# MODELING TIME-SERIES AND SPATIAL DATA FOR RECOMMENDATIONS AND OTHER APPLICATIONS

**VINAYAK GUPTA**

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
APRIL 2023

# MODELING TIME-SERIES AND SPATIAL DATA FOR RECOMMENDATIONS AND OTHER APPLICATIONS

by

VINAYAK GUPTA

Department of Computer Science and Engineering

Submitted

in fulfillment of the requirements of the degree of **Doctor of Philosophy**

to the

**Indian Institute of Technology Delhi**
**April 2023**

# Certificate

This is to certify that the thesis titled **MODELING TIME-SERIES AND SPATIAL DATA FOR RECOMMENDATIONS AND OTHER APPLICATIONS** being submitted by **Mr. VINAYAK GUPTA** for the award of **Doctor of Philosophy** in **Computer Science and Engineering** is a record of bona fide work carried out by him under my guidance and supervision at the **Department of Computer Science and Engineering, Indian Institute of Technology Delhi**. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Srikanta Bedathur
Professor
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
New Delhi- 110016

# Acknowledgements

I am highly indebted to my guide, Prof. Srikanta Bedathur, a brilliant advisor and an invaluable friend over the years. His '*what are we trying to achieve here*' approach to all aspects of research and life will continue to inspire me. I will greatly cherish our discussions on topics ranging from sophisticated neural models trained on tens of GPUs to casual gigs on academic life. I am beyond belief grateful that he took me under his guidance and withstanding my never-ending tantrums around '*what am I here?*'. This thesis is a tiny aspect of the immense knowledge I have gained by working under him.

A special thanks must go to Prof. Abir De for his invaluable encouragement and constant update meetings that always motivated me to work on my projects. Without his support, a significant part of the research presented in this thesis would have never begun. Never had I thought that a small interaction at CODS-COMAD 2019 would result in a collaboration of many years and multiple top-tier publications. I am also grateful to my research committee members Prof. Parag Singla and Prof. Rahul Garg, and Dr. L. V. Subramaniam, for their critical evaluation and suggestions that helped me shape most of the work in this thesis.

Special shout-outs to my SIT 309 gang – Dishant Goyal, Sandeep Kumar, Omais Shafi, Dilpreet Kaur, Ovia Seshadri, and Arindam Bhattacharya for their constant support via poker nights, cricket games, food hunting trips, and innumerable coffee breaks. To Dishant, I would like to know if I'll ever be able to pay him back for being my go-to guy for many years. I am also thankful to have immense support from Apala Shankar Garg and Garima Gaur during my degree's initial and final phases, respectively. I'm also in massive debt to my friends from my IIIT days, specially Vijendra Singh, Ayush Srivastava, Mayur Mishra, Avashesh Singh, Ayushi Jain, and Ovais Malik, for enduring my never-ending cribs about grad school. Lastly, I thank my parents, my brother Kartik, and my cousins Surabhi and Ketan, for standing by me.

I am the author of this thesis, but all of them are surely the authors of me.

**Vinayak Gupta**

# Abstract

A recommender system aims to understand the users' inclination towards the different items and provide better experiences by recommending candidate items for future interactions. These personalized recommendations can be of various forms, such as e-commerce products, points-of-interest (POIs), music, social connections, *etc.* Traditional recommendation systems, such as content-based and collaborative filtering models, calculate the similarity between items and users and then recommending similar items to similar users. However, these approaches utilize the user-item interactions in a *static* way, *i.e.*, without any time-evolving features. This assumption significantly limits their applicability in real-world settings, as a notable fraction of data generated via human activities can be represented as a sequence of events over a continuous time. These continuous-time event sequences or CTES [1] are pervasive across a wide range of domains such as online purchases, health records, spatial mobility, social networks, *etc.* Moreover, these sequences can implicitly represent the time-sensitive properties of events, the evolving relationships between events, and the temporal patterns within and across sequences. For example, (i) event sequences derived from the purchases records in e-commerce platforms can help in monitoring the users' evolving preferences towards products; and (ii) sequences derived from spatial mobility of users can help in identifying the geographical preferences of users, their check-in category interests, and the physical activity of the population within the spatial region. Therefore, we represent the user-item interactions as temporal sequences of discrete events, as understanding these patterns is essential to power accurate recommender systems.

With the research directions described in this thesis, we seek to address the critical challenges in designing recommender systems that can understand the dynamics of continuous-time event sequences. We follow a *ground-up* approach, *i.e.*, first, we address the problems that may arise due to the poor quality of CTES data being fed into a recommender system. Later, we handle the task of designing accurate recommender systems. To improve the quality of the CTES data, we address a fundamental problem of overcoming *missing* events in temporal sequences. Moreover, to provide accurate sequence modeling frameworks, we design solutions for points-

[1] We use the acronym CTES to denote a single and well as multiple continuous-time event sequences.

of-interest recommendation, *i.e.*, models that can handle spatial mobility data of users to various POI check-ins and recommend candidate locations for the next check-in. Lastly, we highlight that the capabilities of the proposed models can have applications beyond recommender systems, and we extend their abilities to design solutions for large-scale CTES retrieval and human activity prediction.

