

Lecture notes

Knowledge Discovery in Databases

Summer Semester 2012

Lecture 11: Q & A - What's next

Lecture: Dr. Eirini Ntoutsi

Tutorials: Erich Schubert

[http://www.dbs.ifi.lmu.de/cms/Knowledge_Discovery_in_Databases_I_\(KDD_I\)](http://www.dbs.ifi.lmu.de/cms/Knowledge_Discovery_in_Databases_I_(KDD_I))

Q&A

- Goals
 - To offer an overview of the KDD and DM areas
 - To introduce basic DM tasks namely
 - o Frequent itemsets & association rules mining
 - o Classification
 - o Clustering
 - o Outlier detection
 - To introduce basic algorithms for each task, their pros and cons
- Still, several issues left uncovered ☹
 - Regression
 - Evaluation of DM models
 - Dealing with data changes
 - Online mining/ stream mining
 - And of course, not all methods were covered in all details, e.g., neural nets

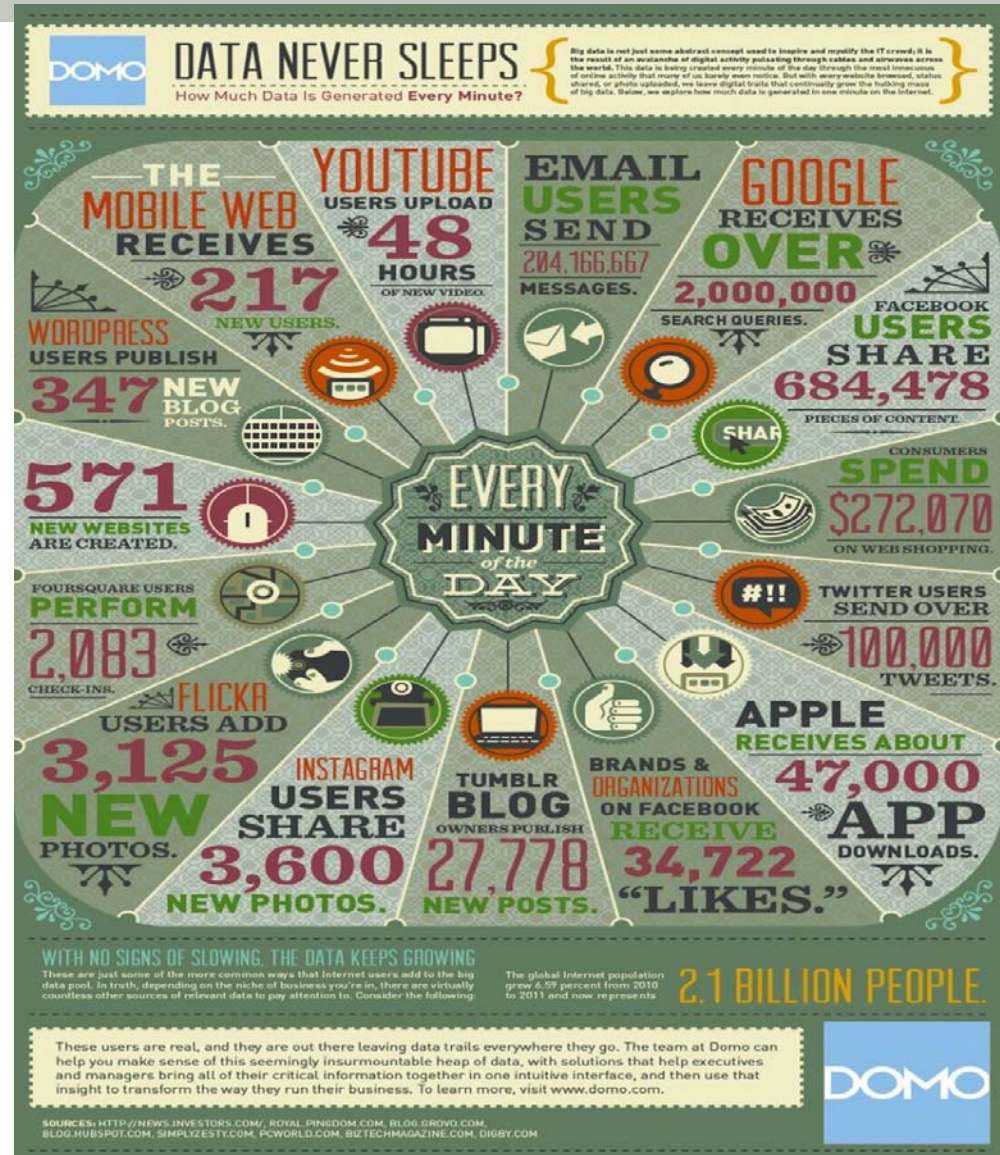
Datum	Vorlesung	Datum	Übung
17.04.2012	Lecture 1: Introduction PDF	-	-
