Jiaxuan Li

Master student, Department of Computer Science
University of Harbin Institute of Technology, Shenzhen, China

Email: jiaxuanliniki@gmail.com
Telephone: 18126352475

Education

Stevens Institute of Technology, State of New Jersey, United States

Ph.D., computer science — Sep. 2020 to Sep. 2021

As a result of the US PP10043 ban, I cannot obtain a visa and withdraw from the program.

Harbin Institute of Technology，Shenzhen, China

M.S., computer science — Sep. 2017 to Jan. 202

GPA: 3.1 / 4.0

Supervisor: Dr. Philippe Fournier-Viger

Hai Nan University, Hainan, China

B.S., computer science — Sep. 2013 to Jul. 2017

GPA: 3.67 / 4.0 (88.2 / 100), Top 3%

Supervisor: Dr. Yucong Duan

Publications

- **Jiaxuan Li**+, Yue Ning*. Anti-Asian Hate Speech Detection via Data Augmented Semantic Relation Inference. *Proc. 16st Intern. (AAAI) Conf. on Web and Social Media (ICWSM 2022), AAAI*. Accepted.
- A Survey of Pattern Mining in Dynamic Graphs. Philippe Fournier-Viger*, Ganghuan He, Chao Cheng, **Jiaxuan Li**, Jerry Chun-Wei Lin, Unil Yun. *WIREs Data Mining and Knowledge Discovery, Wiley*, 2019.
- Efficiently Extracting Cost-Effective patterns from Sequential Event Log. Philippe Fournier-Viger*, **Jiaxuan** Jerry Chun-Wei Lin, Tin Truong Chi, R. Uday Kiran. *Knowledge-Based Systems (KBS,Q1), Elsevier*, 2019**.**
- Discovering and Visualizing Patterns in Utility Sequences. Philippe Fournier-Viger*, **Jiaxuan Li**+, Jerry Chun-W Lin, Tin Truong Chi. *Proc. 21st Intern. Conf. on Data Warehousing and Knowledge Discovery (DAWAK), Springer*, 2019.
- Discovering low-cost high utility patterns. **Jiaxuan Li**+, Philippe Fournier-Viger*, Lin, Jerry Chun-Wei Lin, Truong Chi. *1st International Workshop on Utility-Driven Mining (UDM), in conjunction with the KDD 2018 conference, ACM press*, 2018. **Oral presentation.**

*Academic supervisor, + main student contributor.

Research & Industry Experience

- **Golaxy Data Technologies** — Jul. 2020 to May. 202

  Affiliation with Institute of Computing Technology, Chinese Academy of Sciences

  Mentor: Shaolong Zhou

  – Integrate entity category and relative position information as features and construct a Bert-CRF baseline model for comparison experiments.

  – Construct a multi-head selection with a bilinear layer to solve the entity nesting problem.

- Construct a cascade classification network to identify entities and their corresponding types, and a multi-head selection network to extract complex entity relation pairs.

- **Harbin Institute of Technology, Shenzhen** Mar. 2018 to No
  Mentor: Dr. Philippe Fournier-Viger
  - **Cost-effective pattern mining** Mar. 2018 to Jul. 201
    - Mining cost-effective pattern from event logs in E-learning to provide insights about how to utilize the learn materials.
    - Combined a cost model into high utility sequential pattern mining considering the resources, effort, time cost required to apply the patterns for getting a desirable utility.
    - Designed statistical measures to assess the correlation between utility and cost for the needs of differ applications in terms of the type of utility (binary or numeric).
    - Integrated buffer structure into Prefixspan algorithm and designed pruning strategy to improve algorith performance in terms of memory usage and execution time.
  - **Cost-effective pattern mining from heterogeneous data source** Jul. 2019 to No
    - Mining guidance patterns in e-learning from heterogeneous data source for specific group of users.
    - Combining users' attributes, such as personal information and educational background, with their learn activities.
    - Clustering users based on their attributes and mining cost-effective patterns from their sequences of activi respectively to assist different group of users use materials efficiently.
    - Representing the concatenated features using vector and utilizing a statistical measure to evaluate correlation between the feature and utility.
    - Currently designing the model and searching potential datasets for testing.