To summarize, this thesis includes three directions: (i) Temporal Sequences with Missing Events; (ii) Recommendation in Spatio-Temporal Settings; and (iii) Applications of Modeling Temporal Sequences. In the first part, we present an unsupervised model and inference method for learning neural sequence models in the presence of CTES with missing events. This framework has many downstream applications, such as imputing missing events and forecasting future events. In the second part, we design point-of-interest recommender systems that utilize the geographical features associated with the spatial check-ins to recommend future locations to a user. Here, we propose solutions for *static* POI recommendation, sequential recommendations, and recommendations models that utilize the similarity between physical mobility and the smartphone activities of users. Lastly, in the third part, we highlight the strengths of the proposed frameworks to design solutions for two tasks: (i) retrieval systems, *i.e.*, retrieving relevance sequences for a given query CTES from a large corpus of CTES data; and (ii) understanding that different users take different times to perform similar actions in activity videos. Moreover, in each chapter, we highlight the drawbacks of current deep learning-based models, design better sequence modeling frameworks, and experimentally underline the efficacy of our proposed solutions over the state-of-the-art baselines. Lastly, we report the drawbacks and the possible extensions for every solution proposed in this thesis.

A significant part of this thesis uses the idea of modeling the underlying distribution of CTES via *neural* marked temporal point processes (MTPP). Traditional MTPP models are stochastic processes that utilize a fixed formulation to capture the generative mechanism of a sequence of discrete events localized in continuous time. In contrast, neural MTPP combine the underlying ideas from the point process literature with modern deep learning architectures. The ability of deep-learning models as accurate function approximators has led to a significant gain in the predictive prowess of neural MTPP models. In this thesis, we utilize and present several neural network-based enhancements for the current MTPP frameworks for the aforementioned real-world applications.

# सार

एक अनुशंसाकर्ता प्रणाली का उद्देश्य विभिन्न वस्तुओं के प्रति उपयोगकर्ताओं के झुकाव को समझना और भविष्य की बातचीत के लिए उम्मीदवार वस्तुओं की सिफारिश करके बेहतर अनुभव प्रदान करना है। ये वैयक्तिकृत अनुशंसाएँ विभिन्न रूपों की हो सकती हैं, जैसे ई-कॉमर्स उत्पाद, रुचि के बिंदु (पीओआई), संगीत, सामाजिक संपर्क आदि। पारंपरिक अनुशंसा प्रणाली, जैसे सामग्री-आधारित और सहयोगी फ़िल्टरिंग मॉडल, वस्तुओं और वस्तुओं के बीच समानता की गणना करते हैं। उपयोगकर्ताओं और फिर समान उपयोगकर्ताओं को समान वस्तुओं की सिफारिश करना। हालाँकि, ये दृष्टिकोण उपयोगकर्ता-आइटम इंटरैक्शन को स्थिर तरीके से उपयोग करते हैं, अर्थात, बिना किसी समय-विकसित सुविधाओं के। यह धारणा वास्तविक दुनिया की सेटिंग्स में उनकी प्रयोज्यता को महत्वपूर्ण रूप से सीमित करती है, क्योंकि मानव गतिविधियों के माध्यम से उत्पन्न डेटा का एक उल्लेखनीय अंश निरंतर समय में घटनाओं के अनुक्रम के रूप में प्रदर्शित किया जा सकता है। ये निरंतर-समय घटना अनुक्रम या सीटीईएस[1] ऑनलाइन खरीद, स्वास्थ्य रिकॉर्ड, स्थानिक गतिशीलता, सामाजिक नेटवर्क इत्यादि जैसे डोमेन की एक विस्तृत श्रृंखला में व्यापक हैं। इसके अलावा, ये अनुक्रम घटनाओं के समय-संवेदी गुणों का प्रतिनिधित्व कर सकते हैं, विकसित घटनाओं के बीच संबंध, और अनुक्रमों के भीतर और पार लौकिक पैटर्न। उदाहरण के लिए, (i) ई-कॉमर्स प्लेटफॉर्म में खरीदारी के रिकॉर्ड से प्राप्त घटना क्रम से उत्पादों के प्रति उपयोगकर्ताओं की बढ़ती प्राथमिकताओं की निगरानी में मदद मिलेगी; और (ii) उपयोगकर्ताओं की स्थानिक गतिशीलता से प्राप्त अनुक्रम उपयोगकर्ताओं की भौगोलिक प्राथमिकताओं, उनके चेक-इन श्रेणी के हितों और स्थानिक क्षेत्र के भीतर जनसंख्या की भौतिक गतिविधि की पहचान करने में मदद करेगा। इसलिए, हम असतत घटनाओं के अस्थायी अनुक्रमों के रूप में उपयोगकर्ता-आइटम इंटरैक्शन का प्रतिनिधित्व करते हैं, क्योंकि इन पैटर्नों को समझना सटीक अनुशंसाकर्ता सिस्टम को शक्ति देने के लिए आवश्यक है।

