

Guan-Lin Li

Email: youmu257@gmail.com | Github : <https://github.com/youmu257>

SUMMARY

I am interested in data science and engaged in topic model research which can distill key word from corpus, e.g. LDA[2]. My research[1] can support user to understand quickly a hotel in each aspect. My research had combined DMM[3] and LARAM[4] to automatically summary hotel review. I am also studying programing technique on Spark to deal with massive data. All of my research results have had upload to [github](#).

EDUCATION

National Cheng Kung University

M.S. in Computer Science and Information Engineering

GPA : 3.68 / 4.0

Tainan, Taiwan

Jul 2015 – present

National Taitung University

B.S. in Computer Science and Information Engineering

GPA : 3.67 / 4.0

Taitung, Taiwan

Sep 2011 – Jun 2015

RESEARCH

Latent Aspect Rating Analysis Model in Short Text

LARAM[4] can predicted aspect rating, predicted overall rating and summarize aspect review. My research using topic model for short text to implement aspect segment, and regression model to enhance aspect rating prediction.

- Pros and cons of hotel can be clear at a glance.
- Recommended appropriate hotel for customer by their history.
- Suitable for more data and improve effect of LARAM in short text.

PROJECT EXPERIENCE

2016 MOST Project – Heterogeneous streaming data analysis and visualization environment built on SPARK platform

- Our team's works are building text mining related tool in spark platform. It contains Topic model, NLP tool, and automatically building sentiment lexicon.
- My work is designed Topic model into map reduce structure on spark.

2015 MOST Project – Issue Analytics and Knowledge Discovery Platform for Social Communications on Clean Energy Society

- Our works are analyzed sentiment influences and opinion mining in society medial.
- My work is calculating the sentiment polarity of words in the specific-domain by semi-supervised algorithm[5].

2016 ITRI Project – Using named-entity recognition to support customer service in depot

- Our work is finding key information in customer description to help technician deal with the problem.
- My words are assignments jobs to other group members and analysis experiment result.

AWARDS

- 2015 TSMC Big Data Analytics for Semiconductor Manufacturing[6] - **Honorable Mention** (4~6/67)
- 2013,2014 National Collegiate Programming Contest - **Honorable Mention**[7]
- 2013 Collegiate Programming Examination - Grade A, Ranking : 46 / 819 [8]

SKILL

Technical Languages : Java, Python

REFERENCE

[1] Latent Aspect Rating Analysis Model in Hotel Review[[pdf](#)]

[2] LDA : <https://github.com/youmu257/jLDAGibbsSampling>

[3] DMM : <https://github.com/youmu257/jDMMGibbsSampling>

[4] LARAM : <https://github.com/youmu257/LARAM>

[5] DDSD : <http://140.116.245.104/SentimentLexicon/>

[6] TSMC picture: <https://drive.google.com/drive/folders/0B45UCSqNVGeaTDJyWGdoYlRaajQ?usp=sharing>

➤ Team name : 1KM

[7] National Collegiate Programming Contest

● 2013 NCPC : <http://ncpc2013.nsysu.edu.tw/files/11-1264-9960.php?Lang=zh-tw>

➤ Team name : 波多老師我愛你

● 2014 NCPC : <http://ncpc.ntnu.edu.tw/ncpc/showpage.php?id=11>

➤ Team name : NyanCat

[8] CPE : <https://cpe.cse.nsysu.edu.tw/cpe/scoreboard/2013-10-01>