**“**Spatial Trajectory Generator**”**

***A***

***Project Report***

*submitted in partial fulfillment of the*

*requirements for the award of the degree of*

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE & ENGINEERING**

**With Specialization in**

**Oil and Gas Informatics**

**by**

|  |  |
| --- | --- |
| **Name** | **Roll No.** |
| **Priya Bansal** | **R970217030** |
| **Siddarth Kumar Jha** | **R970217040** |
| Rahul Yadav | **R970217051** |
| Anukriti Rana | **R970217052** |

***under the guidance of***

**Dr Kingshuk Srivastava**

Assistant Professor (S.G.)

Department of Informatics

upes-new-logo

DEPARTMENT OF INFORMATICS  
SCHOOL OF COMPUTER SCIENCE

UNIVERSITY OF PETROLEUM AND ENERGY STUDY

BIDHOLI, DEHRADUN, UTTARAKHAND, INDIA

**JAN,2020**

upes-new-logo

**CANDIDATE’S DECLARATION**

I/We hereby certify that the project work entitled **“Spatial Trajectory Generator”** in partial fulfillment of the requirements for the award of the Degree of BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING with specialization in OIL AND GAS INFORMATICS and submitted to the Department of Informatics at School of Computer Science, University of Petroleum & Energy Studies, Dehradun, is an authentic record of my/ our work carried out during a period from Janu**ary 2020** to **May 2020** under the supervision of **Dr. Kingshuk Shrivastava, Assistance Professor, Department of Informatics**

The matter presented in this project has not been submitted by us for the award of any other degree of this or any other University.

**Priya Bansal, Siddarth Kumar Jha, Rahul Yadav, Anukriti rana**

**(R970217030) ( (R970217040) (R970217051) (R970217052)**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date: \_\_\_/\_\_\_/ 2020 **Dr. Kingshuk Shrivastava**

Project Guide

**Dr. Thipendra Pal Singh**

**Professor and H.O.D**

Department of Informatics

School of Computer Science

University of Petroleum & Energy Studies

Dehradun – 248 001 (Uttarakhand)

**ACKNOWLEDGMENT**

We wish to express our deep gratitude to our guide **Dr. Kingshuk Shristava**, for all advice, encouragement and constant support he has given us throughout our project work. This work would not have been possible without his support and valuable suggestions.

We sincerely thank our respected Head of the Department, **Dr. Thipendra Pal Singh,** for his great support in doing our minor project at **SoCS**.

We are also grateful to **Dr. Manish Prateek, Dean SoCS**, for giving us the necessary facilities to carry out our project work successfully.

We would like to thank all our friends for their help and constructive criticism during our project work. Finally, we have no words to express our sincere gratitude to our **parents** who have shown us this world and for every support, they have given us.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Priya Bansal** | **Siddarth Kumar Jha** | **Rahul Yadav** | **Anukriti Rana** |
| **Roll No.** | **R970217030** | **R970217040** | **R970217051** | **R970217052** |

**ABSTRACT**

Datasets of the same geographic space at different scales and temporalities are increasingly abundant, paving the way for new scientific research. These datasets require data integration, which implies linking homologous entities in a process called data matching that remains a challenging task, despite a quite substantial literature, because of data imperfections and heterogeneities. In this paper, we present an approach for matching spatial networks based on a hidden Markov model (HMM) that takes full benefit of the underlying topology of networks. The approach is assessed using the heterogeneous dataset, showing that the HMM algorithm is robust in regards to data heterogeneities and imperfections and adaptable to match any type of spatial networks. It also has the advantage of requiring no mandatory parameters, as proven by a sensitivity exploration, except a distance threshold that filters potential matching candidates to speed-up the process.

Keywords: spatial networks, data matching, data integration, topology, Hidden Markov model, HMM

**TABLE OF CONTENTS**

1. **Introduction 1**

1. **Literature Review 2**

1. **Problem Statement 3**
2. **Objective 3**

1. **Methodology 4**
2. **Work Done 5-6**

1. **References 6**

**LIST OF FIGURES**

**S.No. Figure Page No**

1. **Chapter 2 2**

Fig. 2.1 Graph of K5 and K3,3 network

1. **INTRODUCTION**

The increasing development of geographical information systems (GIS) especially due to web technologies and collaborative tools is making the integration of such data by researchers and applications more and more challenging. A linear object representing spatial networks such as streets, railways, electrical or hydrographic networks is commonly the subject of many studies in geography, urbanism, sociology or history [7.3]. This project is concerned with the issue faced during the transition in such spatial networks. Spatial networks were the subject a long time ago of many studies in quantitative geography. Objects of studies in geography are locations, activities, flows of individuals and goods, and already in the 1970s, the scientists working in quantitative geography focused on networks evolving in time and space. Advances in complex networks already helped researchers to gain new insights on these difficult and the present review is an attempt to collect modern results on networks and to help researchers in various fields to reach quantitative answers and realistic modeling [7.4]. Statistical models, typically consisting of a collection of probability distributions which are used to describe patterns of variability that random variable or data may display. Markov model is one such model developed by Russian mathematician Andrei Andreyevich Markov in. It is assumed that future states depend only on the current state, not on the events that occurred before it. The simplest Markov model is the Markov chain. It models the state of a system with a random variable that changes through time. Markov chain is a probabilistic model that consists of a finite state. The states are connected through edges and some values are associated with each of these matrices But, Markov Model can't expect to perfectly observe the complete true state of the system. Thus, the Hidden Markov model is the most used model for temporal and sequence data as it observed both the hidden states and the observed states. HMM is represented using Trellis diagrams. [7.7]

1. **Literature Review**

Networks (or graphs) were for a long time the subject of many studies in mathematics, mathematical sociology, computer science, and quantitative geography. In the case of random networks, the first and most important model was proposed by Erdos and Renyi[6.5][6.6], at the end of the 1950s.Various types of related work and ongoing research that has been done on Spatial networks. The simplest example of working of the spatial network in the working of the Internet which is connected by the routers and the physical cables of variable length connected to each router. From the example, we can state that the spatial network is a geometrical figure with nodes and edges in either 2-dimension or 3-dimensional and this has important effects on their topological properties and consequently on processes that take place on them [6.4]. If the nodes are further connected to some central unit hub than the topology aspects of the network are then correlated with the spatial aspects.