इस थीसिस में वर्णित अनुसंधान दिशाओं के साथ, हम अनुशंसाकर्ता प्रणालियों को डिजाइन करने में महत्वपूर्ण चुनौतियों का समाधान करना चाहते हैं जो निरंतर-समय की घटनाओं के अनुक्रमों की गतिशीलता को समझ सकते हैं। हम ग्राउंड-अप दृष्टिकोण का पालन करते हैं, यानी, पहले, हम उन समस्याओं का समाधान करते हैं जो सीटीईएस डेटा की खराब गुणवत्ता के कारण उत्पन्न हो सकती हैं, जो एक अनुशंसाकर्ता प्रणाली में दर्ज की जाती हैं। बाद में, हम सटीक अनुशंसा प्रणाली को डिजाइन करने का कार्य संभालते हैं। सीटीईएस डेटा की गुणवत्ता में सुधार करने के लिए, हम लौकिक अनुक्रमों में लापता घटनाओं पर काबू पाने की मूलभूत समस्या का समाधान करते हैं। इसके अलावा, सटीक अनुक्रम मॉडलिंग ढांचा प्रदान करने के लिए, हम ब्याज के बिंदुओं के लिए समाधान तैयार करते हैं, यानी मॉडल जो विभिन्न पीओआई चेक-इन के लिए उपयोगकर्ताओं के स्थानिक गतिशीलता डेटा को संभाल सकते हैं और अगले चेक-इन के लिए उम्मीदवार स्थानों की सिफारिश कर सकते हैं।

[1]हम परिवर्णी शब्द CTES का उपयोग एकल और अच्छी तरह से एकाधिक निरंतर-समय घटना अनुक्रमों को निरूपित करने के लिए करते हैं।

अंत में, हम इस बात पर प्रकाश डालते हैं कि प्रस्तावित मॉडलों की क्षमताओं में अनुशंसाकर्ता प्रणालियों से परे अनुप्रयोग हो सकते हैं, और हम बड़े पैमाने पर सीटीईएस पुनर्प्राप्ति और मानव गतिविधि की भविष्यवाणी के लिए समाधान तैयार करने के लिए उनकी क्षमताओं का विस्तार करते हैं।

संक्षेप में, इस थीसिस में तीन दिशाएँ शामिल हैं: (i) लापता घटनाओं के साथ टेम्पोरल सीक्वेंस; (ii) स्थान-अस्थायी सेटिंग्स में सिफारिश; और (iii) मॉडलिंग टेम्पोरल सीक्वेंस के अनुप्रयोग। पहले भाग में, हम लापता घटनाओं के साथ सीटीईएस की उपस्थिति में तंत्रिका अनुक्रम मॉडल सीखने के लिए एक अनुपयोगी मॉडल और अनुमान विधि प्रस्तुत करते हैं। इस ढांचे में कई डाउनस्ट्रीम अनुप्रयोग हैं, जैसे लापता घटनाओं को लागू करना और भविष्य की घटनाओं की भविष्यवाणी करना। दूसरे भाग में, हम रुचि के बिंदु अनुशंसाकर्ता सिस्टम डिज़ाइन करते हैं जो उपयोगकर्ता को भविष्य के स्थानों की अनुशंसा करने के लिए स्थानिक चेक-इन से जुड़ी भौगोलिक विशेषताओं का उपयोग करते हैं। यहां, हम स्थैतिक POI अनुशंसा, अनुक्रमिक अनुशंसा और अनुशंसा मॉडल के समाधान प्रस्तावित करते हैं जो भौतिक गतिशीलता और उपयोगकर्ताओं की स्मार्टफ़ोन गतिविधियों के बीच समानता का उपयोग करते हैं। अंत में, तीसरे भाग में, हम दो कार्यों के लिए समाधान डिजाइन करने के लिए प्रस्तावित ढांचे की ताकत पर प्रकाश डालते हैं: (i) पुनर्प्राप्ति प्रणाली, यानी, सीटीईएस डेटा के एक बड़े कोष से किसी दिए गए प्रश्न सीटीईएस के लिए प्रासंगिकता अनुक्रम पुनर्प्राप्त करना; और (ii) विभिन्न उपयोगकर्ताओं द्वारा गतिविधि वीडियो में क्रियाओं के अनुक्रम के समान कार्य करने के लिए आवश्यक समय को समझना। इसके अलावा, प्रत्येक अध्याय में, हम वर्तमान गहन शिक्षण-आधारित मॉडलों की कमियों को उजागर करते हैं, बेहतर अनुक्रम मॉडलिंग ढांचे को डिजाइन करते हैं, और अत्याधुनिक बेसलाइनों पर हमारे प्रस्तावित समाधानों की प्रभावकारिता को प्रयोगात्मक रूप से रेखांकित करते हैं। अंत में, हम इस थीसिस में प्रस्तावित हर समाधान की कमियों और संभावित विस्तार की रिपोर्ट करते हैं।

