International Journal of Engineering & Technology, 7 (4.39) (2018) 564-570

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Season-Wise Long-term Travel Spots Prediction Based on Markov Chain Model in Smart Tourism

Shabir Ahmad[1] and , Do-Hyeun Kim*[2]

Department of Computer Engineering, Jeju National University, Republic of Korea
**Corresponding author E-mail: kimdh@jejunu.ac.kr*

Abstract

Background/Objectives: Tourism plays a pivotal role in the development of a country and prediction of tourist movement can help in devising sustainable policies which in turn benefit in promoting and developing tourism. The movements of sightseers can help in estimating the amount of revenue generated for that particular sight. However, the preferences of these sightseers may change overtime and depends greatly on seasons. Therefore, a sight considered famous for one season is not that much popular for another season.
Methods/Statistical analysis: This paper proposes a probabilistic approach using Markov Chain model to predict the recommended places of Jeju Island for each season. Real data is collected from different locations using routers which are installed on each location. The proposed method is evaluated on the data and the tourist places are recommended for each season based on the steady state probabilities of each spot.
Findings: By applying the Markov Chain Model, the proposed approach accurately analyses the stochastic behavior of the data and recommend the places for each season which are predicted to be the top visiting places. The top places which are found to be equally popular across all seasons are Seongsan Ilchul and Seopjikoji.
Improvements/Applications: The data is analyzed with respect to the collected data and the popularities of the locations are found without considering the reason of the population. An improvement could be considering distance and other factors to correlate with the popularity of a certain location.

Keywords: *Tourism Management, Markov Chain, Stochastic Prediction, Locations Forecasting, Steady-state Vector, Transition Matrix*

1. Introduction

Advancement of the global tourism industry has played a significant role in flourishing economic for a country [1]. For instance, in later 2016, an estimation of 3.1% growth has been indicated by the World Travel and Tourism Council for the tourism industry, which was larger than the estimated global GDP growth (2.3%). The industry mentioned above is playing a pivotal role in providing employment. It has provided more than 100 million jobs. This is more than 3.5% of total employment in 2015. It is forecasted that it will reach 140 million by the end of 2026. In the same year, the capital investment is forecasted to be around USD 1300 billion which would mean a generation of a handful expenditure of USD 2000 billion. Consequently, the foreign tourists' expenditure contributes vitally to the global tourism industry.
The need to have an accurate prediction of foreign tourists is becoming significant for governments because it would help in setting up proper sustainable tourism development and devise strategies for promoting the tourism industry. National authorities should carefully monitor and track the changing number of foreign tourists and account for dynamically increasing and decreasing the number of tourists.
Various international tourism industries have raised a challenging task of predicting the number of foreign tourists [2]. Two categories of methods can be used to develop location forecasting models, deterministic and stochastic methods. The deterministic model assumes that the location transition from one place to another is certain and no alternate route exists, so regression analysis is commonly used to determine the routes [3–5]. These models are often relatively convenient to use but fail to consider uncertainty and randomness of the tourist for long-term [5]. Stochastic models are better in such aspects. Engineering experience can be used to describe the uncertain factors [6]. Discrete probabilistic models based on state and time are effective in predicting the tourists' behaviour [7]. The discrete probability models are represented by a Markov process, which is based on the concept of probabilistic cumulative effort and now commonly used in performance prediction of of random behavior in any data. It is believed that the Markov process has three advantages [8]. Firstly, it can reflect uncertainties in various aspects [9]. Secondly, the prediction of a future state is based on the present state so the model is incremental [7]. And lastly, it can be applied on a network level with many facilities involved which requires calculation efficiency [10]. Recent researchers have adopted the Markov model to predict the performance of safety-critical systems like bridges such as Pontis and BRIDGIT [11, 12] which shows the high reliability of the model even in such scenarios.
Although the Markov process has been widely used in bridge performance prediction, most applications only focus on some bridge component such as a deck or directly consider the bridge as a whole. Besides, the sample size is small in most applications [13]. Every bridge component, as well as the whole bridge, has different deterioration characteristics and the maintenance status usually has varying degrees of impact on different bridge components. Therefore it is necessary to develop a Markov model

based on large samples to predict the deterioration tendency of various components of urban bridges, which is also the objective of this study.

