Topics in Data Engineering

Session 1

Masaomi Kimura

Objective

- We learn methods to extract important information from data existing everywhere in our daily life.
  - Information extraction from structured/unstructured data
    - Data mining
    - Text mining
    - Deep learning based analysis (extension of DM/TM)

Data mining (DM)

1. Store huge amount of data
2. Select method suitable to the objective of the analysis
3. Modify the data format to the one suitable the method (data cleaning)

After applying the method and obtaining results,

4. interpret the results based on business knowledge
5. and improve the business

Text mining (TM)

- A target is text data
- Because text data are *unstructured*, they need to be *transformed into structured data*, which we can apply data mining techniques.
- NLP (natural language processing) is utilized
  - Morphological analysis
  - Dependency analysis
  - Recently, application of deep learning is a hot topic

Deep learning

- Deep learning is just a neural network!
- In practice, if the number of layers in a neural network is more than four, it is called deep learning (neural network).
  - The term, deep learning, has been used from Year 2000.
- Though deep learning belongs to Machine Learning, it gets common to use it in data science topics.

Syllabus (renewed ver.)

	Contents
2	Data mining method (1) Overview of data mining
3	Data mining method (2) Association analysis, memory-based reasoning
4	Data mining method (3) Clustering analysis, genetic algorithm
5	Data mining method (4) Decision Tree analysis, network analysis
6	Data mining method (6) SVM
7	Data mining method (5) artificial neural network
8	Data mining method (7) Deep learning – overview, LSTM
9	Data mining method (8) Deep learning – CNN, Transformer
10	Data mining method (9) Network analysis
11	Text mining method (1) Basics: Natural Language Processing
12	Text mining method (2) Text mining techniques and deep learning
13	XML, Application of XML : Semantic Web
14	Final exam

Evaluation

Minutes paper 30%, Final exam 40%, Final report 30%

- Minutes paper
  - For each session, students are required to submit a minute paper to summarize the explanation in the class via Google classroom.
- Final report
  - Students are expected to summarize a journal paper related data mining, text mining, deep learning and knowledge graph and to submit it as a report.
- Final exam