इस थीसिस का एक महत्वपूर्ण हिस्सा सीटीईएस के अंतर्निहित वितरण को तंत्रिका चिह्नित अस्थायी बिंदु प्रक्रियाओं (एमटीपीपी) के माध्यम से मॉडलिंग के विचार का उपयोग करता है। पारंपरिक एमटीपीपी मॉडल स्टोचैस्टिक प्रक्रियाएं हैं जो निरंतर समय में स्थानीयकृत असतत घटनाओं के अनुक्रम के जनरेटिव तंत्र को पकड़ने के लिए एक निश्चित फॉर्मूलेशन का उपयोग करती हैं। इसके विपरीत, तंत्रिका एमटीपीपी आधुनिक गहन शिक्षण आर्किटेक्चर के साथ बिंदु प्रक्रिया साहित्य से अंतर्निहित विचारों को जोड़ती है। डीप-लर्निंग मॉडल की सटीक फ़ंक्शन सन्निकटन के रूप में क्षमता ने तंत्रिका संबंधी भविष्य कहनेवाला कौशल में एक महत्वपूर्ण लाभ प्राप्त किया है।

# Contents

**Certificate** **i**

**Acknowledgements** **iii**

**Abstract** **v**

**List of Figures** **xvi**

**List of Tables** **xxi**

**1 Introduction** **1**

1.1 Our Research Contributions . . . . . 4

1.2 Organization of Thesis . . . . . 7

**2 Background** **9**

2.1 Marked Temporal Point Process . . . . . 9

2.1.1 Conditional Intensity Function of an MTPP . . . . . 10

2.1.2 Normalizing Flows . . . . . 11

2.2 Neural Temporal Point Process . . . . . 12

2.2.1 Intensity-free formulation of Temporal Point Process . . . . . 13

2.3 Graph Convolution/Attention Networks . . . . . 14
2.3.1 Graphs in Recommendation Systems . . . . . 14
2.4 Self-Attention . . . . . 14

**I Temporal Sequences with Missing Events 17**

**3 Overcoming Missing Events in Continuous-Time Sequences 19**

3.1 Introduction . . . . . 19
3.1.1 Our Contribution . . . . . 20
3.2 Related Work . . . . . 21
3.2.1 Missing Data Models for Discrete-Time Series . . . . . 21
3.2.2 Missing Data Models for Temporal Point Process . . . . . 22
3.3 Problem Setup . . . . . 22
3.3.1 Preliminaries and Notations . . . . . 23
3.3.2 Overcoming Missing Events . . . . . 23
3.4 Components of IMTPP . . . . . 24
3.5 Architecture of IMTPP . . . . . 26
3.5.1 High-level Overview . . . . . 26
3.5.2 Parameterization of $p_\theta$ . . . . . 27
3.5.3 Parameterization of $q_\phi$ . . . . . 28
3.5.4 Prior MTPP model $p_{\text{prior}}$ . . . . . 30
3.5.5 Training $\theta$ and $\phi$ . . . . . 30
3.5.6 Optimal Position for Missing Events . . . . . 31
3.5.7 Salient Features of IMTPP . . . . . 31

3.6 Experiments . . . 33
3.6.1 Experimental Setup . . . 33
3.6.2 Implementation Details . . . 35
3.6.3 Event Prediction Performance . . . 36
3.6.4 Ablation Study . . . 38
3.6.5 Forecasting Future Events . . . 40
3.6.6 Performance with Missing Data . . . 40
3.6.7 Scalability Analysis with PFPP [149] . . . 41
3.6.8 Imputation Performance . . . 43
3.6.9 Evaluating the Performance of IMTPP++ . . . 44
3.7 Conclusion . . . 45

**II Recommendation in Spatio-Temporal Settings 47**

**4 Cross-Region Transfer for Spatial Features 49**

4.1 Introduction . . . 49
4.1.1 Our Contribution . . . 50
4.2 Related Work . . . 51
4.2.1 Mobility Prediction . . . 51
4.2.2 Graph based Recommendation . . . 52
4.2.3 Clustering in Spatial Datasets . . . 52
4.2.4 Transfer Learning and Mobility . . . 53
4.3 Problem Formulation . . . 53
4.3.1 User-Location Graph Construction . . . 54