This paper is an attempt to find a place to predict the top spot for each season inside Jeju Island, South Korea. Markov Chain model is applied on a real dataset which has been collected from routers installed on various picnic spots. The long-term prediction of spots and the probabilities greatly helps in devising adaptable strategies for the advancement of facilities and attracting more tourist across the globe. Thus, this would lead to the overall growth of revenue of the Island.

The remainder of this paper is organised as follows. Section 2 discusses the Markov Chain Modeling for location forecasting and prediction. Section 3 discusses the methodology of the proposed work and outlines various experiments performed. Section 4 describes the data used in the paper and presents preprocessing techniques. Section 5 presents the results, and section 6 finally concludes the paper and identifies future directions.

2. The Markov Chain Model

Markov process is widely utilized in modeling the dynamics of stochastic systems and the state transitions of these complex stochastic systems. A Markov process denoted by $\{M(t),\ t \epsilon T\}$ is a stochastic process with the characteristics that, for a given state of the system at time $\{t\ , M(t)\}$, for a time $\{s > t\ \}$, the system state $M(s)$ is not affected by the system states. $M(u)\ for\ u < t$ that is, prior to the time t. P is a transition matrix; the matrix elements are the probabilities of transition of row to columns, and the sum of all the various elements in a row is equal to 1. The transition matrix P is a square matrix in the form:

$$\begin{bmatrix} p11 & \cdots & p1j \\ \vdots & \ddots & \vdots \\ pi1 & \cdots & pij \end{bmatrix} (0\ \leq pij\ \leq 1) \qquad (1)$$

In other words, the probability of any particular future behavior of the process, if its current state is known, will not be altered by any additional knowledge concerning its past behavior. A discrete-time Markov chain is a Markov process whose state space is a finite or countable set with time interval indexes as $t = (0, 1,2 \ldots)$.

In formal terms, the Markov property is that

$$P\{M(tn) = mn\ |M(t1) = m1,\ M(t2) = m2, \ldots ,\ M(tn-1) = mt - 1\} = P\{M(tn) = mn\ |\ M(\mathrm{tn}-1) = mn - 1\} \qquad (2)$$

The probability of $Mn + 1$ being in state j given that Mn is in state i is called the one-step transition probability and is denoted by Multistep transition probability can be calculated according to one-step transition probability and the Markov property as follows:

$$\mathrm{P(Mn + 1 = mj\ |Mn)}$$
$$P(Mn + 2|\ mn) = \int P(Mn + 2,\ Mn + 1\ \big|\ Mn)dMn + 1 = \ PMn{+}2\ \ Mn{+}1)PMn{+}1\ Mn)dXn{+}1 \qquad (3)$$

If the transition matrix is irreducible and a periodic, the -step transition matrix converges to a stationary distribution π with each column different; that is,

$$\lim_{n\to\infty} p^n = \pi \qquad (4)$$

According to Markov chain properties in finite state space, stationary distribution π can be calculated by

$$\pi p = \ \pi \qquad (5)$$

Markov chain is a special case of Markov process defined as follows.

First, assume $\{M(t),\ t \epsilon T\}$ is a random process, if the state of $\{M(t),\ t \epsilon T\}$ is known at t_0 and is irrelevant to the states of $\{M(t),\ t \epsilon T\}$ at t $>t_0$, then it is said that $\{M(t),\ t \epsilon T\}$ has a Markov property.

Secondly, assume S is a state space of $\{M(t),\ t \epsilon T\}$, for $n\ \geq 2$ and $t < t2 < \cdots . t_n \epsilon T, M(t_i) = m_i, m_i \epsilon S, i = 1,2,3 \ldots n - 1$. The conditional probability for a future state with respect to a past state can be written as follows:

$$P\{M(t_n) \leq\ m_n|M(t_1) =\ m_1\ , M(t_2) =\ m_2, \ldots \ldots . M(t_{n-1})\ = \ m_{n-1}\} = P\{M(t_n) \leq\ m_n\ \mid M(t_{n-1})\ =\ m_{n-1}\ \} \qquad (6)$$

Then, $\{\ M(t),\ t \epsilon T\ \}$ is a Markov process. In this paper, the prediction of the locations is based on the aforementioned Markov Model. The transition matrix is the location transition from one place to another place and the steady state is predicting the probabilities at tn in future.

