（別紙様式７）

専攻分野及び研究計画

Field of Study and Research Plan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name in full,  in your native language | POGAKU |  | ARJUN | CHAKRAVARTHI |
| （姓名（自国語）） | , |
|  | (Surname) |  | (Given name) | (Middle name) |
| Name in Roman capital letters | POGAKU |  | ARJUN | CHAKRAVARTHI |
| (姓名（ローマ字）) | , |
|  | (Surname) |  | (Given name) | (Middle name) |
|  |  |  |  |  |
| Nationality | INDIA | | | |
| （国　籍） |

Proposed study program in Japan (Outline your field of study on this side and the specific of your study program on the reverse side of this sheet. This section is one of the most important references for selection. The statement must be typewritten or written in block letters. Additional sheets of paper may be attached if necessary. If plagiarism or fraud is discovered after selection, the selection will be cancelled retroactively.)

（日本での研究計画；この研究計画は，選考の重要な参考となるので，表面に専攻分野の概要を，裏面に研究計画の詳細を具体に記入すること。記入はタイプ又は楷書によるものとし，必要な場合は別紙を追加してもよい。なお、採用後に不正、盗用等が判明した場合は遡って採用を取り消す。）

If you have Japanese language ability, write in Japanese. （相当の日本語能力を有する者は，日本語により記入すること｡）

　１　Present field of study（現在の専攻分野）

My present field of study is computer science with a specialization in Temporal Big Data Analytics & Data Science. During my undergraduate and master’s programs, I have studied various basic and advanced courses on signal communications, computer science, data science, and machine learning, which empowered me to investigate novel techniques in the field. I graduated with “First class with Distinction” in both my bachelor's and master's degrees. During my master's program at Asia University, Taiwan, I received the Innovative Academic Achievement award for scoring top marks in the department. Currently, I hold the position of assistant professor in the Department of Electronics and Communications Engineering at Madanapalle Institute of Technology & Science. I teach students in various subjects, including Statistics & Probability, Signals & Systems, Machine Learning in Communications, and Mobile Communications. Over the past year, our research group has been actively exploring and developing deep learning algorithms for implementation in Internet of Things (IoT)-based communication devices.

Collaborating with the researchers at the University of Aizu (UoA), Japan, I have expanded my knowledge in the field of pattern mining, a powerful tool for identifying complex patterns in various domains [1]. Over the past six months, I have been dedicated to studying and researching pattern mining, with a specific focus on traffic patterns. This research has enabled me to develop a deeper understanding of the application of pattern mining in detecting and analyzing complex traffic patterns. As part of this learning experience, I am currently working on a novel IoT-based project utilizing Arduino nano 33 BLE sense to dynamically identify traffic congestion on Indian roads. Indian roads are highly sensitive to traffic congestion. The roads are smaller and the number of vehicles is huge. To balance this condition, people need to travel with more organization and discipline. However, a smaller disturbance can create a huge congestion. This project initiative aims to develop innovative solutions for real-time traffic monitoring and management, possibly reducing the congestion and majorly focusing on gathering the congestion data. The project is nearing completion. Furthermore, my study of pattern mining and traffic congestion has also led me to explore its applications in predicting traffic congestion. I have authored a conference paper titled “A Novel Multi-Task Learning Framework for Predicting Traffic Congestion,” which has been accepted for presentation at the “Big Data & Artificial Intelligence 2024” conference, scheduled to take place in December 2024. This paper presents a novel algorithm based on multi-task learning neural networks for predicting traffic congestion, demonstrating the potential of pattern mining in enhancing traffic management systems. The paper talks about mining various patterns from the JARTIC dataset and feeding these patterns into a Multi-Task Learning Neural Networks framework to predict the traffic congestion of frequently congested roads all at a time. In collaboration with researchers at UoA, I have also worked on a manuscript to develop an air pollution dataset to understand and extract interesting patterns. These patterns will reveal the locations or sensors that are highly exposed to air pollution (PM2.5). The manuscript has been submitted for review in the Scientific Data Journal.