4.4 AXOLOTL Framework . . . . . 55
4.4.1 Basic Model . . . . . 56
4.4.2 Model Prediction . . . . . 58
4.4.3 AXOLOTL: Information Transfer . . . . . 59
4.5 Experiments . . . . . 64
4.5.1 Experimental Settings . . . . . 64
4.5.2 Methods . . . . . 65
4.5.3 Performance Comparison . . . . . 67
4.5.4 Ablation Study . . . . . 68
4.5.5 Transfer of Weights across Regions . . . . . 69
4.6 Conclusion . . . . . 70

**5 Sequential Recommendation using Spatio-Temporal Sequences 73**

5.1 Introduction . . . . . 73
5.1.1 Our Contribution . . . . . 74
5.2 Related Works . . . . . 75
5.2.1 Sequential POI Recommendation . . . . . 75
5.2.2 Marked Temporal Point Processes with Spatial Data . . . . . 75
5.3 Problem Setup . . . . . 75
5.4 Model Description . . . . . 76
5.4.1 Region-Specific MTPP . . . . . 76
5.4.2 Flow-based Transfer . . . . . 78
5.5 Evaluation . . . . . 79
5.5.1 Experimental Settings . . . . . 79

5.5.2 Baselines . . . 80
5.5.3 Prediction Performance . . . 81
5.5.4 Qualitative Analysis . . . 82
5.5.5 Advantages of Transfer . . . 83
5.5.6 Product Recommendation . . . 83
5.6 Conclusion . . . 84

**6 Learning Spatial Behaviour using User Traces on Smartphone Apps 85**

6.1 Introduction . . . 85
6.1.1 Limitations of Prior Works . . . 86
6.1.2 Our Contribution . . . 87
6.2 Related Work . . . 88
6.2.1 Modeling Smartphone and Mobility . . . 88
6.2.2 Sequential Recommendation . . . 89
6.2.3 Relative Positional Encodings and Self-Attention . . . 89
6.3 Problem Formulation . . . 90
6.4 REVAMP Framework . . . 91
6.4.1 High-level Overview . . . 91
6.4.2 Embedding Initiator (EI) . . . 93
6.4.3 Relative or Inter-check-in Variations . . . 95
6.4.4 Sequential Recommender (SR) . . . 97
6.4.5 REVAMP: Training . . . 100
6.5 Experiments . . . 100
6.5.1 Experimental Setup . . . 101

6.5.2 Performance Comparison . . . . . 104
6.5.3 Ablation Study . . . . . 106
6.5.4 App and Location Prediction Category . . . . . 107
6.5.5 Scalability of REVAMP . . . . . 108
6.6 Conclusion . . . . . 109

**III Applications of Modeling Temporal Sequences 111**

**7 Large-Scale Retrieval of Temporal Sequences 113**

7.1 Introduction . . . . . 113
7.1.1 Our Contribution . . . . . 114
7.2 Preliminaries . . . . . 115
7.2.1 Notations and MTPP . . . . . 115
7.2.2 Problem setup . . . . . 116
7.3 NEUROSEQRET Model . . . . . 117
7.3.1 Components of NEUROSEQRET . . . . . 117
7.3.2 Neural Parameterization of NEUROSEQRET . . . . . 119
7.3.3 Parameter estimation . . . . . 122
7.4 Scalable Retrieval with Hashing . . . . . 122
7.4.1 Random hyperplane based hashing method . . . . . 123
7.4.2 Trainable Hashing for Retrieval . . . . . 123
7.5 Experiments . . . . . 125
7.5.1 Experimental Setup . . . . . 126
7.5.2 Results on retrieval accuracy . . . . . 129

7.5.3 Results on Retrieval Efficiency . . . 131
7.5.4 Analysis at a Query Level . . . 133
7.5.5 Runtime Analysis . . . 133
7.5.6 Query Length . . . 134
7.5.7 Qualitative Analysis . . . 134
7.6 Conclusion . . . 134

**8 Modeling Human Action Sequences 137**

8.1 Introduction . . . 137
8.1.1 Our Contribution . . . 138
8.2 Preliminaries . . . 139
8.2.1 Activity Prediction . . . 139
8.2.2 Problem Formulation . . . 139
8.3 PROACTIVE Model . . . 140
8.3.1 High Level Overview . . . 140
8.3.2 Neural Parameterization . . . 141
8.3.3 Early Goal Detection and Action Hierarchy . . . 144
8.3.4 Optimization . . . 145
8.3.5 Sequence Generation . . . 146
8.4 Experiments . . . 147
8.4.1 Datasets . . . 148
8.4.2 Baselines . . . 148
8.4.3 Evaluation Criteria . . . 149
8.4.4 Experimental Setup . . . 149

