Milchvieh-Informationsmanagement auf Versuchsbetrieben – Beispielanwendungen und Nutzen für Praxisbetriebe

¹Wolfgang Büscher, ¹Kathrin Hendriksen, ²Ute Müller, ²Peter Müller,
³Andreas Behrend, ⁴Eckhard Stamer

¹Universität Bonn
Institut für Landtechnik
buescher@uni-bonn.de

²Universität Bonn, Institut für Tierwissenschaften

³Universität Bonn, Institut für Informatik III

⁴TiDa Tier und Daten GmbH

Abstract: An den milchwirtschaftlichen Versuchsbetrieben der Universitäten und anderen Forschungseinrichtungen werden unterschiedlichste wissenschaftliche Fragestellungen bearbeitet. Bei den Forschungsthemen fallen viele zusätzliche Daten an, die oft weit über die üblichen Daten des klassischen Herdenmanagements hinausgehen. Das Kuhdatenmanagementsystem „KuhDaM“ (ein relationales Datenbanksystem) ermöglicht eine Zusammenführung und Vernetzung der anfallenden Daten. Mit Hilfe dieser relationalen Datenbanken auf Betriebsservern können die Informationen zusammen geführt werden, wodurch komplexe Analyse- und Auswertungsmethoden möglich sind. Dazu werden unter anderem bewährte Datenstromanalysetechniken direkt in den relationalen Datenbankkontext übertragen, um dort neben der etablierten Analyse historischer Daten auch die Analyse von dynamischen Stromdaten zu ermöglichen. Da langfristig eine hohe Übereinstimmung zwischen Forschungs- und Praxiszielen besteht, ist die „Datenstromanalyse“ ein wichtiges Zukunftsthema für die Wissenschaft und Industrie.

1 Milchwirtschaftliche Versuchsbetriebe: Situationsbeschreibung

Milchwirtschaftliche Versuchsbetriebe haben gemeinsam, dass es immer eine Trennung zwischen einem Wirtschafts- und einem Forschungsbetrieb gibt. Der Wirtschaftsbetrieb befasst sich mit den täglichen Aufgaben und stellt die Basis für zeitlich befristete Forschungsfragen dar. Ein Teil des Tierbestandes bewegt sich themenbezogen im Forschungsbetrieb und wird ggs. besonders behandelt oder intensiv untersucht. Bei den Forschungsthemen fallen viele zusätzliche Daten an, die oft weit über die üblichen Daten des klassischen Herdenmanagements hinausgehen. Am Beispiel der Milchviehherde des Forschungsbetriebes Frankenforst und der Zusammenarbeit mit einem an der Universität Kiel entwickelten Datenbank-System („KuhDaM“) soll im Vortrag erläutert werden, wie

die Datensituation auf Versuchsbetrieben typischerweise aussieht und welche Anwendungsziele verfolgt und bisher umgesetzt werden.

2 Beispiele für hohe Individualität in den Merkmalsausprägungen

Durch die auf dem Versuchsbetrieb Frankenforst vorhandenen Futterwiegetröge für die Grundfütterration, die Wasserwiegetröge und die Krafftutterstationen fällt eine Vielzahl von Daten im Bereich der Futter- und Wasserversorgung der Tiere an. Es zeigt sich in diesen Daten eine hohe Individualität der Tiere. Versucht man zum Beispiel Grenzwerte zu definieren, welche die Kontrolle der täglichen Verzehrmenngen und Besuchshäufigkeiten der Tiere unterstützen sollen, ist dies auf Herdenebene nur schwer möglich. Beispielsweise fressen zwei Kühe am Tag annähernd die gleiche Frischmasse der Grundfütterration, zeigen jedoch ein sehr unterschiedliches Besuchsverhalten an den Trögen.

