

Mobile kontextadaptive Dienste und Systeme

Data Mining & Analyse von Sensordaten

Felix Ortmann

Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Department Informatik
Zentrum für Verteilte Informations- und Kommunikationssysteme
Arbeitsbereich Verteilte Systeme und Informationssysteme

10. Dezember 2016



Gliederung

- 1 Einführung
- 2 Context & Sensoren
- 3 Datamining
- 4 Real World Mobile Sensor Data Mining
- 5 Literatur

Motivation

- Durch mobile Geräte werden zunehmend mehr Daten erhoben
- „Context“ Wissen
- Unterschiedlichste Nutzungsszenarien
- Nutzbar für wen, wozu?
- Besondere Verfahren notwendig, um nutzbare Informationen aus Datenflut zu extrahieren

Zielsetzung

- Datamining und Machinelearning mit/auf mobilen Geräten
- Herausstellung diverser Problematiken bezogen auf den mobilen Bereich
- Datamining, Techniken und Algorithmen
- „Real World“ Beispiel

Context & Sensoren

- 1 Einführung
- 2 Context & Sensoren**
 - Context
 - Context Awareness
 - Context Challenges
- 3 Datamining
- 4 Real World Mobile Sensor Data Mining
- 5 Literatur

Context

„Context“ [Rahmati et al., 2015]

...eines Gerätes oder einer Anwendung:

Wissen über Umgebung – intern und extern

Datenerhebung aus verschiedenen Quellen erlaubt Rückschluss auf
Context

Context

Unterscheidung verschiedener Contexte [Rahmati et al., 2015]

Nutzungsdaten (Beispiele)

- Besuchte Webseiten
- Getätigte Anrufe
- Installierte Apps, Nutzungsstatistiken

Sensordaten (Beispiele)

- Helligkeit
- Beschleunigung
- GPS

Context Awareness [Rahmati et al., 2015] [Orsini et al., 2016]

„Context Awareness“

- Gerät / Applikation erhebt und nutzt Context Daten
- App-Hersteller nutzen Context Daten

„Context Awareness“ ist Nutzbar für:

- Userfeatures (zB. „Recommendations“, Lokalisierung, Health-Care, „Ambient Display“)
- App-Hersteller (zB. Verbesserungen der Produkte, gezielte Werbung)

Context Challenges

Challenges [Rahmati et al., 2015] [Orsini et al., 2016]

Mobile-Charakteristika erschweren die Nutzung der Context Daten

- Heterogene Datenquellen (sehr viele Sensoren, zB. Auto)
- Limitierte Ressourcen
 - Stromversorgung (Akku, Batterie)
 - Rechen-Ressourcen
 - Hoher Verbrauch durch sowohl Sensoren als auch Analyse der Daten!

„Datamining to the rescue“

Datamining

1 Einführung

2 Context & Sensoren

3 Datamining

- Mining „auf“ Mobilgeräten
- Mining Prozesse
- Data Preparation
- Learning Strategien
- Linear Regression
- K-Nearest-Neighbor (KNN)
- Support Vector Machine – SVM
- K-Means Clustering

Mining „auf“ Mobilgeräten

Zwei Kategorien von mobilem Mining [Gaber et al., 2014]

„Mobile Interface“

- Entfernter Mining Prozess (zB. Cloud)
- Mobiles Gerät als Zugang und Datenlieferant

„On-board Mining“

- Mining Prozess auf dem Gerät
- Visualisierung (oder andere Auswertungen) auch auf Gerät
- → wachsende Leistungs-Anforderungen an mobile Geräte!

Mining Prozesse – KDD

Knowledge Discovery in Databases – KDD [Fayyad et al., 1996]

- Selection
- Pre-processing
- Transformation
- Data Mining
- Interpretation/Evaluation

Mining Prozesse – KDD

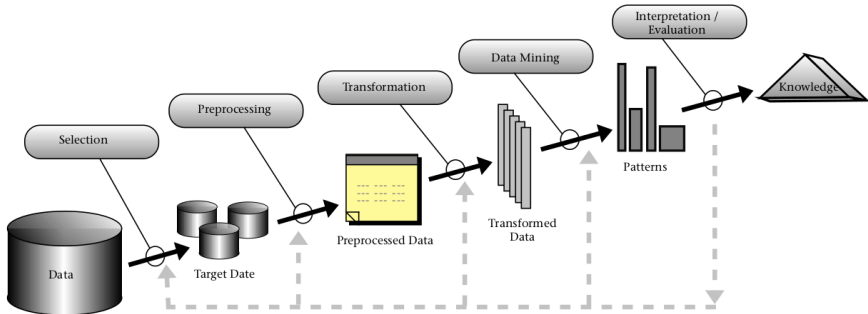


Abbildung: [Fayyad et al., 1996]