24.04.2012	Lecture 2: Data preprocessing / Feature spaces PDF	26.04.2012	Blatt 01 PDF
01.05.2012	entfällt (Feiertag)	03.05.2012	entfällt
08.05.2012	Lecture 3: Association Rules and Frequent Itemsets Mining PDF (updated: 08.05.2012)	10.05.2012	Blatt 02 PDF
15.05.2012	Lecture 4: Classification PDF (updated: 15.05.2012)	17.05.2012 (Feiertag)	Blatt 03 PDF (Übung entfällt)
22.05.2012	Lecture 5: Classification PDF (updated: 22.05.2012)	24.05.2012	Blatt 04 PDF
29.05.2012	entfällt	31.05.2012	Besprechung Blatt 3+4
05.06.2012	Lecture 6: Classification PDF (updated: 05.06.2012)	07.06.2012 (Feiertag)	Blatt 05 PDF (Übung entfällt)
12.06.2012	Lecture 7: Clustering PDF (updated 12.06.2012)	14.06.2012	Blatt 06 PDF
19.06.2012	Lecture 8: Clustering PDF (updated 19.6.2012)	21.06.2012	Blatt 07 PDF
26.06.2012	Lecture 9: Clustering PDF	28.06.2012	Blatt 08 PDF
03.07.2012	Lecture 10: Outlier Detection PDF (updated 3.7.2012)	05.07.2012	Blatt 09 PDF
10.07.2012	Q&A	12.07.2012	Q&A
14.07.2012	Klausur (Exam)		