3. Proposed Travel Spot Prediction Methodology

The goal of the paper is to perform a series of experiments to refine and preprocess a dataset and extract features. Once it has been preprocessed, the next step is to apply the model. The main parts of the preprocessing phase is tagging, feature extraction and season-wise splitting.

Figure 1 shows the proposed methodology for the prediction. A series of experiments are performed to pre-process the data, extract the features of more interests and to filter out the irrelevant data. Once final data is achieved, the Markov model is applied to the data and on the basis of long-term probabilities, the spots are predicted for all the seasons. In order to find long-term probabilities, the transition matrix is computed. A transition Matrix is an m x n matrix showing the transition probability from one location to another location. If the transition matrix is irreducible and periodic, then the n-step transition matrix converges to a stationary distribution. This steady transition helps to predict the location in the long term because it predicts for infinite location as shown in the equation 4.

4. Experimental Results and Discussion

A real-time data based on the mobile tourist is gathered for the year 2017-2018 for different locations of Jeju Island, South Korea. The data is collected from the routers installed on different locations. The data has attributes like date of the connection. The date is split across month, half and quarter for granularity purposes. The quarter is showing the season code. Apart from this, the travel path is recorded showing the tourist movements across different locations. Additionally, the number of tourists is counted for the route, and the duplicate count is also noted down which the number of tourists who travels part of the recorded route is. The difference between total path count and partially met vehicle count is that in case of total path count the mobile tourist covered the whole trajectory while in a later case the tourist covered part of the trajectory. The structure of the dataset is shown in figure 2.