3 KuhDaM

Seit November 2009 ist das Kuhdatenmanagementsystem „KuhDaM“ in die Betriebsroutine des Versuchsbetriebes Frankenforst eingebunden. Dieses relationale Datenbanksystem ermöglicht die effiziente und widerspruchsfreie Speicherung großer Datenmengen und die bedarfsgerechte Bereitstellung unterschiedlichster Teilmengen der Daten für Benutzer und Anwendungsprogramme, um ein verlässliches Monitoring des Versuchsablaufes und eine zeitnahe Versuchsauswertung realisieren zu können. [BMMS10]. Die „KuhDaM“-Datenbank umfasst die Bereiche Stammdaten, Fütterung, Fruchtbarkeit, Milch und Gesundheit. Das Datenbankschema ist so konzipiert, dass zukünftige Entwicklungen in der Rinderhaltung (z.B. Erhöhung der Messwertdichte durch neue Sensortechniken) nachträglich abgebildet und einbezogen werden können. Vom Versuchsbetrieb Frankenforst werden folgende Daten (täglich, monatlich oder sporadisch anfallend) übernommen: Stammdaten, Standorte, Gesundheit, Milchleistung, Milchflusscharakteristika, Selektionen, Körperkondition, Aktivität, Futter- und Wasseraufnahme sowie –verhalten, Lebendgewicht, Daten der Milchleistungsprüfung, Futterzusammensetzung sowie projektabhängige Zwischenkontrollen. Die täglich anfallenden Daten werden jede Nacht aus der Prozesstechnik heraus automatisch in definierte Listen (ASCII-Format) geschrieben und in ein Exportverzeichnis kopiert. Der anschließende Datentransfer erfolgt verschlüsselt auf der Basis des SSH-Protokolls mit asymmetrischer Verschlüsselung auf einen Fileserver im Rechenzentrum der Universität Kiel; zurzeit beträgt der Umfang der automatisch erstellten und transferierten Daten ca. 10 MB je Tag. Das System „KuhDaM“, aktuell auf vier weiteren Milchviehversuchsbetrieben eingesetzt, gewährleistet die Verknüpfung und strukturierte Speicherung dieser Vielzahl von Datenquellen; die unterschiedlichsten Dateien werden jeden Morgen mit Hilfe entsprechender, automatisch angestoßener Schnittstellenprogramme unter Prüfung umfangreicher Plausibilitätsbedingungen in das System eingespielt. In den Fehlerprotokollen ausgewiesene Dateninkonsistenzen können dann unverzüglich im Dialog mit Mitarbeitern des Versuchsbetriebes analysiert und korrigiert werden. Nicht elektronisch vorliegende Daten (z.B. Futteranalysen und –rationen) werden mit Hilfe internetbasierter Oberflächen ma-

nuell und standortunabhängig eingegeben. Die Erfahrungen zeigen, dass grundsätzlich ein hoher Automatisierungsgrad in der Datenbereitstellung des Versuchsbetriebes anzustreben ist; alle (manuell oder elektronisch) erfassten Daten sollten in gleich bleibender Auflösung und Qualität und zeitnah zu klar definierten Zeitpunkten in elektronischer Form bereitgestellt werden. Der Zugriff auf die Datenbank erfolgt über internetbasierte Hilfsprogramme zur permanenten Überwachung von Versuchsdaten (tabellarisch oder grafisch) und über die Softwarepakete SAS und Excel unter Nutzung der ODBC-Schnittstelle. Innerhalb des Versuchsmonitorings werden z.B. aufgenommene Futtermengen oder tägliche Milchleistungen je Einzeltier oder Versuchsgruppe grafisch abgerufen und kontrolliert. Ebenso stehen herunterladbare, definierte Exportdateien für die Weiterverarbeitung durch den Nutzer zur Verfügung. Alternativ können direkt aus dem Statistikpaket SAS heraus zu analysierende Messwerte zusammen mit den relevanten Einflussfaktoren zusammengestellt und analysiert werden, um innerhalb weniger Stunden oder Tage nach Abschluss der Versuchsdatenerhebung fundierte Ergebnisse auf der Basis konsistenter und vollständiger Daten präsentieren zu können.

4 Datenstromanalyse

Die verschiedenen Messdaten, die vom Versuchsstall Frankenforst generiert und mit Hilfe von KuhDaM verwaltet werden, repräsentieren einen Strom von zeitgestempelten Tupeln, die mit herkömmlichen Methoden nur schwierig zu analysieren sind. Seit mehr als 10 Jahren wird in der Informatik an sogenannten Datenstromsystemen (DSMS) gearbeitet, die eine effiziente Bearbeitung solcher dynamischen Tupelströme erlauben. Es gibt mittlerweile eine Vielzahl von prototypischen Systemen (z.B. Aurora, STREAM oder TelegraphCQ [GÖ03]), mit deren Hilfe man Muster in Datenströmen – wie etwa große Kursschwankungen in einem Strom von Börsendaten oder bestimmte kritische Wertekombinationen in einem Strom von medizinischen Messdaten – effizient finden kann. Diese Muster werden mit Hilfe von gespeicherten Anfragen im DSMS bestimmt, die kontinuierlich neue Anfrageergebnisse liefern, sobald neue Stromdaten eintreffen. Allerdings eignen sich DSMS nicht zur dauerhaften Speicherung von Daten; historische Daten werden höchstens in den Zuständen von Operatoren im DSMS gehalten, und Anfragen über frühere Daten sind mit einem DSMS oft nicht möglich. Somit sind sie nur bedingt für die Analyse von Milchkuhdaten geeignet, in denen neben dem aktuellen Messdatenstrom insbesondere die historischen Messungen tierindividuell für die Analyse berücksichtigt werden müssen. Deshalb verfolgen wir das Ziel, bewährte Stromanalysetechniken direkt in den relationalen Datenbankkontext zu übertragen, um dort neben der etablierten Analyse historischer Daten eben auch die Analyse von dynamischen Stromdaten zu ermöglichen. Zu diesem Zweck werden insbesondere inkrementelle Methoden zur Änderungs-propagierung gezielt eingesetzt, um so den Einsatz von spezialisierten DSMS und die damit verbundenen Abstriche bzgl. der persistente Speicherung und der Historisierung von Daten zu vermeiden. Die wichtigste Methode stellt dabei der Magic Updates-Ansatz [BM04] dar, in dem die theoretischen Grundlagen für eine inkrementelle Verarbeitung von Stromdaten gelegt werden. Die prinzipielle Idee ist es, die aufwändige Komplettneuberechnung von SQL-Anfragen über Datenströmen zu vermeiden und nur die neuen Anfrageresultate zu bestimmen, die sich aus den neu eingegangenen Stromwerten ergeben. Dabei erlaubt diese Methode insbesondere eine effiziente Ver-