　２　Your research topic in Japan: Describe articulately the research you wish to carry out in Japan.

（渡日後の研究テーマ：日本においてどういった研究がしたいかを明確に記入すること）

**Discovering Periodic Patterns in Uncertain Temporal Data**

**Research theme:** Uncertainty is the inherent nature of big data. Useful patterns that can empower the end-users to achieve socio-economic development are hidden in this uncertain big data. Unfortunately, finding these patterns is quite challenging due to the scale of big data and the probabilistic nature of items’ occurrences. When confronted with this problem in real-world applications, researchers primarily focused on developing clustering and classification techniques by disregarding the existence of interesting patterns hidden in uncertain data. My research work aims to develop pattern-mining techniques to uncover useful information in big data. The purpose of this research is two-fold. First, my research aims to develop a theoretical framework and novel mathematical models to discover various types of patterns hidden in big data. Second, my research focuses on developing fast algorithms that can effectively reduce the vast search space of pattern mining is *2n-1*, where *n* represents the total number of items in a database.

Outdoor air pollution is one of the world's largest health and environmental problems. World Health Organization (WHO) data show that almost all of the global population (99%) breathe air that exceeds WHO guideline limits. The Global Burden of Disease study estimates that 4.14 million deaths were attributed to outdoor air pollution in 2019. Every year, Japan records at least 60,000 premature deaths due to air pollution. To tackle the pollution problem, the Atmospheric Environmental Regional Observation System (AEROS) [5], a nationwide sensor network (see Figure 1a), was set up in Japan to monitor air pollution. The hourly big data (see Figure 1b) generated by this system naturally exists as uncertain temporal data due to erosion of sensors. Using the Extract, Transformation, and Load (ETL) tools, we will transform the air pollution data into a temporal database (see Figure 1c). Next, we will implement the pattern mining techniques developed by us to discover useful information (see Figure 1d). The generated patterns will be later visualized on the maps to provide location-specific information to the users.

A diagram of a pattern mining

Description automatically generated

Figure : Pattern discovery framework to uncover interesting information hidden in nationwide air pollution data generated by AEROS

**Previous research:** Pattern mining is an important knowledge discovery technique in data science. It is an unsupervised learning technique that aims to find all user interest-based patterns that may exist in a database. Most previous studies on pattern mining have primarily focused on finding interesting patterns hidden in certain datasets. These approaches include finding frequent patterns, correlated patterns, weighted frequent patterns, periodic-frequent patterns, utility patterns, and fuzzy frequent patterns. A key limitation of these studies is that they fail to discover patterns in uncertain (or noisy) data. When confronted with this problem, few researchers have tried to tackle the problem by finding only frequently occurring patterns in uncertain data of small size. To the best of our knowledge, no work exists to find periodically occurring patterns in uncertain big data. Our research primarily aims to fill this gap in the field of big data analytics by exploring novel pattern discovery models and fast algorithms.

**Research Goals:** (1) Study the nature of uncertainty in the temporal occurrences of items in the big data, (2) develop a mathematical model of periodic-frequent pattern that may exist in uncertain temporal data, (3) design an efficient algorithm to find all periodic-frequent patterns in uncertain data, and (4) demonstrate the usefulness of our model by uncovering the geographical areas in which people living in Japan were regularly exposed to high levels of air pollutants.

３　Study program in Japan: (Describe in detail and with specifics - particularly concerning the ultimate goal(s) of your research in Japan)

（研究計画：詳細かつ具体に記入し、特に研究の最終目標について具体的に記入すること。）

**Research Approach:** To the best of our knowledge, there exists no prior work on finding periodic-frequent patterns in uncertain temporal big data. In this context, I intend to review the existing literature on ‘finding frequent patterns in uncertain transactional data’ and ‘discovering periodic-frequent patterns in certain temporal big data.’ Next, I plan to develop a unified mathematical model to find all periodic patterns in uncertain temporal big data. I will also develop an efficient algorithm for my mathematical model.

