

Quantitative Data Analysis—SERM 2014

Purpose

In this study unit, students will be introduced to basic techniques of data measurement, acquisition, mining, and analysis. Students will apply these techniques to large datasets in the various case studies.

Goals

By the end of this unit students should be able to:

- Acquire a basic understanding of Data Measurement basics & Definitions
- Gain a high level understanding of Data Acquisition Systems
- Gain an appreciation of the implications of measuring system calibration
- Acquire an overview of the Accreditation and Certification processes
- Acquire a high level knowledge of transducers and sensors
- Gain a clear and concise knowledge of Data Measurement analysis fundamentals (including data processing, presentation, etc.)
- Acquire a basic understanding of the fundamentals of signal & data transmission (including remote sensing, wireless sensor networks, etc.)
- Gain a strong insight into specific data acquisition & control applications
- Describe what data mining is
- Identify scenarios in which data mining is likely to be a useful problem solving tool
- Explain the steps involved in data mining
- Describe the technological components necessary to do data mining, i.e. hardware and software
- Demonstrate the ability to use of at least one tool or set of tools to complete a data mining analysis

Text

Kamber, Micheline and Jiawei Han (2005) *Data Mining: Concepts and Techniques, 2nd Ed.* Morgan Kaufmann. ISBN: 1558609016

Evaluation

Assignment	%
Data Mining Quiz	5%
Data Mining Project & Presentation	45%
Data Measurement Exercises/Homework	10%
Data Measurement Exam	40%
Total	100%

The primary means of evaluating students' understanding of **data measurement and acquisition** concepts will be through an exam to be delivered during the end-of-term exam period. The primary means of evaluating students' understanding of **data mining** concepts and practice will be through the data mining analysis report and presentation that they will write and turn in at the end of this unit.

As a secondary means of evaluation, students receive one quiz given midway through the unit and will also have various exercises to complete as homework during the unit.

Schedule for Data Mining Content

Day 1: Data Rich, Information Poor—What is Data Mining and Why is it Useful?

“Data mining” has many different meanings and is a multi-disciplinary practice that has been influenced by many fields. This introductory day will bring out the underlying and unifying themes of KDD (Knowledge Discovery from Data) via a series of illustrative cases. The cases will be chosen in fields relevant to the course of study of the SERM participants, but will also be broad in focus to show the power and flexibility of the analytical methods that data mining represents. The steps in the data mining process will be introduced.

Relevant Readings

Kamber & Han Chapters 1 & 11

Day 2: Stage 1—Data Cleaning and Integration

The next several days of the unit will be spent working together through a case study. The beginning of this day's study will be spent introducing the case. Students will be asked to help develop a “plan of attack” for the data analysis. Students will get familiar with the analysis tools and the rest of this day will be spent preparing the data for analysis.

Relevant Readings

Kamber & Han Chapter 2

Day 3: Stage 2—Data Selection and Transformation

This day will be spent in continued pre-processing of data in preparation for the mining step and getting familiar with the software tools that will be used to complete the mining step.

Relevant Readings

Kamber & Han Chapter 3

Day 4: Stage 3—Data Mining

Although there are many different approaches to mining the data, the methods to be used for the case study will be chosen ahead of time and this day's lecture will quickly focus on explaining these. Once

introduced, the methods will be used right away to complete the data mining step on the data cleaned and pre-processed during the first two days.

Relevant Readings

Kamber & Han Chapters 5 & 6

Day 5: Stage 4—Pattern Evaluation and Knowledge Presentation

This will be the last day of the group case study. We will spend time discussing the patterns that emerged from the data mining step, and potentially return to earlier stages, re-running the data mining to discover new knowledge. We will spend time as a group discussing what these patterns may mean and outlining our presentation of the results.

Relevant Reading

Kamber & Han Chapter 7

Day 6: Further Discussion of Analysis Methods

Now that we will have completed a complete cycle, we will spend a little more time looking at alternative methods of mining, as well as perhaps looking at other sources of data, particularly sources such as real-time data streams from data being gathered in the field, e.g. energy production from wind turbines. There will also be lab time for students to work on their unit projects and ask questions about them.

Day 7: Lab and Discussion

This day will be essentially the same as Day 6.

Day 8: Class Presentations

Students will present their projects and findings to the class. There will be time for Q&A after each presentation. Students will turn in their completed reports in class on this day.

Schedule for Data Measurement Content

Week ONE: Measurement & Instrumentation

Introduction to Measurement & Measuring System (Instrument)

Measurement: Units and Metric System (SI)

Instruments: Measuring System Types

Instruments: Static and Dynamic characteristics of Measuring System (including Accuracy, Resolution)

Measurement: Uncertainty in measurement Type A & B – Sources & Estimation

Instrument: Calibration of Measuring System, Traceability, Accreditation & Certification

Week TWO: Measuring Systems as Data Acquisition Systems

Introduction to Data & Signals

Sensors and transducers - Sensors are EVERYWHERE!

Signal Conditioning – Variable Conversion Elements

Signal Conditioning: Signal to Noise, Filtering, Amplification, linearization, isolation, etc.

Signal Conditioning - analogue to digital converters (ADCs) and (DACs), sampling, quantization, aliasing, etc.

Signal Processing - Frequency domain and fast Fourier transform (FFT), digital filtering, etc.

Signal Transmission: Wireline and Wireless Data Transmission basics

Signal Presentation : plots, graphs, etc.