8.4.5 Action Prediction Performance . . . . . 150

8.4.6 Goal Prediction Performance . . . . . 152

8.4.7 Sequence Generation Performance . . . . . 153

8.5 Conclusion . . . . . 154

**9 Conclusion and Future Work** **155**

**Biography** **177**

# List of Figures

3.1 The overall neural architecture of IMTPP. The figure illustrates the notations, the observed, and missing point processes in IMTPP. The components concerning observed events and missing events are marked with blue and red, respectively. The figure also illustrates the generation process for events $e_{k+1}$ and $\epsilon_r$. . . . . 24

3.2 Architecture of different processes in IMTPP. Panel (a) shows the neural architecture of the MTPP of observations $p_\theta$. Panel (b) shows the neural architecture of the posterior MTPP of missing events $q_\phi$. The information of $e_{k+1}$ is truncated to the lognormal distribution for missing data generation, whereas the lognormal distribution for observed is non-truncated. . . . . 26

3.3 Real life examples of true and predicted inter-arrival times $\Delta_{t,k}$ of different events $e_k$, against $k$ for $k \in \{k+1, \ldots, N\}$. Panels (a) and (b) show the results for Movies and Toys datasets, respectively. . . . . 38

3.4 Performance gain in terms of AE(baseline) $-$ AE(IMTPP)— the gain (above x-axis) or loss (below x-axis) of the average error per event $\mathbb{E}[|t_k - \hat{t_k}|]$ of IMTPP— with respect to two competitive baselines: RMTPP and PFPP. Events in the test set are sorted by decreasing gain of IMTPP along $x$-axis. Panels (a) and (b) show the results for Movies and Toys datasets, respectively. . . . . 38

3.5 Variation of the forecasting performance of IMTPP and RMTPP in terms of MAE and MPA at predicting next $i$-th event, against $i$ for Movies and Toys dataset. Panels (a–b) show the variation of MAE, while panels (c–d) show the variation of MPA. They show that as $n$ increases, the performance deteriorates for both the metrics and both datasets as the prediction task becomes more and more difficult. . . . . 40

3.6 Impact of missing observations on model performance for Movies and Toys dataset. We randomly delete 40% and 60% events from the observed sequence and then train and test IMTPP and best performing baselines on the rest of the observed events. Panels (a–b) show the results for time prediction, while panels (c–d) show the results for mark prediction. . . . . 41

3.7 Runtime performance of PFPP and IMTPP for the largest dataset, *i.e.*, Movies, with complete sequences. Panel (a) shows the time vs. length of the training sequence, and panel (b) shows the time vs. the number of epochs. . . . . 42

3.8 Missing event imputation performance of IMTPP and PFPP for Movies and Toys datasets. Panels (a–b) show the results for time prediction while panels (c–d) show the results for mark prediction. . . . . 43

3.9 Predicting observed events using IMTPP++ across different values of $\overline{N}$. The results show that as we increase $\overline{N}$, the performance gap between IMTPP++ and IMTPP is decreased. . . . . 43

3.10 Compared the imputation performance of IMTPP++ and PFPP across different numbers of missing events. Note that this setting differs from Figure 3.8, as here the events are missing at random positions. . . . . 44

4.1 Skew in the volume of mobility data across different states in the US (Figure 4.1a) and the large variation in region-specific density of mobility data between California and Washington in Figures 4.1b and 4.1c respectively (based on Gowalla dataset [183]). . . . . 50

4.2 Different node and edge types in our graph model (users are in yellow and locations are in blue). The dashed arrows represent various influences and corresponding GATs ($\Phi_i, i = 1, 2, 3, 4$) in AXOLOTL. . . . . 56

4.3 System architecture of AXO-basic with level-wise embedding computation and affinity prediction. User-latent ($\boldsymbol{U}_l$) and location conditioned ($\boldsymbol{U}_s$) embeddings are combined to a *final* user embedding ($\boldsymbol{U}_f$), and similarly for location, $\boldsymbol{L}_l$ and $\boldsymbol{L}_s$ are combined to $\boldsymbol{L}_f$. . . . . 58

4.4 The difference between MAML(fig 4.4a) and our SSML (fig 4.4b) is that the latter optimizes parameters in a hierarchy, *i.e.*, user and location parameters are updated for neighborhood prediction and then combined for POI recommendation. (Best viewed in color). . . . . 59

4.5 Ablation Study across different graph architectures possible in AXOLOTL along with a GCN variant. . . . . 69

4.6 Contribution of novel spatio-social meta-learning in AXOLOTL and the corresponding gains over MeLU. . . . . 69

4.7 Bucket-wise Mean Attention Transfer Weights between Source and Target regions. . . . . 70

5.1 Architecture of REFORMD with flow-based transfer between source region (red) and target region (blue). . . . . 77

5.2 Real life *true* and *predicted* inter-arrival times $\Delta_{t,k}$ of different events $e_k$ for (a) Virginia and (b) Aichi. . . . . 82