knüpfung dynamischer Stromdaten mit quasi statischen Anwendungsdaten, was mittels eines DSMS nur schwer möglich ist. Als Beispiel hierfür betrachten wir die Verschlechterung des Gangverhaltens einer an Mortellaro erkrankten Kuh. Am 248. Laktationstag wurde die Huferkrankung Mortellaro registriert. Die damit verbundene Verschlechterung des Gangverhaltens der Kuh in den nachfolgenden 77 Tagen führte gleichzeitig zu einer zeitweise deutlichen Erhöhung der Liegezeiten. Daneben erhöhte sich die durchschnittliche Menge der Wasseraufnahme. Um diese Abweichung von den normalen Durchschnittswerten dieser Kuh in der Laktationsperiode zeigen zu können, war es zunächst erforderlich, für die historischen Sensordaten repräsentative Mittelwerte zu bestimmen. Diese können nun mit den aktuellen Stromwerten verglichen werden, um so signifikante Abweichung möglichst zeitnah signalisieren zu können. Um repräsentative Mittelwerte zu erhalten, müssen Zeitintervalle in den historischen Daten gefunden werden, in denen vergleichbare Bedingungen für das Tier bestanden. So sind z.B. für eine repräsentative Bestimmung der durchschnittlichen Wasseraufnahme des Tieres die Außentemperaturwerte zu berücksichtigen. Nach der Analyse werden die Stromdaten selbst zu historischen Werten, die zur Anpassung der repräsentativen Mittelwerte herangezogen werden müssen. In dieser Weise werden Alterungsprozesse und die damit verbundenen physiologischen Änderungen der Kuh berücksichtigt.

5 Fazit

Die kontinuierliche Verarbeitung und Analyse der verschiedensten Datenmengen und -strukturen von immer neu entwickelten Sensortechniken ist eine große Herausforderung für milchviehhaltende Versuchsbetriebe. Neben der zunehmenden Vielfalt der Sensor- und Prozessordaten spielt auch die hohe Individualität der Datenausprägungen je Tier eine herausfordernde Rolle bei der Entwicklung von automatisierten Auswertungsverfahren. Die unterschiedlichen Anforderungen an die Informationssysteme auf Versuchsbetrieben definieren die Architekturen der Datenbanken, wobei sich das KuhDaM-Datenbanksystem als hilfreiches Instrumentarium herausgestellt hat. Mit Hilfe dieser relationaler Datenbanken auf Betriebsservern können die Informationen zusammengeführt werden, wodurch komplexe Analyse- und Auswertungsmethoden möglich sind. Aufgrund der datentechnischen Gegebenheiten auf Versuchsbetrieben wird dabei neben der bisherigen Analyse historischer Daten auch die Analyse von dynamischen Stromdaten eine wesentliche Rolle spielen.

Literaturverzeichnis

- [BM04] Behrend, A.; Manthey, R.: Update Propagation in Deductive Databases Using Soft Stratification. ADBIS: 22-36, LNCS 3255, Springer, 2004.
- [BMM10] Büscher, W.; Müller, U.; Müller, P.; Stamer, E.: Informations-management in der Milchviehhaltung - am Beispiel der Milchviehherde "Frankenforst" der Universität Bonn, 40. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 27.09.-01.10.2010, Leipzig, Germany, 2010; S.179-183.
- [GÖ03] Golab, L.; Özsu, M. T.: Issues in Data Stream Management. ACM SIGMOD Record 32 (2), 2003: S.5-14.