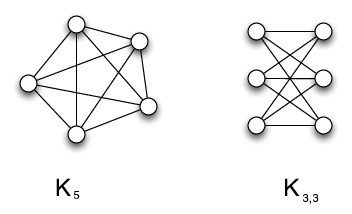


Fig 2.1 Graph of K5 and K3,3 network

Hidden Markov modeling is a powerful statistical machine learning technique that is just beginning to gain use in information extraction tasks. Hidden Markov models (HMMs) offer the advantages of having strong statistical foundations that are well-suited to natural language domains, handling new data robustly, and being computationally efficient to develop and evaluate due to the existence of established training algorithms [6.7].

1. **Problem Statement**

With the increase in geographical technology, the number of spatial networks is gradually increasing. So, it becomes more and more difficult for researchers, cartographers, historians to pinpoint the exact location and to work on such network topologies. The moment from one transition state to another is easy when the exact co-ordinates of the next transition states are provided. But, what about Traverse to the next location if the location is encoded? This project deal with such cases of transition in spatial networks.

1. **Objective**

It's time-consuming to find the traversing probability of a location from another location as there can be a single transition or multiple transitions from a state. So, using mathematical calculation we create a model to find an efficient way of transition between such states/locations.

**4.1 Sub- Objectives: -**

1. Design the Markov chains.

2. Find the Transition probabilities and matrix.

3. Find Emission probabilities.

4. Matching the spatial networks with the

5. Implement the Hidden Markov Model to find the next feasible transition.

1. **Methodology**

The overall study includes:

* Step 1. Literature Review.
* Step 2. Designing The process flow diagrams.
* Step 3. Designing of Markov Model, Markov Chains.
* Step 4. A detailed study of all the algorithms under the Hidden Markov Model
* Step 5. Implementation of the Forward-Backward algorithm for the HMM.
* Step 6. Development and Implementation.
* Step 7: Matching the spatial networks.
* Step 7. Testing HMM implementation on spatial networks.
* Step 8. Report writing.

1. **Work Done**

**Step-1 Implementation of Markov Chains.**

* In this, we have created a 2D Array that stores the encoded states codes in the Row/Column Matrix.
* Objective achieved.

**Step-2 Implementation of Forward algorithm**.

* According to the given sequence, the probability of likelihood is calculated. All the individual probabilities are summed and resultant probability is returned.
* Objective Achieved

**Step-3 Implementation of Backward Algorithm.**

* According to the given sequence, the probability of likelihood is calculated. But, unlike forward in this, it's calculated from backward.
* Objective yet to be Achieved.

**Step-4 Implementation of Viterbi Algorithm.**

* The initial, transition, emission probability are calculated using Viterbi. Unlike forward, it doesn't add up the probability instead it finds the maximum among the multiplication results and assigns to the Viterbi variable.
* Objective Achieved.

**Step-5 Calculations of Transitions probabilities between states from the algorithm.**

* The sequence achieved from Viterbi and likelihood from the Forward & Backward will then be combined to find the transition probabilities of each state of spatial network available.
* Objective yet to be Achieved.

**Step-6 Mapping of Markov chain on Spatial Networks.**

* The Markov chains are mapped on the spatial network available using Map Mapping.
* Objective yet to be achieved.

**Step-7 Combining all the above algorithms to find the trajectory.**

* The final step will be to train the algorithm to use the probabilities and find the trajectory from one state to another.
* Objective yet to be Achieved.

1. **Reference**

[7.1]. BAUM, L. E., AND PETRIE, T. Statistical inference for probabilistic functions of finite-state Markov chains. *The annals of mathematical statistics 37*, 6 (1966), 1554–1563. doi:10.1214/aoms/1177699147.

[7.2]. JIANG, B., ZHAO, S., AND YIN, J. Self-organized natural roads for predicting traffic flow: A sensitivity study. Journal of Statistical Mechanics: Theory and Experiment (July 2008). doi:10.1088/1742-5468/2008/07/P07008.

[7.3]. Costes, Benoit and Perret, Julien (2019) "A hidden Markov model for matching spatial networks," Journal of Spatial Information Science: No. 18, 57-89.

[7.4]. Marc Barth ́elemy Spatial Networks, Institut de Physique Th ́eorique, CEA, IPhT CNRS, URA 2306 F-91191 Gif-sur-Yvette France

[7.5]. P. Erdo ̈s and A. R ́enyi. On random graphs. Publications Mathematicae, 6:290 – 297, 1959

[7.6]. P. Erdo ̈s and P. R ́enyi. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci., 5:17–60, 1960.

[7.7]. K Seymore, A McCallum, R Rosenfeld, Learning hidden Markov model structure for information extraction, 1999 - aaai.org