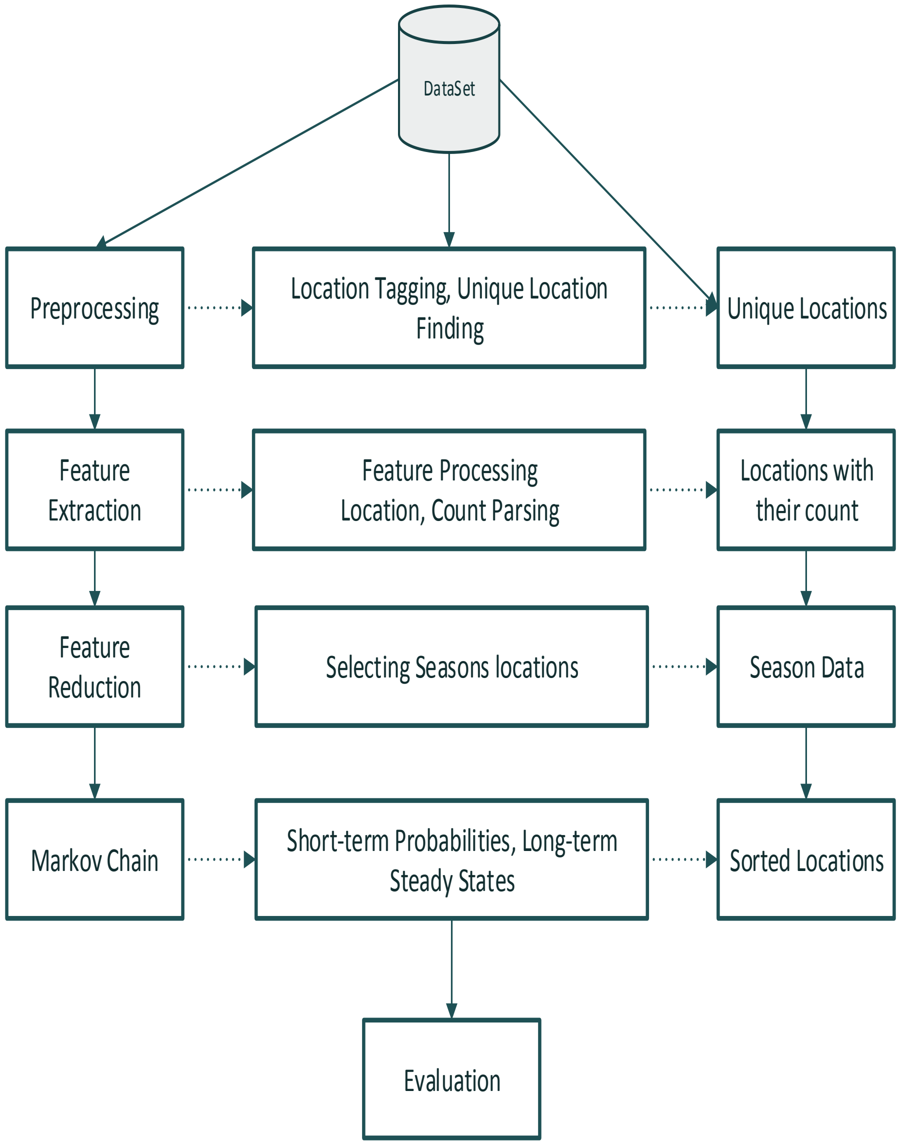


Figure 1: Proposed Long-term Travel Spot Prediction Methodology

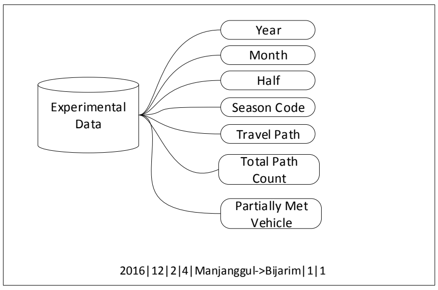


Figure 2. Dataset Snapshot

4.1 Data Preprocessing and Season-wise Data Filtering

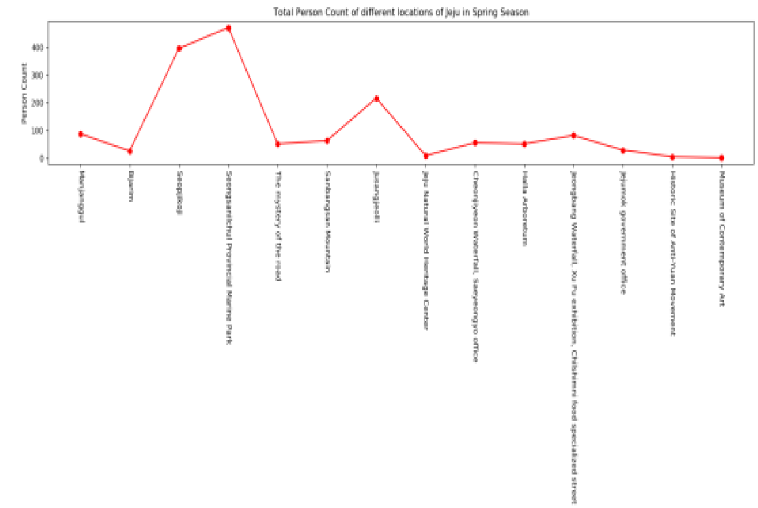
In this part, a number of experiments have been performed to present the data in a form which can be easily consumed by the Markov Chain model. In this stage, locations have been tagged for efficient processing. A location is tagged using the index of the location and the first three letters of the location. For instance, Seopjikoji is tagged as SEO-002. Once the locations are tagged the tagged locations are replaced in the preprocessed dataset. Secondly, the trajectories are in an arrow separated string, so another experiment is performed to parse the list and make a panda data frame to show the location and the total tourists passed through the location. Table 1 shows the result of tagging and data preprocessing. The data is then split based on season code. The dataset is filtered much like select database operators. As a result, four sub-datasets are achieved.