5.3 Training curves of REFORMD and RMTPP for time prediction with *best* MAE for (a) Virginia (VI) and (b) Aichi (AI). . . . . 83

6.1 The probability of a smartphone-app category – among '*Social*', '*Travel*', '*Shopping*', '*Navigation*', and '*Music*' – to be used at ten most popular locations from Shanghai-Telecom dataset. The plot indicates that the smartphone app usage depends on the check-in location. . . . . 86

6.2 REVAMP learns the dynamics of a check-in sequence via inter-check-in time, app, and POI variations. Here, $\Delta^a_{\bullet,\bullet}$ and $\Delta^d_{\bullet,\bullet}$ denote the smartphone-app and POI-based differences respectively. . . . . 87

6.3 Overview of the neural architecture of REVAMP. . . . . 91

6.4 Architecture of different components in REVAMP. Panel (a) illustrates the setup of the Embedding Initiator (EI) that learns the category representations. Panel (b) shows the self-attention architecture in Sequential Recommender (SR). Note that the input to the self-attention is an aggregation of all past events and relative positional encodings. . . . . 91

6.5 Word-cloud for POI Categories for both Shanghai-Telecom and TalkingData datasets. The larger the font size indicates a larger frequency of location of the category. . . . . 102

6.6 POI recommendation performance of REVAMP with different methods for obtaining the relative positional encodings, *i.e.*, the inter-check-in differences between app- and location embeddings. Here, the time-based representations are kept consistent across all the models. . . . . 105

6.7 Ablation study with different *relative* positional encodings used in REVAMP and their comparison with SASRec [100] and TiSRec [116]. . . . . 107

6.8 Root-mean squared distance between the user-preference vector estimated by the model and the mean of the app- and location-category vectors of events in the test set. It shows that REVAMP is the best performer for all the datasets. . . 107

6.9 Epoch-wise recommendation performance of REVAMP for both datasets in terms of Hits@1, 5, 10. . . . . 108

7.1 Effect of unwarping on a *relevant* query-corpus pair in Audio. $U_\phi(\cdot)$ learns to transform $\mathcal{H}_q$ in order to capture a high value of its latent similarity with $\mathcal{H}_c$. . . 131

7.2 Tradeoff between NDCG@10 vs. Reduction factor, *i.e.*, % reduction in number of comparisons between query-corpus pairs w.r.t. the exhaustive comparisons for different hashing methods. The point marked as $\star$ indicates the case with exhaustive comparisons on the set of corpus sequences. . . . . 132

7.3 Query-wise performance comparison between NEUROSEQRET and best baseline methods – `Rank`-THP, `Rank`-SAHP. Queries are sorted by the decreasing gain in AP. . . . . 133

7.4 Qualitative examples of inter-event times of events in a query sequence and the top-search results by NEUROSEQRET for all datasets. . . . . 135

8.1 Real life *true* and *predicted* inter-arrival times $\Delta_{t,k}$ of different events $e_k$ for (a) Multi-THUMOS and (b) Activity-Net datasets. The results show that the true arrival times match with the times predicted by PROACTIVE. . . . . 149

8.2 Sequence goal prediction performance of PROACTIVE, its variants – PROACT-m and PROACT-$\gamma$, and other baseline models. The results show that PROACTIVE can effectively detect the CTAS goal even with smaller test sequences as input. . . . . 151

8.3 Sequence Generation results for PROACTIVE and other baselines in terms of APA for action prediction. . . . . 153

8.4 Sequence Generation results for PROACTIVE and other baselines in terms of MAE for time prediction. . . . . 153

# List of Tables

1.1 List of first-author research contributions made in this thesis. . . . . 4

2.1 Drawbacks and advantages of neural MTPP frameworks. . . . . 13

3.1 Statistics of all real datasets used in this chapter. . . . . 34

3.2 Performance of all the methods in terms of mean absolute error across all datasets on the 20% test set. Numbers with bold font (boxes) indicate the best (second best) performer. Results marked † are statistically significant (two-sided Fisher's test with $p \leq 0.1$) over the best baseline. . . . . 37

3.3 Performance of all the methods in terms of mark prediction accuracy (MPA). Numbers with bold font (boxes) indicate the best (second best) performer. Results marked † are statistically significant (two-sided Fisher's test with $p \leq 0.1$) over the best baseline. . . . . 37

3.4 Mark prediction performance of Markov Chains and IMTPP across all datasets. We use MC of orders 1,2, and 3 and report results for the best-performing model. 39

3.5 Time prediction performance of IMTPP and its variants – IMTPP$_{\mathcal{S}}$and IMTPP$_{\mathcal{R}}$in terms of MAE on the 20% test set. Numbers with bold font indicate the best performer. . . . . 39

