

Implementierung von partitionierenden Clusteralgorithmen zur automatisierten Verdichtung von Ortungsdaten

Frederik Kulpi, Bernhard Haidn

Institut für Landtechnik und Tierhaltung
Bayerische Landesanstalt für Landwirtschaft
Prof.-Dürrwaechter-Platz 2
85586 Poing
frederik.kulpi@lfl.bayern.de
bernhard.haidn@lfl.bayern.de

Abstract: Im Zuge der sich stetig vergrößernden Tierbestände in Milchviehställen wird ein kontinuierliches Erfassen von Tierbewegungen zunehmend erforderlich, um damit das individuelle Tierverhalten ableiten und deren tägliche Aktivitäten beschreiben zu können. Im Rahmen eines Versuchs wurde über einen mehrmonatigen Zeitraum in einem Laufstall ein Ortungssystem installiert, das in einem extrem großen Frequenzbereich operiert. Die Aufzeichnungsrates von etwa einer Position pro Tier und Sekunde hat zur Folge, dass die generierten Daten mit den heute verfügbaren Computern für die Informationsverarbeitung in Behörden nur bedingt geeignet sind. Daher wird das Erfasste unter Zuhilfenahme des partitionierenden Clusterungsalgorithmus von Lloyd, stallbereichsbezogen und vollautomatisiert vorverarbeitet.

1 Datenerfassung

Zur Datenerfassung werden Ultrabreitbandortungssensoren verwendet, die im Frequenzbereich sechs bis acht Gigahertz mit aktiven Transpondern kommunizieren und dabei Laufzeit- und Einfallswinkeldifferenzen zur Positionsbestimmung nutzen. Das sekundlich erfasste Koordinatentupel der einzelnen Transponder wird durch einen Zeitstempel erweitert und anschließend strukturiert in einer Datenbank abgespeichert.

Die zonenspezifische Verarbeitung der Daten setzt voraus, dass die Geometrie der sich im Stall befindlichen Objekte digital in Form von Polygonen vorliegt. Aufgrund einer bereits vorhandenen Konstruktionszeichnung des Versuchsstalls konnte auf eine manuelle Digitalisierung weitestgehend verzichtet werden. Die Abbildung 1 zeigt die Zonen und die Nummerierung der darin enthaltenen Stallobjekte, wie beispielsweise die einzelnen Liegeboxen oder Fressplätze.

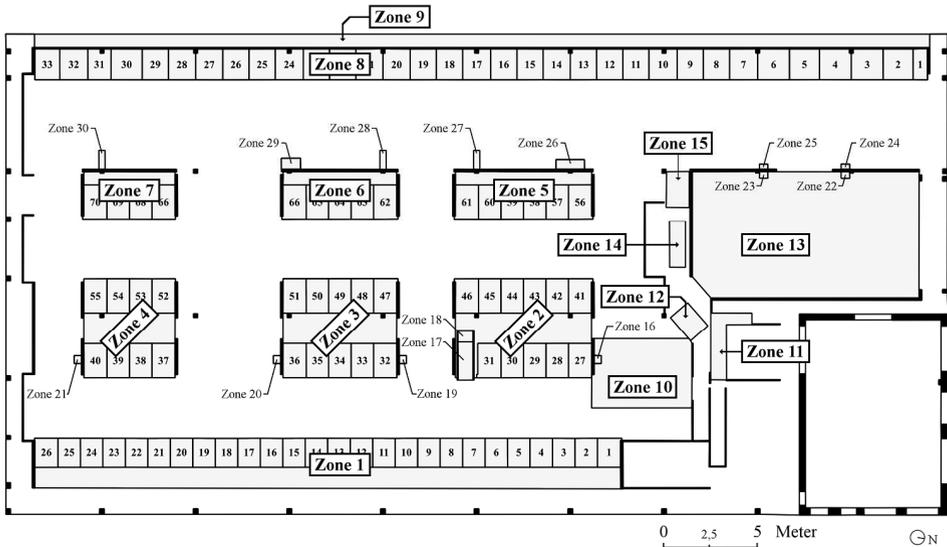


Abbildung 1: Clusterzonen und Raumindizes

2 Zonenbezogene Selektion

Zunächst wird jede der Zonen gesondert betrachtet und verarbeitet. Die Selektion der Eingangsdaten erfolgt auf der Grundlage des Jordanschen Kurvensatzes, der die Ränder eines Polygons in einen Innen- und einen Außenraum unterteilt [GP74]. Damit werden alle Elemente die sich innerhalb einer bestimmten Zone befinden selektiert und zur Weiterverarbeitung freigegeben.

3 Partitionierender Clusteralgorithmus

Bei partitionierenden Clusterungsverfahren gilt, dass die betrachteten Objekte selbstständig in eine endliche Menge von Gruppen eingeteilt werden. Der für die Reduktion der Positionsaufzeichnungen verwendete Algorithmus² ordnet bei dessen Initialisierung die Elemente hinsichtlich ihrer Nähe zu zufällig gewählten Zentren. Jedes neue Element wird dann der Gruppe zugeordnet, bei dem die geringste Erhöhung der Varianz zum Zentrum zu vernehmen ist. Anschließend wird das Zentrum der gebildeten Gruppe aktualisiert [LI57].

Das Ergebnis der Clustering ist eine Tabelle, die die statistischen Mittelwerte der Elemente der jeweiligen Gruppe beinhaltet. Die x- und y-Koordinate beschreibt hierbei das

¹ Die Elemente werden jeweils durch eine Position im zweidimensionalen Raum und durch die Zeit beschrieben.