Table 1: Data Preprocessing and Tagging

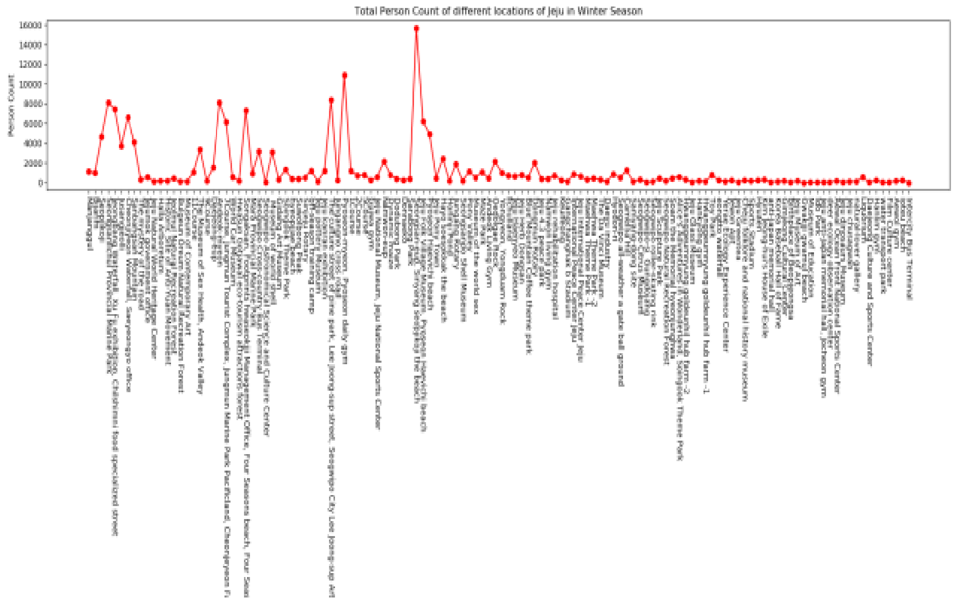
Index	Location	Total Count
1	Mang-001 <Mangangul>	2120
2	Seop-002 <Seopjikoji>	1110
3	Song-003 <Songaksan>	1222
4	Samb-004 <Sambangsan>	3770
5	Bija-005 <Bijarim>	2901
6	Hall-006 <Halla Arboretum>	2160
….	………	………

4.2 Season-Wise Data Collection

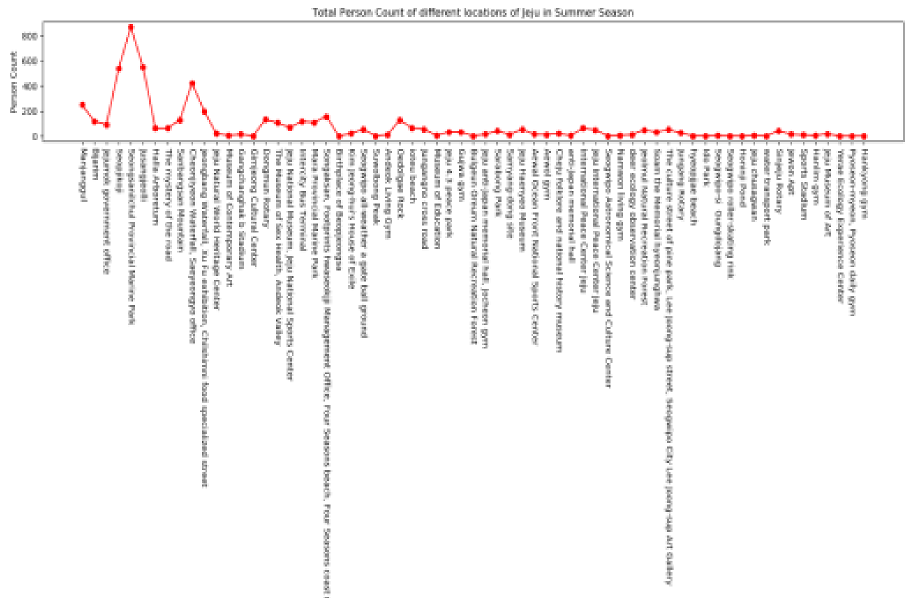
In this paper, the focus is the find the hot tourism spots for every season based on the dataset. So after preprocessing the data splitting it across four seasons. We assume season_code represents season and it has four distinct values. For simplicity, 1 represent spring, 2 represents summer, 3 represents fall and 4 represents winter season. Figure 3 shows the season-wise location and their respective counts of tourists.