3.6 Mark prediction performance of IMTPP and its variants in terms of MPA on the 20% test set. Numbers with bold font indicate the best performer. . . . . 39

3.7 Runtime comparison between IMTPP and PFPP in a streaming setting. Here, DNF indicates that the code did not finish within 24:00hrs. . . . . 42

4.1 Summary of Notations Used. . . . . 55

4.2 Statistics of datasets used in this chapter. The source region columns are highlighted, followed by target regions. The datasets are further partitioned based on the country of origin (US, Japan, and Germany). . . . . 65

4.3 Performance comparison between state-of-the-art baselines, AXOLOTL and its variants (AXO-f and AXO-m). The first column represents the *source* and the corresponding *target* regions. The grouping is done based on baseline details in Section 4.5.2. Numbers with bold font indicate the best-performing model. All results marked † are statistically significant (*i.e.*, two-sided Fisher's test with $p \leq 0.1$) over the best baseline. . . . . 67

5.1 Statistics of datasets used in this chapter. The source region columns are highlighted, followed by target regions, and are partitioned based on the country of source (the US and Japan). . . . . 79

5.2 Performance of all the methods in terms of mean absolute error (MAE) on the 20% test set. Numbers with bold font (boxes) indicate the best (second best) performer. Results marked † are statistically significant (two-sided Fisher's test with $p \leq 0.1$) over the best baseline. . . . . 81

5.3 Performance in terms of mark prediction accuracy (MPA) on the 20% test set. Numbers with bold font (boxes) indicate the best (second best) performer. Results marked † are statistically significant (two-sided Fisher's test with $p \leq 0.1$) over the best baseline. . . . . 81

5.4 Prediction performance of all the methods for product recommendation in Amazon datasets. Results marked † are statistically significant as in Tables 5.2 and 5.3. . . . . 82

6.1 Summary of Notations Used. . . . . 93

6.2 Statistics of all datasets used in this chapter. . . . . 101

6.3 Next check-in recommendation performance of REVAMP and state-of-the-art baselines for the Shanghai-Telecom dataset. Here, we exclude a comparison with STGN [246] as it requires the precise geographical coordinates for check-in locations, which we lack in the Shanghai-Telecom dataset. All results are statistically significant (*i.e.*, two-sided Fisher's test with $p \leq 0.1$) over the best baseline. Numbers with bold font (boxes) indicate the best (second best) performer. . . . . 104

6.4 Next check-in performance comparison between REVAMP and state-of-the-art baselines for the TalkingData dataset. All results are statistically significant over the best baseline as in Table 6.3. . . . . 104

6.5 Run-time Statistics of training REVAMP in minutes on a 32GB Tesla V100 GPU with 256 batch-size. . . . . 108

7.1 Statistics of the search corpus for all datasets. $|\mathcal{C}_{q+}|/|\mathcal{C}|$ denotes the ratio of positive corpus sequences to the total sequences sampled for training. The ratio is kept the same for all queries. . . . . 126

7.2 Hyper-parameter values used for different datasets. The values are determined by fine-tuning the performance on the validation set. . . . . 127

7.3 Retrieval quality in terms of Mean Average Precision (MAP in %) of all the methods across five datasets on the test set. Numbers with bold font (underlined) indicate the best (second best) performer. Boxed numbers indicate the best-performing state-of-the-art baseline. Results marked $^{\dagger}$ are statistically significant (two-sided Fisher's test with $p \leq 0.1$) over the best performing state-of-the-art baseline (`Rank`-THP or Sharp). The standard deviation for MASS and UDTW are zero since they are deterministic retrieval algorithms. . . . . 129

7.4 Retrieval quality in terms of NDCG@10 (in %) of all the methods. . . . . 129

7.5 Retrieval quality in terms of mean reciprocal rank (MRR in %). . . . . 130

7.6 Retrieval quality in terms of NDCG@20 (in %) of all the methods. . . . . 130

7.7 Ablation study of CROSSATTN-NEUROSEQRET and its variants in terms of Mean Average Precision (MAP) in %. . . . . 131

7.8 Training-times of NEUROSEQRET for all datasets. . . . . 133

7.9 Retrieval quality in terms of mean average precision (MAP) for query sequence lengths sampled between 10 and 50. . . . . 134

7.10 Retrieval quality in terms of mean average precision (MAP) for query sequence lengths sampled between 50 and 100. . . . . 134

8.1 Performance of all the methods in terms of action prediction accuracy (APA). Bold (underline) fonts indicate the best performer (baseline). Results marked $^{\dagger}$ are statistically significant (i.e. two-sided Fisher's test with $p \leq 0.1$) over the best baseline. . . . . 150

8.2 Performance of all the methods in terms of mean absolute error (MAE). . . . . 150