Mining Prozesse – CRISP

Cross Industry Standard Process for Data Mining – CRISP [Shearer, 2000]

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Mining Prozesse – CRISP

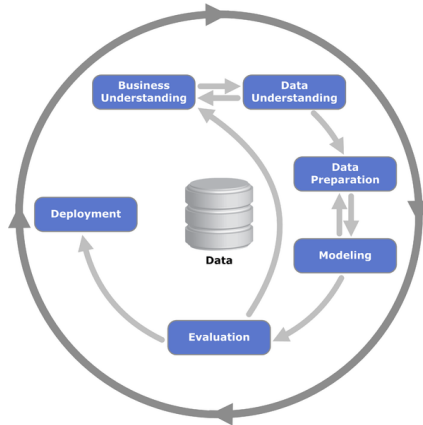


Abbildung: https://commons.wikimedia.org/wiki/File%3ACRISP-DM_Process_Diagram.png

Data Preparation

„Sehr aufwändiger, wenn nicht aufwändigster Teil-Prozess!“

[Gaber et al., 2014]

- Datenerhebung: Sammeln aus heterogenen Quellen
- Unifikation: Vereinheitlichen, soweit möglich
 - Formate bzw Attribute (ISO Standards, JSON, XML...)
 - „Sparse Data“ kann entstehen
- „Outlier“ Entfernen (nicht immer sinnvoll)
- „Preaggregation“ – zB. Mittelwerte verwenden (nicht immer sinnvoll)

Learning Strategien [Ian H. Witten and Hall, 2011]

„Supervised Learning“

- Klassifizierung *vom Menschen / durch Datenpunkte* vorgegeben
- Input und (gewünschter) Output bekannt
- Algorithmen *verwenden* Klassifikator
- Trainingsdaten & Testdaten

„Unsupervised Learning“

- Klassifizierung *vom Algorithmus* bestimmt
- Nur Input bekannt
- Output des Algorithmus muss Mensch zufrieden stellen

Linear Regression – [Ian H. Witten and Hall, 2011]

Supervised → Klassifikation durch aufgestellte Formel anhand der Attribute bereits bekannter Punkte

- Klasse \approx lineare Kombination der Attribute (des zu klassifizierenden Punktes)
- Algorithmus lernt, Attribute zu wichten
- $f(x) = w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k$
 - $f(x)$: Vorhersage-Funktion der Klasse
 - a_k : Bekannte Attribute
 - w_k : Wichtung der Attribute

K-Nearest-Neighbor (KNN)

Supervised → Klassifikation durch bereits bekannte Klassifikation der Nachbarn [Ian H. Witten and Hall, 2011]

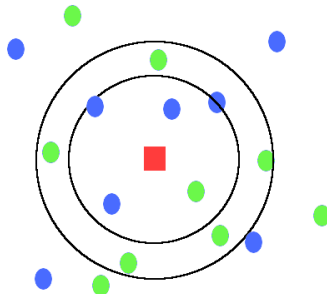


Abbildung: K-Nearest-Neighbor, vereinfachte Darstellung. Klassifikation der Roten Box: Wähle $K = 5 \rightarrow \text{blau}$. Wähle $K = 10 \rightarrow \text{grün}$

Support Vector Machine – SVM –

[Ian H. Witten and Hall, 2011]

Supervised → Klassifikation durch bereits bekannte Klassifikation anderer Punkte

- Betrachtet eine Menge „kritischer Punkte“
- Punkte so weit wie möglich von einander entfernt unterteilen
- „Hyperplane“: (Lineare) Unterscheidungsfunktion, eine Dimension geringer als Features in Punktwolke
- → Binärer (linearer) Klassifikator

Support Vector Machine – SVM

B klassifiziert „am besten“:

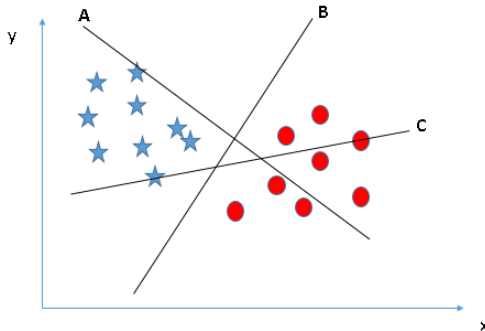


Abbildung: Online Material von [Ray, 2015],
[https://www.analyticsvidhya.com/blog/2015/10/
understaing-support-vector-machine-example-code/](https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/)

Support Vector Machine – SVM

Und nun?

