

Data Mining in der Landwirtschaft

Vortrag zum Doktorandentag

Georg Ruß, IWS

02. Februar 2010

Gliederung

Motivation

Daten

Wissenschaftliche Fragestellungen

Zusammenfassung

Data Mining in der Landwirtschaft

Grundidee

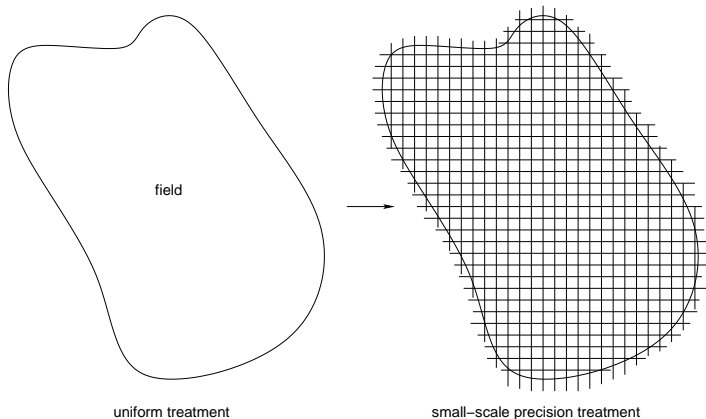


Abbildung: *Präzisionslandwirtschaft* = datengetriebene Herangehensweise

Präzisionslandwirtschaft

Grundlagen / Motivation

- ▶ Ursprünge:
 - ▶ günstige Datenquellen
 - ▶ GPS-Technologie für räumliche Daten
 - ▶ Feldunterteilung in Teilflächen
 - ▶ teilflächenspezifische Bewirtschaftung
- ▶ große Datenmengen
- ▶ Nutze Data Mining, um u.a.:
 - ▶ Effizienz zu erhöhen,
 - ▶ Ertrag zu erhöhen,
 - ▶ nützliche Daten herauszufiltern.

Gliederung

Motivation

Daten

Stickstoffdünger und Ertrag

Vegetation und elektrische Leitfähigkeit

Wissenschaftliche Fragestellungen

Zusammenfassung

Stickstoffdünger und Ertrag

- ▶ Stickstoffdünger
 - ▶ Menge kann beim Düngen relativ einfach gemessen werden
 - ▶ Dünger wird zu drei Zeitpunkten während der Vegetationsphase ausgebracht
 - ▶ Drei Attribute: N_1 , N_2 , N_3
- ▶ Ertrag 2007/2008
 - ▶ Ertrag wird bei der Ernte vollautomatisch geocodiert aufgezeichnet
 - ▶ Daten aus 2007 (Vorjahr) und 2008 (aktuelles Jahr)
 - ▶ Zwei Attribute: Yield07, Yield08

Vegetation und elektrische Leitfähigkeit

- ▶ REIP (Red Edge Inflection Point)
 - ▶ Wert der zweiten Ableitung des spektralen Rotbereichs eines Bildes
 - ▶ kann aus Luftbildern oder Satellitenfotos gewonnen werden
 - ▶ höherer Wert bedeutet mehr Vegetation
 - ▶ wird vor der Ausbringung von N_2 und N_3 gemessen
 - ▶ Zwei Attribute: $REIP_{32}$, $REIP_{49}$
- ▶ Elektromagnetische Leitfähigkeit
 - ▶ scheinbare Bodenleitfähigkeit wird per Sensor gemessen, bis zur Tiefe von etwa 1,50m
 - ▶ starke Korrelation mit verschiedenen Bodeneigenschaften wird erwartet
 - ▶ kommerzielle Sensoren sind verfügbar
 - ▶ Ein Attribut: EC_a

