17th Meeting of Consortium for Globalization of Chinese Medicine (CGCM 2016)

Bioinformatics: "Omics" Approach and Data Analysis Session

(Meeting Room 4+5, BCC)

August 10, 9:30-12:00

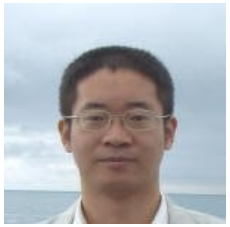
Summary

Organizing Team

Chair: Nevin L. Zhang

Hong Kong UST

Co-chair: Defang OUYANG



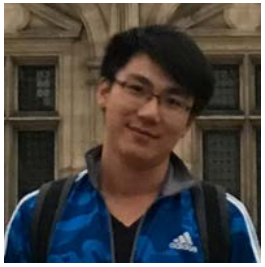
U of Macao

Panelists

Zhennan She

Kunming M. U

Zeyuan Wang

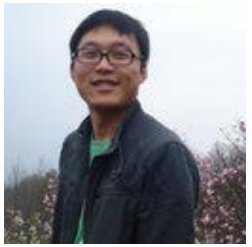


U of Sydney

Qihe Xu

King's College London

Mingxiao Yang



U of Hong Kong

Poster Selection

- 17 abstracts submitted to the OADA session
- 11 are selected by the panel for poster highlights
- 2 did not respond
- 9 poster highlights in program

Program

- Part I: 9:35 – 11:00: 9 poster highlights. Each talk is allocated 8 minutes. Q&A possible if the time not used up

- Part II 11:00 – 12:00
  - Panel-lead discussions on research directions in the OADA area
  - Audience is strongly encouraged to participate the discussions
  - Feel free to interrupt the panelists at any time
  - Feel free to raise new issues

- The session is well attended

OMICS Papers

- Qihe Xu: Proteomic analysis of TGF-β1-induced fibrogenesis and antifibrotic mechanisms of Scutellariae Radix, baicalein and an Alk5 inhibitor

- Lie-Fen Shyur: Deciphering biosynthetic pathway of bioactive polyacetylenes in Bidens pilosa L. using integrative omics and transgenic approaches

- Yuhui Hu: Genome-wide Identification of the Molecular Targets of TCM for Functional Elucidation and Quality Control

Discussions on Omics

- How to integrate different omics data?
  - Have a clear problem statement

- How to deal with small data
  - Domain knowledge
  - Dimensionality reduction
  - Bayesian inference: prior + data => posterior
  - Semi-supervised learning
    - Limited labelled data + abundant unlabelled data => classifier

Data Analysis Papers

- Zeyuan Wang: TCM Translator for Symptoms to Herbs
- Nevin L. Zhang: Comparisons of Three Data-Drive Approaches for Determining TCM Syndrome Subclasses in a Patient Population
- Zhuyifan Ye: Prediction of formulations of oral fast disintegrating films by deep learning
- Zhennan She: Real-World Evidence promote development of medical products and internationalization of TCM

Zeyuan Wang: TCM Translator for Symptoms to Herbs

Example

Symptoms	舌苔黄 双下肢发沉 舌苔厚 腰痛 口苦 发沉 脉数
Translations	a tongue with thick and yellow, legs feel heavy, bitter taste, lumbago, rapid pulse
Prescription from the Clinical Doctor	黄连 茯苓 车前草 黄柏 泽泻 益母草
Prescription from the Model	黄连 黄柏 茯苓 干姜 知母 炒酸枣仁 红参

Conclusion

Considering symptoms and herbs as sentences and constructing TCM translator by Recurrent Neural Network with attention mechanism shows promise in TCM prescription generation tasks.

Nevin L. Zhang: Patient Classification (Syndrome Differentiation)

- Judgment by individual doctors

- Judgement by panel of experts'

- Problem: Lack of consistency, subjective

Machine Learning: Supervised Learning

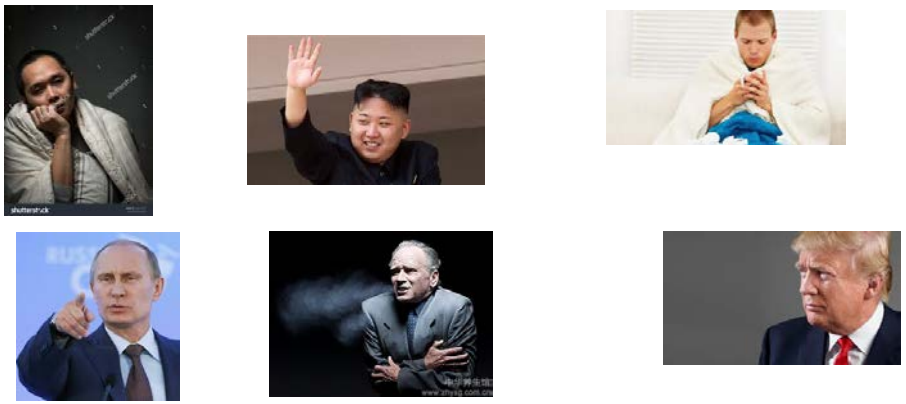
- Illustrative example: Yang Deficiency
- Labelled data:

