IMT Nord Europe
École Mines-Télécom
IMT-Université de Lille

Proposal for internship - 2021/2022

Mining Frequent Gradual Itemsets From Noise Data

Supervisor: Jerry Lonlac, jerry.lonlac@imt-nord-europe.fr

Research laboratory: IMT Nord Europe, Centre for Digital Systems
https://imt-lille-douai.fr/recherche-et-innovation/systemes-numeriques/

Research team: Hide / Mcleod
https://imt-nord-europe.fr/recherche-et-innovation/systemes-numeriques/humain-interaction-decision/

Duration: 5 months – Asap (March - April 2022)

Key words: Data mining, Machine learning, Data quality, Noise data, Gradual itemset

OBJECTIVE:
Frequent Itemset Mining (FIM, for short) is an active part of data mining field and an important first step in data analysis. In the last decades, FIM has been applied in a broad range of applications such as e-commerce [4], e-learning [10], malware detection [3]. The application of FIM in a wide range of domains with a proliferation of different type of data has inspired the development of numerous other pattern-mining techniques. Recently, gradual itemsets [6, 2, 11, 7, 9] have then been proposed for analysing numerical data and different algorithms have been designed to automatically extract gradual itemsets from different data model [12, 13, 1, 5]. Gradual itemsets aroused great interest for extracting frequent complex co-variations between numerical attributes in a multitude of areas. However, in some real-world applications, data are subject to noise and measurement error. To date, the effect of noise on classical frequent gradual itemset mining algorithms has been not addressed.

The goal of this work is to propose a noise tolerant gradual itemset model, which unlike classical gradual itemsets [2, 8] tolerates a controlled fraction of errors on the extent of the gradual itemset. By allowing noise, the proposed models will generalize the level-wise enumeration of different forms of frequent gradual itemsets [2, 12, 8, 7] that can be extracted from different types of complex numerical data but obscured by noise.

Candidate profile:

- 2nd year student of a Master's or Engineering of Computer Science degree.
- Goods skills in Artificial Intelligence, more particularly in pattern mining.
- Goods skills in programming language (C++, Python).

Application: Send your CV, transcript of grades, and motivation letter to jerry.lonlac@imt-nord-europe.fr

---

References

[1] Aymeric Côme and Jerry Lonlac. Extracting frequent (closed) seasonal gradual patterns using closed itemset mining. In *IEEE International Conference on Tools with Artificial Intelligence, ICTAI*, pages 1442–1448, 2021.

[2] Lisa Di-Jorio, Anne Laurent, and Maguelonne Teisseire. Mining frequent gradual itemsets from large databases. In *IDA*, pages 297–308, 2009.

[3] Yiheng Duan, Xiao Fu, Bin Luo, Ziqi Wang, Jin Shi, and Xiaojiang Du. Detective: Automatically identify and analyze malware processes in forensic scenarios via dlls. In *ICC*, pages 5691–5696, 2015.

[4] Philippe Fournier-Viger, Jerry Chun-Wei Lin, Bay Vo, Tin Chi Truong, Ji Zhang, and Hoai Bac Le. A survey of itemset mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 7(4), 2017.

[5] Amel Hidouri, Saïd Jabbour, Jerry Lonlac, and Badran Raddaoui. A constraint-based approach for enumerating gradual itemsets. In *IEEE International Conference on Tools with Artificial Intelligence, ICTAI*, pages 582–589, 2021.

[6] Eyke Hüllermeier. Association rules for expressing gradual dependencies. In *PKDD*, pages 200–211, 2002.

[7] Jerry Lonlac, Arnaud Doniec, Marin Lujak, and Stéphane Lecoeuche. Mining frequent seasonal gradual patterns. In *Big Data Analytics and Knowledge Discovery - DaWaK*, volume 12393, pages 197–207, 2020.

[8] Jerry Lonlac, Yannick Miras, Aude Beauger, Vincent Mazenod, Jean-Luc Peiry, and Engelbert Mephu Nguifo. An approach for extracting frequent (closed) gradual patterns under temporal constraint. In *FUZZ-IEEE*, pages 878–885, 2018.

[9] Jerry Lonlac and Engelbert Mephu Nguifo. A novel algorithm for searching frequent gradual patterns from an ordered data set. *Intell. Data Anal.*, 24(5):1029–1042, 2020.

[10] Espérance Mwamikazi, Philippe Fournier-Viger, Chadia Moghrabi, and Robert Baudouin. A dynamic questionnaire to further reduce questions in learning style assessment. In Lazaros Iliadis, Ilias Maglogiannis, and Harris Papadopoulos, editors, *Artificial Intelligence Applications and Innovations*, pages 224–235, 2014.

[11] Benjamin Négrevergne, Alexandre Termier, Marie-Christine Rousset, and Jean-François Méhaut. Para miner: a generic pattern mining algorithm for multi-core architectures. *DMKD*, 28(3):593–633, 2014.

[12] NhatHai Phan, Dino Ienco, Donato Malerba, Pascal Poncelet, and Maguelonne Teisseire. Mining multi-relational gradual patterns. In *SDM*, pages 846–854, 2015.

[13] Faaiz Shah, Arnaud Castelltort, and Anne Laurent. Extracting fuzzy gradual patterns from property graphs. In *FUZZ-IEEE*, pages 1–6, 2019.