What's next?

Is Data Mining important nowadays?

- Huge amounts of data are generated nowadays.
- Data Mining and Knowledge Discovery are required in order to exploit all this wealth of information.



- Algorithmen und Datenstrukturen
- Anfragebearbeitung und Indexstrukturen in Datenbanksystemen
- Datenbanksysteme I
- Datenbanksysteme II
- Einführung in die Informatik: Systeme und Anwendungen
- Einführung in die Programmierung
- Geo-Informationssysteme
- Knowledge Discovery in Databases I (KDD I)
- Knowledge Discovery in Databases II (KDD II)
- Maschinelles Lernen und Data Mining
- Spatial, Temporal and Multimedia Databases (STMD I)
- Spatial, Temporal and Multimedia Databases II (STMD II)
- Managing Massive Multiplayer Online Games

Vorlesung		Übung	
Datum	Thema	Datum	Blatt
18.04.12	entfällt	20.04.12	entfällt
25.04.12	Einführung	27.04.12	entfällt
02.05.12	Das Perceptron / Wiederholung: Lineare Algebra / Lineare Regression	04.05.12	optional: R -Einführung (Sources)
09.05.12	Basisfunktionen / Neuronale Netze	11.05.12	Blatt 1 Lösungen: Lineare Algebra , Perceptron
16.05.12	Rekurrente Neuronale Netze und Zeitreihen	18.05.12	Blatt 2 Lösung: Lineare Regression ,
23.05.12	Kerne / Wiederholung: Wahrscheinlichkeitslehre	25.05.12	Blatt 3
30.05.12	Hauptkomponentenanalyse (Zusammenfassung)	01.06.12	Blatt 4 Lösung: Standard-normalverteilung ,
06.06.12	Reinforcement Learning, Teil 2	08.06.12	Blatt 5 (Daten: numberMatrix.RTable Lösung: PCA)
13.06.12	Frequentistische Statistik und Bayes'sche Statistik / Lineare Klassifikatoren	15.06.12	Blatt 6 (Daten: soccerBall , facesEasy ; Lösung)
20.06.12	Fortsetzung: Lineare Klassifikatoren / Support Vector Machines	22.06.12	Blatt 7 (Daten: body_sizes , bayesianData R-Lösungen zu 7-1 , 7-2 und 7-3)
27.06.12	Modellvergleich	29.06.12	Blatt 8 R-Lösung zu 8-1
04.07.12	Instanzbasiertes Lernen	06.07.12	Blatt 9
11.07.12	Bayes'sche Netze	13.07.12	Wiederholung + Fragen
18.07.12	Klausur	20.07.12	Klausurbesprechung

Datum	Vorlesung	Datum	Übung
18.10.2011	Kapitel 1: Überblick Kapitel 2: Feature Selection und Feature Reduction (Teil 1)	21.10.2011	entfällt
25.10.2011	Kapitel 2: Feature Selection und Feature Reduction (Teil 2)	28.10.2011	Blatt 1
1.11.2011	entfällt wegen Feiertag	4.11.2010	entfällt wegen der ausgefallenen Vorlesung
08.11.2011	Kapitel 2: Feature Reduktion Kapitel 3: Clustering in hochdimensionalen Räumen	11.11.2011	Blatt 2
15.11.2011	Kapitel 3: Clustering in hochdimensionalen Räumen	18.11.2011	Blatt 3
22.11.2011	Kapitel 3: Clustering in hochdimensionalen Räumen	25.11.2011	Blatt 4 Daten , gnuplot Befehle , Folien zur Houghtransformation Tabelle zur Aufgabe 4.3
29.11.2011	Kapitel 4: Parallels, Verteiltes und Privacy Preserving Data Mining	2.12.2011	Blatt 5
06.12.2011	Kapitel 5: Ensemble Techniken	9.12.2011	Blatt 6
13.12.2011	Kapitel 6: Multirepresentiertes Data Mining	16.12.2011	Blatt 7
20.12.2011	Kapitel 7: Multi-Instanz Data Mining	23.12.2011	Zusammenfassung des ersten Teils der Vorlesung
10.01.2012	Kapitel 8: Graph-Strukturierte Daten	13.01.2012	Blatt 8
17.01.2012	Kapitel 9: Link Mining	20.01.2012	Blatt 9
24.01.2012	Kapitel 10: Dynamische Daten	27.01.2012	Wiederholung
30.01.2012	Prüfungen	03.02.2012	Entfällt wegen Prüfungen

- Abstract

Volatile data, that is, data that change over time, are met in many modern applications. A prominent example is social networks such as Twitter and Facebook which tend to be the new trend of our society. The content published in such a network evolves over time as member interests change over time either slightly (drift) or drastically (shift). The interactions among the members also change over time as new people enter/quit the network or the communication between the group members becomes more/ less intense.

Volatile data impose new challenges for data management due to their size and high degree of variability. Among these challenges are: efficient pattern extraction, adapting patterns to the underlying population changes, discovery of patterns that span more than one time points, change detection and monitoring.

In this seminar, we will study methods and algorithms for knowledge extraction and data mining in volatile data environments. An emphasis will be put on the area of social networks.

- A variety of different topics are offered
 - Clustering
 - Outlier detection
 - Query processing
 - Sensor networks
 - Social networks
 - Mining over dynamic/ stream data
 - Change detection and Evolution Monitoring
- A complete list at: http://www.dbs.ifi.lmu.de/cms/Offene_Themen
 - Or, we can discuss your own interesting ideas

Exam info:

Saturday 14.07.2012

14:00 – 16:00

lecture hall B 101 in the main building

Good luck!!!