- Learn classification rule:

- Problem: Class labels still lack consistency, subjective

Machine Learning: Unsupervised Learning

- No class labels in data



- Classification rule:

Machine Learning: Unsupervised Learning

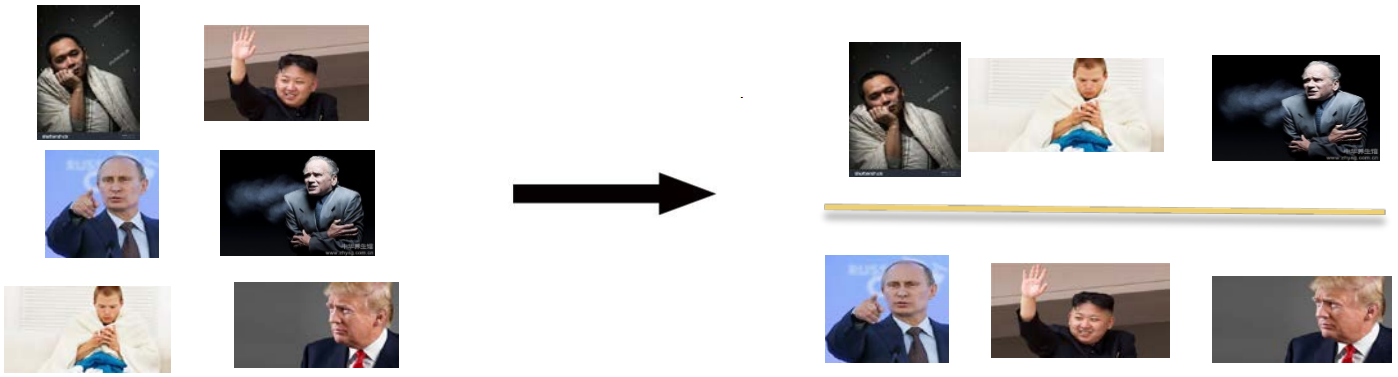
- Cluster Analysis

- Cluster interpretation: Top class is Yang Deficiency

- Classification rule: f () = Yang Deficiency

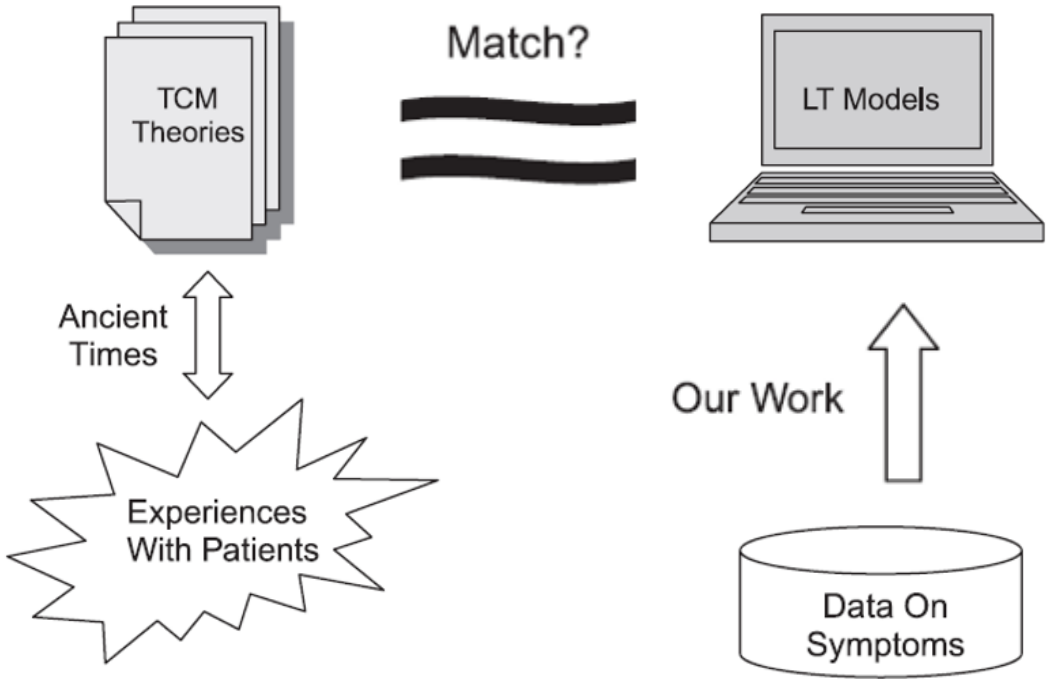
Machine Learning: Unsupervised Learning

- Cluster Analysis



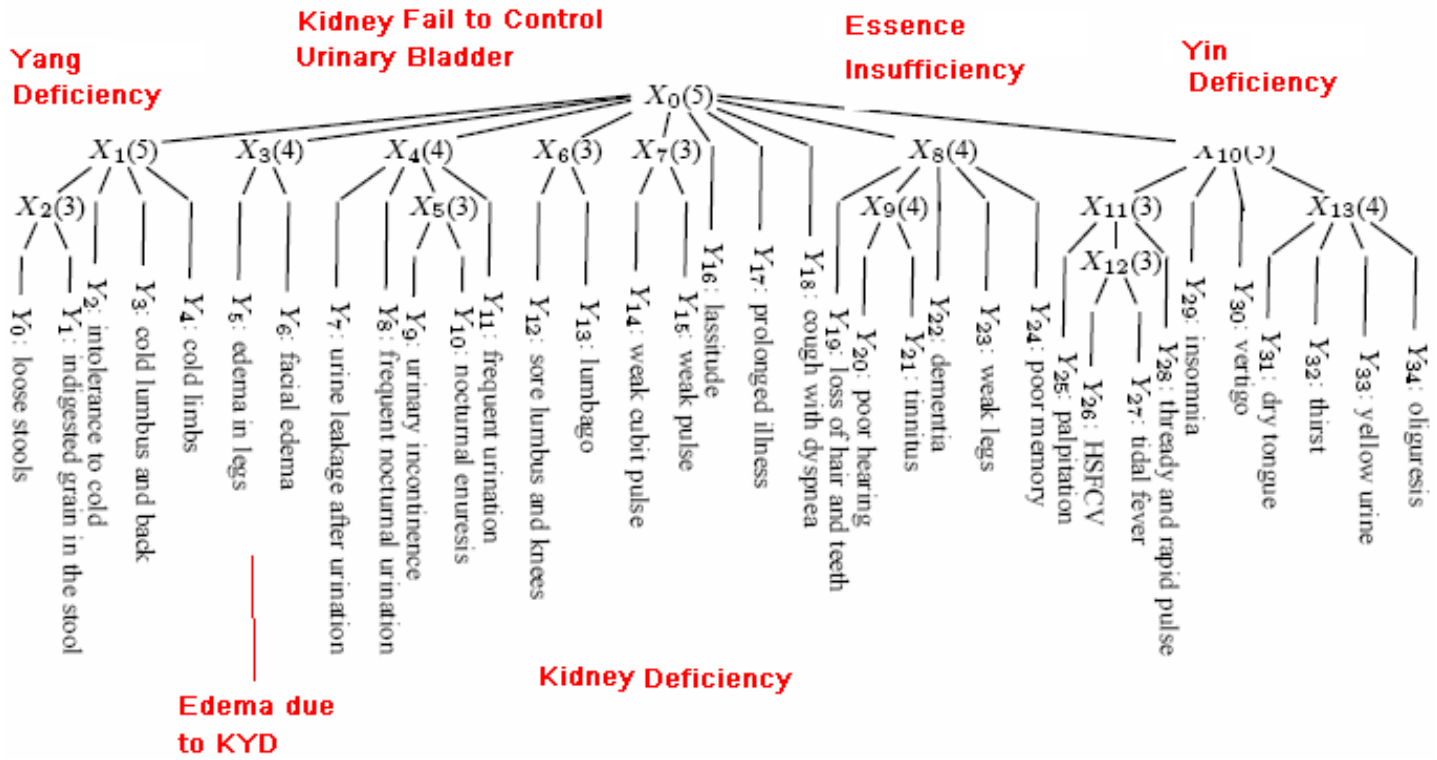
- Three Approaches to cluster analysis
  - Latent class analysis
  - One-phase latent tree analysis
  - Two-phase latent tree analysis

Providing Statistical Validation to TCM Theories

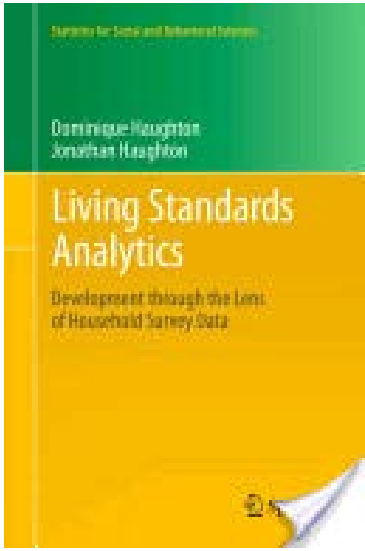


More than 20 Datasets Analyzed

- In all cases, the results match relevant TCM Theory well.



What Others Think of Our Work?



D. Haughton and J. Haughton. Living Standards Analytics: Development through the Lens of Household Survey Data. Springer. 2012

- Zhang et al. provide a very interesting application of latent tree models to diagnoses in traditional Chinese medicine (TCM).
- The results tend to confirm known theories in Chinese traditional medicine.
- This is a significant advance, since the scientific bases for these theories are not known.
- The model proposed by the authors provides at least a statistical justification for them.