² Zur Reduktion der Daten wird der partitionierende Algorithmus k-Means verwendet.

zweidimensionale Zentrum aller Elemente der Gruppe. Der gemittelte Zeitwert wird bei der weiteren Betrachtung nicht berücksichtigt, stattdessen werden die Elemente während des Clusterverfahrens kategorisiert, sodass sie nach Abschluss der Clustering zur Bestimmung der minimalen und maximalen Zeitwerte beziehungsweise zur Berechnung der Verweildauer der jeweiligen Gruppe herangezogen werden können.

Die Anzahl der Zentren und damit verbunden die generierte Datenmenge ist abhängig von der Vorgabe der Anzahl der zu bildenden Gruppen. Die ersten Auswerterversuche haben gezeigt, dass die Betrachtung von längeren Zeitspannen und mehreren Tieren eine zusätzliche Reduktion der Daten erfordert.

4 Datenfusion basierend auf zeitlichen Nachbarschaften

Als Vorgabe wird hierbei im Gegensatz zur partitionierenden Clustering – bei der eine Anzahl an zu bildenden Gruppen benötigt wird – ein zeitlicher Schwellenwert definiert. Von den zeitlich sortierten Gruppen wird fortlaufend überprüft, ob die Summe des zeitlichen Maximums und dem definierten Schwellenwert, das zeitliche Minimum der nachfolgenden Gruppe überschreitet. Ist dies der Fall, so werden die beiden Gruppen zusammengefasst, andernfalls bleiben beide Gruppen unverändert. Beim Zusammenfassen werden aus den Elementen der beiden Gruppen, das zeitliche Minimum und das zeitliche Maximum extrahiert. Ergänzend erfolgt die Neuberechnung des räumlichen Zentrums auf der Grundlage der Koordinatentupel der beiden betrachteten Gruppen.

Der Abbildung 2 sind die beiden beschriebenen Verdichtungsstufen (Abbildungen 2b und 2c) ausgehend von den Positionsaufzeichnungen (Abbildung 2a) zu entnehmen.

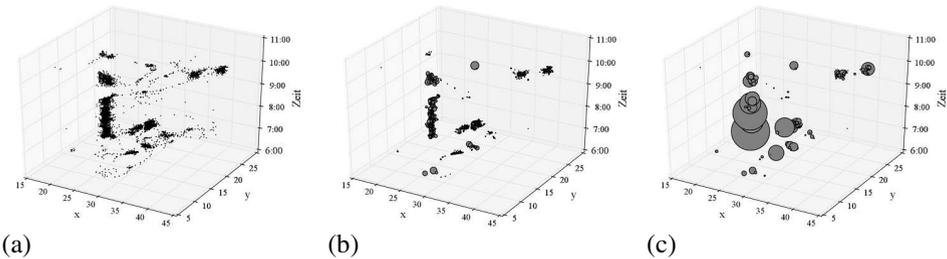


Abbildung 2: Aufgezeichnete und nachverarbeitete Positionen eines Transponders in einem Zeitraum von fünf Stunden: (a) Rohdaten; (b) Daten nach partitionierender Clustering; (c) Daten nach Fusionierung.

5 Räumliche Indizierung der Zentren

Die schnelle Lokalisierung der verdichteten Daten für die Ableitung des Tierverhaltens wird über einen räumlichen Index erzielt. Hierbei wird das stallobjektbeschreibende Attribut mit den Positionen der errechneten Clusterzentren basierend auf einer räumlichen Beziehung verknüpft. Jede Position wird indiziert und kann für bereichsspezifische Analysen wie zum Beispiel die Beschreibung von Bereichsübergängen, die Anzahl der Fressperioden oder zur Ermittlung der bevorzugten Liegebucht verwendet werden.

6 Datenanalyse

Die ersten Auswertungsergebnisse mit den verdichteten Daten deuten darauf hin, dass die bereichsbezogenen Aufenthaltsdauerberechnungen mit den in der Literatur zu findenden Angaben übereinstimmen [Ho09]. Die Ermittlung der Aufenthaltsperioden, insbesondere im Liege- und im Fressbereich, ist beim nachträglichen Kontrollieren mit Videoaufzeichnungen größtenteils deckungsgleich. Darüber hinaus lassen sich die Daten für die Beschreibung von Zonenübergängen verwenden, auch wenn hierbei noch ein Anpassungsbedarf bei nahe beieinanderliegenden Zonen notwendig ist.

7 Zusammenfassung

Der beschriebene Versuchsaufbau und die Methode zur Vereinfachung der Daten auf die wesentlichen Kerninformationen zeigen, dass sich Ultrabreitbandortungssysteme durchaus eignen, automatisiert das Verhalten von Tieren in Laufställen abzuleiten und dass sie ein großes Potenzial haben, das Tierherden- aber auch das individuelle Tierverhalten zu optimieren. Es ist in einem weiteren Schritt denkbar durch aktive Eingriffe die lokalisierten Tiere beim Eintreten bzw. Verlassen von vordefinierten Gebieten gezielt mit entsprechenden Impulsgebern zu steuern.

Literaturverzeichnis

- [GP74] Guillemin, V.; Pollock, J.: Differential Topology. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.
- [Ho09] Hoy, S.: Nutztierethologie. Verlag Eugen Ulmer, Stuttgart, 2009.
- [LI57] Lloyd, S.: Least Squares Quantization in PCM. Technical Report, Bell Telephone Laboratories, Murray Hill, New Jersey, 1957.