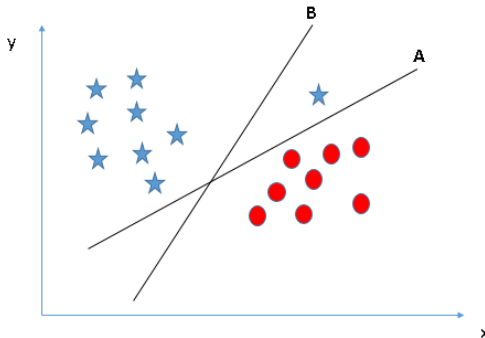


Abbildung: Online Material von [Ray, 2015],
[https://www.analyticsvidhya.com/blog/2015/10/
understaing-support-vector-machine-example-code/](https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/)

K-Means Clustering

Unsupervised → Einzelne Punkte haben keine bekannte Klassifikation
[Ian H. Witten and Hall, 2011]

```
def K-Means(K: int, toCluster: Set):  
    Place K centroids randomly  
    repeat until convergence:  
        for each Point in 'toCluster':  
            find closest centroid  
            add point to cluster  
        for each cluster in 1..K:  
            new centroid: mid of all points
```

[Lavrenko, 2013]

K-Means Clustering

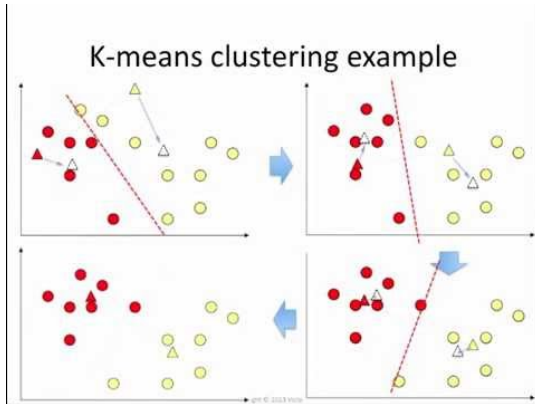


Abbildung: Online Material von [Lavrenko, 2013],
https://i.ytimg.com/vi/_aWzGGNrcic/hqdefault.jpg

Real World Mobile Sensor Data Mining

- 1 Einführung
- 2 Context & Sensoren
- 3 Datamining
- 4 Real World Mobile Sensor Data Mining**
 - Ein Markt Entsteht
 - iOS Health
- 5 Literatur

Ein Markt Entsteht – [Lockhart, 2011]

- Menschliches Verhalten in Daten messbar? – Universitärer Interessenanstieg an dem Thema
- Forschungsprojekte wie zB. „WISDM Lab“ Wireless Sensor Data Mining Laboratory entstehen¹
- Mehr Wissen über Nutzer – monetäre Interessen wachsen
- Neue Möglichkeiten, Produkte & Anwendungen zu verbessern

¹Fordham University, Bronx, NY <http://www.cis.fordham.edu/wisdm/>

iOS Health

„The all-new Health app has been redesigned to make it easier to learn about your health and start reaching your goals. It consolidates health data from iPhone, Apple Watch, and third-party apps you already use, so you can view all your progress in one convenient place. And it recommends other helpful apps to round out your collection — making it simpler than ever to move your health forward.” [Apple, 2016b]

iOS Health

Aus dem iOS Security Whitepaper [Apple, 2016c] – HealthKit aggregiert...

- „Health Data“: *Größe, Gewicht, vergangene Distanz, Blutdruck uvm.*
- „Management Data“: *Zugriffsberechtigungen, Angeschlossene Geräte, „Scheduling Daten“ wenn Apps gestartet werden uvm.*

Aus der Apple HealthKit API Reference [Apple, 2016a]:

- „Store“ in der Cloud
- Gerät überträgt (Sensor-) Daten

iOS Health

Brainstorming: was passiert bei HealthKit alles?