Every year, Japan records at least 60,000 premature deaths due to air pollution. The Ministry of Environment, Japan has set up an Atmospheric Environmental Regional Observation System [5] to monitor air pollution. The data generated by this system naturally exists as uncertain temporal data. However, finding useful information in this uncertain temporal data is non-trivial and challenging due to the lack of mathematical models. I intend to demonstrate the usefulness of my mathematical model by finding useful information that is hidden in the uncertain air pollution temporal big data. In particular, my model aims to identify the geographical areas in which people living in Japan were regularly subjected to high levels of air pollution. Such a work was not possible with the existing air pollution analytical models [2,3,4].

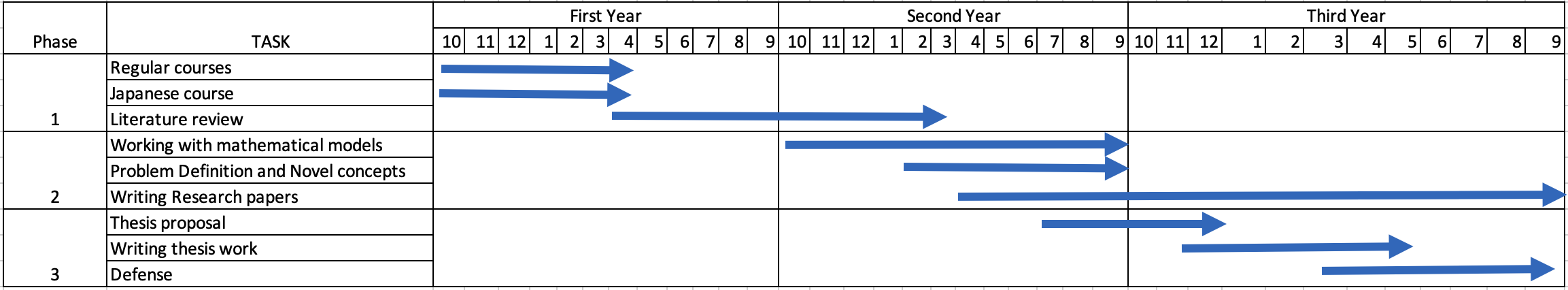
**Research Timeline:**

**Phase I:** As part of my coursework, I would like to complete the courses related to Big Data Science and some other technical courses offered by the university. I have to familiarize myself with the software/programming languages related to my research work. As part of the literature review, I would like to identify and utilize several research articles from peer-reviewed journals, books, and any other resources during the entire stipulated schedule. I would like to identify whether there will be any modification/further optimization of the existing research work by thoroughly inspecting the present literature for the same. (During First Year)

**Phase II**: Besides continuing the regular coursework, I will propose a novel method and compare with the existing methods. (During the second year, first semester)

**Phase III**: Analysis of results and Publications. (During the second semester in the second year)

**Phase IV**: Thesis write-up. (During the first semester of the third year)



References

1. Qiang Yang & Xindong Wu, 2006. “10 Challenging Problems In Data Mining Research,” International Journal of Information Technology & Decision Making (IJITDM), World Scientific Publishing Co. Pte. Ltd., vol. 5(04), pages 597-604.
2. Ravikumar, P., Likhitha, P., Venus Vikranth Raj, B., Uday Kiran, R., Watanobe, Y., & Zettsu, K. (2021). Efficient discovery of periodic-frequent patterns in columnar temporal databases. Electronics, 10(12), 1478.
3. Kiran, R. U., Zettsu, K., Toyoda, M., Fournier-Viger, P., Reddy, P. K., & Kitsuregawa, M. (2019, July). Discovering spatial high utility itemsets in spatiotemporal databases. In Proceedings of the 31st International Conference on Scientific and Statistical Database Management (pp. 49-60).
4. Likhitha, P., Ravikumar, P., Kiran, R. U., & Watanobe, Y. (2023, January). Discovering Top-k Periodic-Frequent Patterns in Very Large Temporal. In Big Data Analytics: 10th International Conference, BDA 2022, Hyderabad, India, December 19–22, 2022, Proceedings (Vol. 13773, p. 200). Springer Nature.