Knowledge Engineering and Data Mining

Knowledge Engineering

- The process of building intelligent knowledge based systems is called knowledge engineering

- Knowledge engineering has 6 basic phases:

  1. Problem assessment
  2. Data and knowledge acquisition
  3. Development of a prototype system
  4. Development of a complete system
  5. Evaluation and revision of the system
  6. Integration and maintenance of the system

Phase 1: Problem assessment

- Determine the problem’s characteristics

- Identify the main participants in the project

- Specify the project’s objectives

- Determine the resources needed for building the system
- Type of problems addressed by intelligent systems

<table>
<thead>
<tr>
<th>Problem type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>Inferring malfunctions of an object from its behaviour and recommending solutions.</td>
</tr>
<tr>
<td>Selection</td>
<td>Recommending the best option from a list of possible alternatives.</td>
</tr>
<tr>
<td>Prediction</td>
<td>Predicting the future behaviour of an object from its behaviour in the past.</td>
</tr>
<tr>
<td>Classification</td>
<td>Assigning an object to one of the defined classes.</td>
</tr>
<tr>
<td>Clustering</td>
<td>Dividing a heterogeneous group of objects into homogeneous subgroups.</td>
</tr>
<tr>
<td>Optimisation</td>
<td>Improving the quality of solutions until an optimal one is found.</td>
</tr>
<tr>
<td>Control</td>
<td>Governing the behaviour of an object to meet specified requirements in real-time.</td>
</tr>
</tbody>
</table>

Phase 2: Data and knowledge acquisition

- Collect and analyse data and knowledge

- Make key concepts of the system design more explicit

- The three main issues in this phase:
  - Incompatible data: store in different format or coding
  - Inconsistent data: represented differently in different data base
  - Missing data: data records contain blank fields
Phase 3: Development of a prototype system

- Choose a tool for building an intelligent system
- Transform data and represent knowledge
- Design and implement a prototype system
- Test the prototype with test cases

Phase 4: Development of a complete system

- Prepare a detailed design for a full-scale system
- Collect additional data and knowledge
- Develop the user interface
- Implement the complete system

Phase 5: Evaluation and revision of the system

- Evaluate the system against the performance criteria
- Revise the system as necessary
Phase 6: Integration and maintenance of the system

- Make arrangements for technology transfer
- Establish an effective maintenance program

How do we choose an expert system development tool?

- Tools range from high-level programming languages such as LISP, PROLOG, OPS, C and Java, to expert system shells.
- High-level programming languages offer a greater flexibility, but they require high-level programming skills
- Shells provide us with the build-in inference engine, explanation facilities and the user interface. No programming skill is required to use a shell.
Data Mining

- **Data** is what we collect and store, and **knowledge** is what helps us to make informed decisions

- The extraction of knowledge from data is called data mining

- Data mining can also be defined as the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules

- The ultimate goal of data mining is to discover knowledge

Data Warehouse

- Modern organisations must respond quickly to any change in the market. This requires rapid access to current data normally stored in operational databases

- However, an organisation must also determine which trends are relevant. This tasks is accomplished with access to historical data that are stored in large databases called data warehouse

- The main characteristic of a data warehouse is its capacity. A data warehouse is really big – it includes millions, even billions, of data records
The data stored in a data warehouse is

- **Time dependent** – linked together by the times of recording – and
- **Integrated** – all relevant information from the operational databases is combined and structured in the warehouse

A data warehouse is designed to support decision marking in the organisation. The information needed can be obtained with query tools

Query tools are assumption-based – a user must ask the right questions

How is data mining applied in practice?

- Many companies use data mining today, but refuse to talk about it
- In direct marketing, data mining is used for targeting people who are most likely to buy certain products and services
- In trend analysis, it is used to determine trends in the marketplace, for example, to model the stock market.
- In fraud detection, data mining is used to identify insurance claims, cellular phone calls and credit card purchases that are most likely to be fraudulent
Where do you go from here?

- If you want to do more research in intelligent research, we have an Intelligent System Research Group within the school of IT.

- You can get more information from the web site at:

  http://www.it.murdoch.edu.au/research/groups.html

under Intelligent Systems
<table>
<thead>
<tr>
<th>Staff</th>
<th>Visitors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prof Tim Gedeon</td>
<td>Prof Linda Holley</td>
</tr>
<tr>
<td>Dr Shayan Kiar</td>
<td>Andrew Coward</td>
</tr>
<tr>
<td>Dr Graham Mann</td>
<td>Engkorn Sookhambhitama</td>
</tr>
<tr>
<td>Dr Da Rising</td>
<td></td>
</tr>
<tr>
<td>Theo Hung Long</td>
<td></td>
</tr>
<tr>
<td>Sun Rui</td>
<td></td>
</tr>
<tr>
<td>Dr Nicole Sitter</td>
<td></td>
</tr>
<tr>
<td>Dr Kevin Wong</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ph.D.</th>
<th>Honorary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex Chong</td>
<td>Mark Alexander</td>
</tr>
<tr>
<td>Thatha Thayathake</td>
<td>Liz Cunningham</td>
</tr>
<tr>
<td>Rane Makower</td>
<td>Sebastian Kher</td>
</tr>
<tr>
<td></td>
<td>Daniel Patock</td>
</tr>
<tr>
<td></td>
<td>Seow Wai Young</td>
</tr>
</tbody>
</table>
Some of the projects undertaken in this research group:

**Staff Projects:**

- 3D Model Retrieval from Image Sequences
- ANN for Digital Signal Processing
- Autogeneration of reflexive agent shell scripts
- BEELINE Natural Language Agent
- Content-based Image Retrieval using Shape Information
- Development of methods of judging 'goodness' of image registration
- Development of Automated System for Information Extraction
- Encapsulation of good OO strategies
- Face Recognition
- Hierarchical Fuzzy System
- Influence of Personal Cognitive Styles in Online Learning Environments
- Intelligent Data Analysis
- Intelligent Data Mining
- Intelligent Routing for Network
- Intelligent Web Content Categorisation
- Keyboard Dynamics in Biometrics
- Pattern Classification using Orthogonal Moments
- Registration of Images of the Cornea
- Shape calculation from slit lamp images of the Cornea
- Starchaser Marsupial Rover prototype
- TarBaby bipedal robot
- Use of Fuzzy Cognitive Maps for Simulating Dynamic Systems
PhD Projects:

- Aspects of Using Visualisation to Obtain Information from Text.
- Finding Similarities in Text with an Alternative Connectionist Architecture
- Image Feature Extraction Invariant to Colour Intensity, Rotation and Scaling Using Eigenvector-Guided Self-Organizing Mapping Neural Network
- Hierarchical Fuzzy System in the Bioinformatics Domain

Honour Projects

- Application of Artificial Neural Networks for Digital Signal Processing
- Character Recognition using a Multilayer Perceptron Classifier and Orthogonal Moments on the Unit Disk
- Identity Verification Based on Computer Interaction
- Intelligent Web Content Categorisation
- Ranking Web Documents using Clustering, based on Content Relevance