Kartierung

kleinräumig, hochaufgelöst

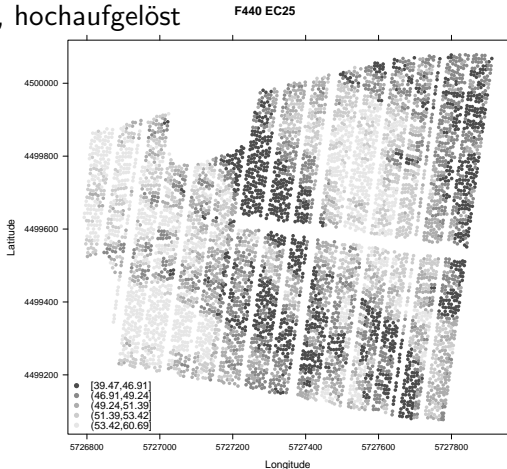


Abbildung: Scheinbare Bodenleitfähigkeit für Feld F440

Kartierung

kleinräumig, hochaufgelöst



Abbildung: Stickstoffdünger (N1) für Feld F440

Gliederung

Motivation

Daten

Wissenschaftliche Fragestellungen

Ertragsvorhersage

Interessantheit von Datenquellen

Management-Zonen

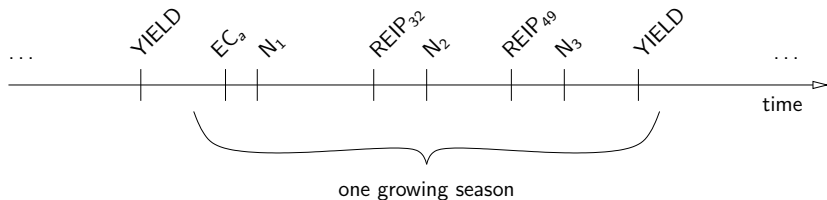
Zusammenfassung

Fragestellungen

Es besteht wissenschaftlicher und praktischer Bedarf an der Lösung folgender spezifischer Fragestellungen:

- ▶ Ertragsvorhersage
- ▶ Nützlichkeit zusätzlicher Daten
- ▶ Bestimmung von Management-Zonen

Ertragsvorhersage



- ▶ Versuche, den diesjährigen Ertrag basierend auf den vorhandenen Datenattributen vorherzusagen!
- ▶ bisher nur als Regressionsaufgabe unter Vernachlässigung der räumlichen Komponente gelöst

Ertragsvorhersage

Problem:

- ▶ nicht-räumliche Regressionsmodelle werden aufgrund der vorliegenden räumlichen Autokorrelation *überlernt*
- ▶ der Vorhersagefehler wird unterschätzt

Lösung:

- ▶ Zerlegung des Feldes in Teilflächen
- ▶ Durchführen einer räumlichen Kreuzvalidierung auf diesen Teilflächen

Ertragsvorhersage

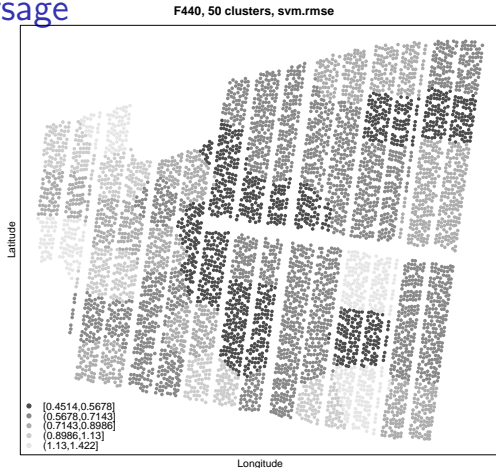


Abbildung: Räumliche Kreuzvalidierung, 50 Teilflächen, Support Vector Regression

Interessantheit von Datenquellen

Wie nützlich sind zusätzlich eingeführte Sensordaten im Hinblick auf die Ertragsvorhersage?

- ▶ Fragestellung taucht bei der Entwicklung neuer Sensoren oder Datenquellen auf
- ▶ neue Datenquellen sollen so schnell wie möglich evaluiert werden

Interessantheit von Datenquellen

- ▶ nutze das entwickelte räumliche Ertragsvorhersagemodell:
 - ▶ im Zusammenhang mit Standard-Ansätzen zur Merkmalsauswahl (*feature selection*)
 - ▶ im Zusammenhang mit neuentwickelten Ansätzen zur Variablen-Vertauschung (*variable permutation*)
- ▶ vergleiche beide Ansätze und evaluiere diese in der Praxis

Management-Zonen

- ▶ Gibt es Teilflächen, die aufgrund der Datenlage besonders hervorstechen?
- ▶ Gibt es dort bisher unbekannte Zusammenhänge zwischen Datenattributen?
- ▶ praktisch und wissenschaftlich sehr relevante Fragestellung im Bereich der Grunddüngung
- ▶ Beantwortung der obigen Fragen aus Data-Mining-Sicht möglich

Management-Zonen

- ▶ teile Feld in Teilflächen (k-means, wie vorher)
- ▶ bestimme Nachbarschaft von Teilflächen
- ▶ verschmelze benachbarte (ähnliche) Teilflächen
- ▶ stoppe, wenn festgelegte Anzahl von Zonen erreicht ist
- ▶ werte entstandene Zonen aus, beispielsweise mit Assoziationsregeln oder Konfigurationen