„Context“ umfasst...

- Biologische Nutzer Daten

iOS Health

Brainstorming: was passiert bei HealthKit alles?

„Context“ umfasst...

- Biologische Nutzer Daten
- Persönliche Nutzer Daten

iOS Health

Brainstorming: was passiert bei HealthKit alles?

„Context“ umfasst...

- Biologische Nutzer Daten
- Persönliche Nutzer Daten
- Gerät der Anwendungsausführung (Phone, Uhr etc.)

iOS Health

Brainstorming: was passiert bei HealthKit alles?

„Context“ umfasst...

- Biologische Nutzer Daten
- Persönliche Nutzer Daten
- Gerät der Anwendungsausführung (Phone, Uhr etc.)
- Daten anderer Apps auf dem Gerät

iOS Health

Brainstorming: was passiert bei HealthKit alles?

„Context“ umfasst...

- Biologische Nutzer Daten
- Persönliche Nutzer Daten
- Gerät der Anwendungsausführung (Phone, Uhr etc.)
- Daten anderer Apps auf dem Gerät
- Geografische Bewegungsdaten des Geräts / Nutzers

iOS Health

Brainstorming: was passiert bei HealthKit alles?

„Context“ umfasst...

- Biologische Nutzer Daten
- Persönliche Nutzer Daten
- Gerät der Anwendungsausführung (Phone, Uhr etc.)
- Daten anderer Apps auf dem Gerät
- Geografische Bewegungsdaten des Geräts / Nutzers
- ...

iOS Health

Die Nutzbarmachung dieses Contexts wird als API Feature für Developer angeboten.

iOS Health

Die Nutzbarmachung dieses Contexts wird als API Feature für Developer angeboten.

Ist das eine „biologische Query-API“ auf einem Menschen?

Habt ihr Fragen?

Danke für eure Aufmerksamkeit! :)

Literatur

- 1 Einführung
- 2 Context & Sensoren
- 3 Datamining
- 4 Real World Mobile Sensor Data Mining
- 5 Literatur**

|



Apple, I. (2016a).

Apple api reference.

Accessed: 2016-12-07.



Apple, I. (2016b).

ios health apple.

Accessed: 2016-12-07.



Apple, I. (2016c).

ios security whitepaper.

Accessed: 2016-12-07.

II

 Dinh, H. T., Lee, C., Niyato, D., and Wang, P. (2013).

A survey of mobile cloud computing: architecture, applications, and approaches.

Wireless communications and mobile computing, 13(18):1587–1611.

 Do, T.-M.-T. and Gatica-Perez, D. (2010).

By their apps you shall understand them: mining large-scale patterns of mobile phone usage.

In *The 9th International Conference on Mobile and Ubiquitous Multimedia*.

 Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996).

From data mining to knowledge discovery in databases.

AI magazine, 17(3):37.

III



Gaber, M. M., Gomes, J. B., and Stahl, F. (2014).

Pocket data mining.

Big Data on Small Devices. Series: Studies in Big Data.



Ian H. Witten, E. F. and Hall, M. A. (2011).

Data Mining: Practical Machine Learning Tools and Techniques (Third Edition).

The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, Boston, third edition.



Lavrenko, V. (2013).

Dr. victor lavrenko, university of edinburgh, online lectures.

Accessed: 2016-12-07.

IV



Lockhart, J. (2011).

Mobile sensor data mining.

Fordham Undergraduate Research Journal, 1:67–68.



Orsini, G., Bade, D., and Lamersdorf, W. (2016).

Generic context adaptation for mobile cloud computing environments.

Procedia Computer Science, 94:17 – 24.

The 11th International Conference on Future Networks and Communications (FNC 2016) / The 13th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2016) / Affiliated Workshops.

V



Rahmati, A., Shepard, C., Tossell, C., Zhong, L., and Kortum, P. (2015).

Practical context awareness: Measuring and utilizing the context dependency of mobile usage.

IEEE Transactions on Mobile Computing, 14(9):1932–1946.



Ray, S. (2015).

Analytics vidhya – understanding support vector machine algorithm from examples (along with code).

Accessed: 2016-12-07.



Shearer, C. (2000).

The crisp-dm model: the new blueprint for data mining.

Journal of data warehousing, 5(4):13–22.