebnis interaktiv mittels automatisch berechneter Facetten einschränken und so genau an sein Informationsbedürfnis anpassen. Der Vorteil der Suche gegenüber bestehenden Lösungen ist, dass die Suche die Zusammenhänge in den Daten erkennt und für die Berechnung der Ergebnisse ausnutzen kann. Abbildung 1 zeigt einen Ausschnitt aus den Ergebnissen für die Suchanfrage nach Anschaffungspreisen von Maschinen, die für die Aussaat von Weizen geeignet sind (als Stichwortanfrage: "maschine preis weizen saat").

Hilfe

Anfragen an KTBL-Datenbank (4) [Alle anzeigen...](#)

Feldarbeit **genutzte Maschine** **KTBL Standardmaschine** [Alle Ergebnisse anzeigen ▶](#)

Feldarbeit **Bezeichnung enthält** "weizen saat"

KTBL Standardmaschine **Anschaffungspreis**

Ergebnisvorschau (192 insgesamt):

| KTBL Standardmaschine | KTBL Standardmaschine | Anschaffungspreis | Feldarbeit |
|--|-----------------------|-------------------|---|
| Standardtraktor mit Allradantrieb und Stufengetriebe - 93-111 kW | 85500 | | Mulchsaat von Weizen, Ackerbohnen, Erbsen - 4,5 m; 102 kW |
| Saatgutförderschnecke - zum Überladen mit Hydro-Motor; 3 m; 20 t/h | 1100 | | Mulchsaat von Weizen, Ackerbohnen, Erbsen - 4,5 m; 102 kW |
| Mulchsaatmaschine - 4,5 m; 2500 l | 75000 | | Mulchsaat von Weizen, Ackerbohnen, Erbsen - 4,5 m; 102 kW |

Abbildung 2: Ausschnitt eines Screenshots. Zu sehen sind Ergebnisse für die Suchanfrage nach Maschinen und deren Anschaffungspreis, die für die Aussaat von Weizen geeignet sind.

3 Ergebnisse

Sowohl der Linked Open Data Service als auch die semantische Suche ließen sich sehr einfach in die bestehende Infrastruktur integrieren. Auf Datenbankebene waren keine Anpassungen an bestehenden Zugangsmechanismen, Datenbankschemas oder Tabellen notwendig. Mittelfristig kann die Implementation durch die Erstellung zusätzlicher

³ <http://aims.fao.org/standards/agrovoc/about>, abgerufen am 14.10.2013

Views optimiert werden, ein zwingender Bedarf ergab sich hierfür jedoch im prototypischen Betrieb nicht. Mit überschaubarem Aufwand konnte innerhalb eines Zeitraumes von etwa drei Monaten unter Aufwendung von etwa 2-3 Personenmonaten eine vollständig spezifikationsgemäß funktionierende Implementation erzielt werden.

Derzeit startet der interne Probetrieb und die ersten Indikatoren lassen darauf schließen, dass die KTBL-Mitarbeiter von den Möglichkeiten, sich schnell einen Überblick über vorhandene Daten verschaffen zu können, profitieren. Es ist zu erwarten, dass Berater und Landwirte innerhalb des KTBL-Datenangebotes schneller und zielgerichteter Anwendungen, Daten und Veröffentlichungen finden, die für ihre Fragestellungen Antworten liefern können. Zusätzlich bekommen sie bei weiterer Verbreitung der Technik weitere Informationen anderer Anbieter, die zu der Anfrage passen, angezeigt. Die maschinenlesbare Schnittstelle kann von Entwicklern bei Herstellern von Farmmanagement-Informationssystemen und Entscheidungsunterstützungssystemen genutzt werden, um neue Funktionalitäten zu implementieren. Auch seitens des KTBL können so neue Anwendungen durch Daten externer Einrichtungen angereichert werden.

In Detailpunkten wurden auch Limitationen gefunden. So ist es beispielsweise in der d2rq mapping language nicht möglich, in RDF übliche Angaben zur Sprache an Bezeichnungen, die aus Inhalten mehrerer Datenbanktabellen zusammengesetzt wurden, anzuhängen. Auch die Abbildung, Interpretation und Sortierung von physikalischen Einheiten hat noch Raum für Verbesserungen. Insgesamt stellen die Arbeiten einen zukunftsweisenden Ansatz zur Etablierung von innovativen und flexibleren Lösungen zur Nutzung des KTBL-Datenangebots und insbesondere zur Beantwortung von Fachfragen dar, die langfristig die integrierte Nutzung von Daten, die in verschiedenen Organisationen erhoben wurden, erlauben.

Literaturverzeichnis

- [Be09] Berners-Lee, T.: Linked Data. World Wide Web Consortium, 2009. <http://www.w3.org/DesignIssues/LinkedData.html>, aufgerufen am 10.10.2013.
- [BG04] Brickley, D., Guha, R. V.: RDF Vocabulary Description Language 1.0. World Wide Web Consortium, 2004. <http://www.w3.org/TR/rdf-schema/>, aufgerufen am 10.10.2013.
- [HB11] Heath, T., Bizer, C.: Linked Data – Evolving the Web into a Global Data Space. Morgan & Claypool Publishers, 2011.
- [HS13] SPARQL 1.1 Query Language. World Wide Web Consortium, 2013. <http://www.w3.org/TR/sparql11-query/>, aufgerufen am 10.10.2013.
- [K09] Kuratorium für Technik und Bauwesen in der Landwirtschaft (Hrsg.): Faustzahlen für die Landwirtschaft (14. Auflage). KTBL e. V., Darmstadt 2009.
- [K12] Kuratorium für Technik und Bauwesen in der Landwirtschaft (Hrsg.): Betriebsplanung Landwirtschaft 2012/13 – Daten für die Betriebsplanung in der Landwirtschaft (23. Auflage). KTBL e. V., Darmstadt, 2012.
- [LT10] Ladwig G., Tran T.: Combining Keyword Translation with Structured Query Answering for Efficient Keyword Search, Proceedings of the 7th Extended Semantic Web Conference (ESWC '10), Springer, 2010
- [PC13] Prud'hommeaux, E., Carothers, G.: Turtle – Terse RDF Triple Language. World Wide Web Consortium, 2013. <http://www.w3.org/TR/turtle/>, aufgerufen am 10.10.2013.