- **Noah'Ark Lab, Huawei Technologies** Aug. 2019 to No
  Mentor: Dr. Min Zhou
  - Spatial-temporal sequence pattern mining in telecommunication network to compress alarm records, iden important alarms, and locate root-cause alarms.
  - Using dynamic attributes graph as an approach to mine important sequences of alarms that have a hig priority to be responded from various network equipment, meanwhile keeping the topology of the network
  - Designing correlation measure and generating potential correlated sequence rules of alarms to identify the r cause alarms.
  -

- **2012 Lab, Huawei Technologies** Jun. 2019 to Aug. 20
  Mentor: Dr. Zixian Zhang
  - Designed, implemented and tested a function for automatically extracting and checking the CAD drawin content about servers to improve the manual inspection's accuracy.
  - Analyzed about 5 types of drawings. Extracted their components' data structure using ActiveX and mi different components' crucial features, respectively.
  - Based on those features, implemented algorithms to structurally extract contents in specified areas, compa those information with the official documents, and finally generated a detailed verification report.
  - Checked the operating specifications of the drawings, such as the intersection between texts and lines, miss arrows, and manual errors, such as the absence of a component's description or missing a component should be contained in the drawing.
  - Tested about 100 drawings, and now this feature was integrated into their production system.

Selected Awards

Formatted: Font: (Default) Times New Roman, Bold
Deleted: →
Formatted: Font: (Default) Times New Roman
Formatted: Indent: First line: 0.74 cm
Deleted: <#>¶
Mining guidance patterns in e-learning for specific group of us from heterogeneous data source.¶
Combining users' attributes, such as personal information and educational background, with their learning activities. ¶
Clustering users based on their attributes and mine cost-effecti patterns, provided to assist users use materials efficiently, for different group of users. ¶
Using standard variance of each pattern's cost to evaluate how costly this pattern is useful for slow and fast learners. ¶
Currently designing a model to build connection between different data source and searching potential datasets for testing ¶
Deleted: <#>→
Deleted: <#>→
Formatted: Font: Not Bold
Moved (insertion) [1]
Formatted: Indent: Left: -0.5 cm, Hanging: 0.5 cm, Bulleted + Level: 1 + Aligned at: -0.49 cm + Indent at: 0.25 cm
Deleted: ¶
Moved up [1]: Integrated buffer structure into Prefixspan algorithm and designed pruning strategy to improve algorithms performance in terms of memory usage and execution time.¶
Formatted: Font: (Default) Times New Roman
Formatted: List Paragraph, Indent: Left: -0.5 cm, Hanging: 0.5 cm, Bulleted + Level: 1 + Aligned at: -0.49 cm + Indent at: 0.25 cm
Deleted: <#>¶
¶
¶
¶
¶
¶
Deleted: identify the important alarms
Deleted:
Deleted: replace the manual inspection.
Deleted: using ID3 classification algorithm

- Outstanding graduates of Hainan University, 2017.
- Mathematical Contest in Modeling Certificate of Achievement, Honorable Mention, 2016.

Formatted: List Paragraph, Bulleted + Level: 1 + Aligned at: -1.5 cm + Indent at: -0.76 cm

Deleted: ¶

¶

¶

¶

¶

¶

¶

¶

¶

¶

mportant sequences of alarms and filter that of trivial from

arious network equipmentk ¶

keep the topology of the network¶

Motivation: this work is extended from a research area callec

igh utility sequential pattern mining (HUSPM), which assess

he utility or benefits that each pattern provides. However,

HUSPM ignores the resources, effort, time or cost required t

pply these patterns. Therefore, this work considered both a

tility model and a cost model to provide insightful patterns

nd addresses the needs of various applications.¶

Defined three problems, first, information about utility is

ncoded as a binary label representing a desirable or

ndesirable outcome for each sequence; second, utility is

ncoded as a positive number evaluating the performance;

hird, the utility is a binary value and records are available fo

he positive class. ¶

¶

Designed two statistical measures to assess the correlation

etween utility and cost for above problems, since utility and

ost are measured using different units (not a simple

ubtraction relation).¶

¶

ntegrated buffer structure into Prefixspan algorithm and

designed pruning strategy to improve algorithms'

erformance in terms of memory usage and execution time.'

¶

Taking users' attributes, such as personal information and

ducational background, into consideration to mine guidanc

patterns for specific group of users from heterogeneous dat

ource.¶

Currently designing a model to build connection between

different data source and searching potential datasets for

... [1]

Page 3: [1] Deleted Win10 10/29/19 10:01:00 AM