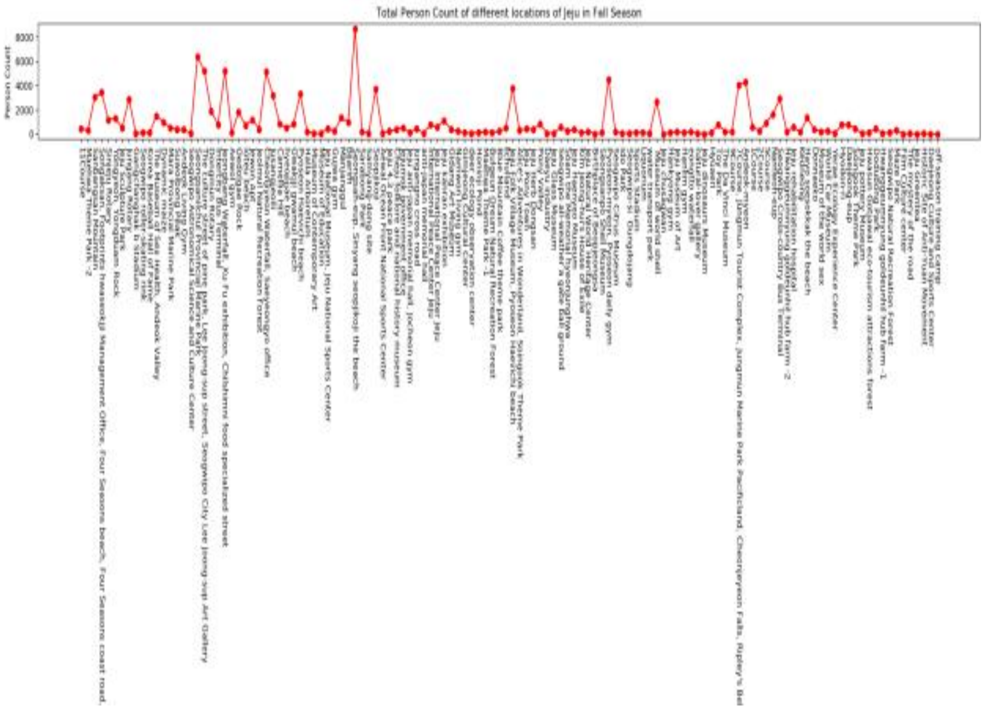
(a) Spring Season



(b) Winter Season



(c) Summer Season



(d) Fall Season

Figure 3. Total Visitors per Location in different Seasons; (a) Spring Season, (b) Winter Season, (c) Summer Season, (d) Fall Season

4.3 Transition Matrix Calculation

Transition Matrix is one of the preliminary steps in applying the Markov Model on a dataset. In this paper, the states are the existence of tourist on a location and the transition is the movement of the tourist from that location to a next location. Thus, the rows and the columns show the locations and the values of the matrix are the transition probabilities. For this use case, the input data is the season-wise locations with their respective counts. The data is a two-dimensional data frame. The indices show the locations while the column represents the counts. In order to make a transition matrix, the total count is found for each transition. A transition is represented as Loc A -> Loc B, so a string matching algorithm is applied to search for every location dynamically and if found the respective count against that index is recorded. In the end, the duplicate transition is summed up to find the unique transitions. Table 2 shows the portion of transition Matrix for Spring season.

Table 2: Transition Matrix Sample from Spring Season

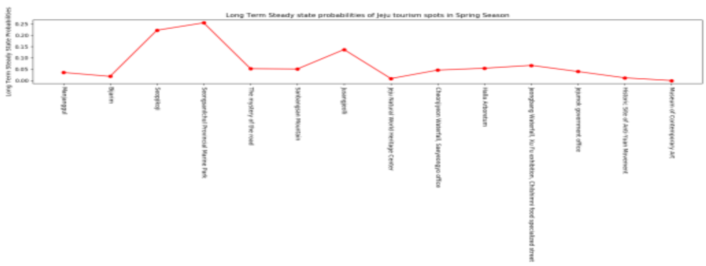
	Manjang gul	Bijari m	Seopjik oji	Seongsanilc hul	Sanbangs an
Manjanggul	0	0.035	0.107	0.732	0
Bijarim	0.166	0	0.416	0.017	0
Seopjikoji	0.057	0.050	0	0	0.006
		9			
Seongsanilc hul	0.066	0.015 8	0.733	0	0.009
Sanbangsan	0	0	0	0.126	0