Management-Zonen

F440, hypothetical management zones

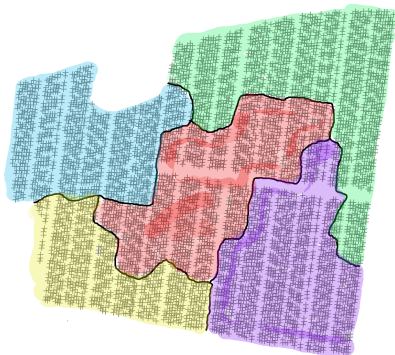


Abbildung: Vorschlag: konsekutives Verschmelzen der Teilflächen zu einer festgelegten Anzahl von Zonen

Gliederung

Motivation

Daten

Wissenschaftliche Fragestellungen

Zusammenfassung

Zusammenfassung (0)

- ▶ Arbeit mit in der Landwirtschaft anfallenden kleinräumigen / hochaufgelösten Daten
- ▶ Behandlung dieser Daten mit Methoden, die die räumliche Komponente mit einbeziehen (räumliche Autokorrelation!)
- ▶ Bearbeitung drei wissenschaftlich und praktisch aktueller Fragestellungen, die aus Data-Mining-Sicht und mit Data-Mining-Mitteln beantwortet werden können (...)

Zusammenfassung (1)

(...)

- ▶ Ertragsvorhersage
 - ▶ neuentwickelter Ansatz
 - ▶ modellübergreifend, statistisch gültig
- ▶ Nützlichkeit hinzukommender Datenquellen
 - ▶ *feature selection* (alt)
 - ▶ *variable permutation* (neu)
- ▶ Erstellung von Management-Zonen aus Datenmaterial
 - ▶ neuer Ansatz zum räumlichen Clustern
 - ▶ derzeit in der Evaluation

Publikationen

- ▶ Ertragsvorhersage: [1, 5, 6, 9, 10] (publiziert)
- ▶ Interessantheit: [3] (publiziert), [2, 4, 7] (eingereicht)
- ▶ Management-Zonen: [8] (publiziert) + zwei Deadlines bis März 2010

Abschluß...

Fragen & Antworten

- ▶ Research Blog: <http://research.geogruss.de/>
- ▶ “Data Mining in Agriculture” Workshop 2010:
<http://dma2010.de/>

Bibliography I



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 24–37. Springer, July 2009.



Georg Ruß and Alexander Brenning.

Data mining in precision agriculture: Management of spatial information.

In *Proceedings of IPMU'2010*, pages –, July 2010.
submitted for review.



Georg Ruß and Rudolf Kruse.

Feature selection for wheat yield prediction.

In Tony Allen, Richard Ellis, and Miltos Petridis, editors, *Research and Development in Intelligent Systems XXVI, Incorporating Applications and Innovations in Intelligent Systems XVII*, volume 26 of *Proceedings of AI-2009*, pages 465–478. BCS SGAI, Springer, January 2010.



Georg Ruß and Rudolf Kruse.

Regression models for spatial data: An example from agriculture.

In Petra Perner, editor, *Advances in Data Mining*, Lecture Notes in Computer Science, pages –. Springer, July 2010.
submitted for review.



Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.

Estimation of neural network parameters for wheat yield prediction.

In Max Bramer, editor, *AI in Theory and Practice II*, volume 276 of *Proceedings of IFIP-2008*, pages 109–118. Springer, July 2008.

Bibliography II



Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.

Optimizing wheat yield prediction using different topologies of neural networks.

In J. Verdegay, M. Ojeda-Aciego, and L. Magdalena, editors, *Proceedings of IPMU-08*, pages 576–582. University of Málaga, June 2008.



Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.

Using advanced regression models for determining optimal soil heterogeneity indicators.

In Hermann Locarek-Junge and Claus Weihs, editors, *Classification as a Tool for Research, Proceedings of the 11th IFCS Biennial Conference and 33rd Annual Conference of the Gesellschaft für Klassifikation e.V.*, Studies in Classification, Data Analysis, and Knowledge Organization, Berlin-Heidelberg-New York, mar 2009. Springer.
to appear.



Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.

Visual data mining of agriculture data.

In Petra Perner, editor, *Machine Learning and Data Mining in Pattern Recognition, 6th International Conference, MLDM 2009, Leipzig, Germany, July 18-20, 2009, Poster Proceedings*, Machine Learning and Data Mining in Pattern Recognition, MLDM, pages 30–44, Leipzig, July 2009. ibai publishing.



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.

Bibliography III



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.