4.4 Steady State Convergence

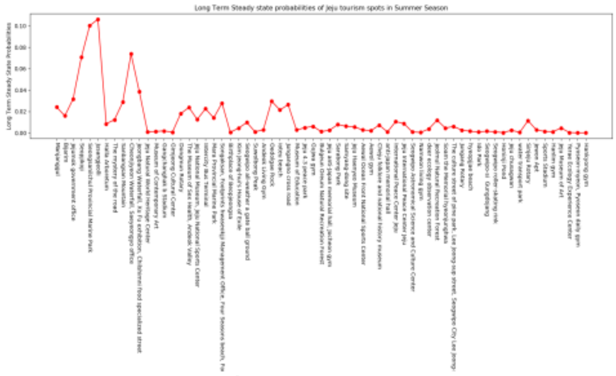
Once the transition matrix is computed, the probabilities distributions are converged after a finite interval of time. After this state, the probabilities are not changed if another vector is multiplied. This state shows the long-term prediction results. It should be noted that the results of this convergence may be different from the current result which is based on the current data and it shows no stochastic behaviour. However, for long-term, the prediction can be found using the convergence matrix.

5. Execution Results and Discussions

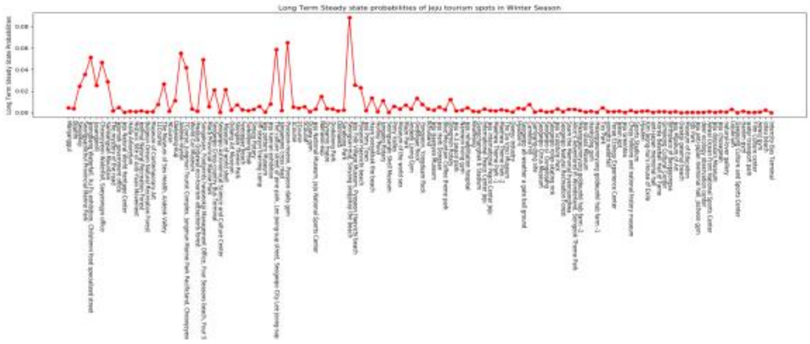
The steady-state probabilities of various locations predict the forecasting of tourists in the future. The steady state is found based on the equation $\pi p = \pi$, and the steady-state probabilities are recorded for different seasons. Figure 4 shows the steady-state probabilities for various seasons.



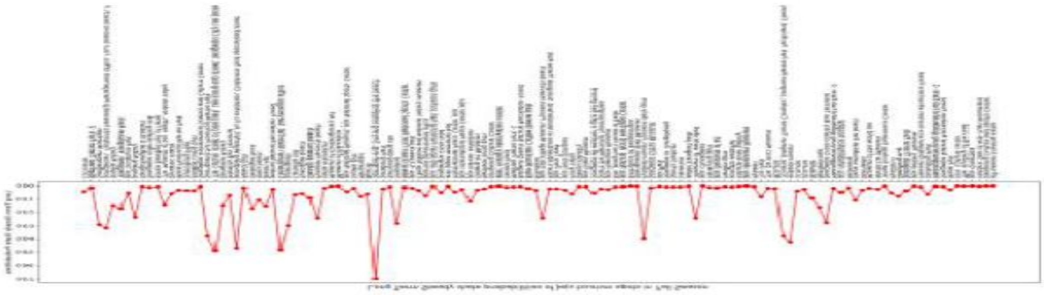
(a) Spring Season



(b) Winter Season



(c) Summer Season



(d) Fall Season

Figure 4: Season-wise location prediction based on steady state probabilities; (a) Spring Season, (b) Winter Season, (c) Summer Season, (d) Fall Season

The top 3 spots are those spots which have the highest probabilities value for that season. Based on the graph the recommended places for all four seasons are summarized in table 3, table 4, table 5 and table 6 respectively.

Table 3: Spring Season

Rank	Location	Probability
1	Seongsan-eup, Sinyang seopjikoji	0.255039
2	Pyoseon-myeon, Pyoseon daily gym	0.222229
3	The culture street of pine park	*0.136936*

Table 4: Summer Season

Rank	Location	Probability
1	Jusangjeolli	0.106
2	Seongsanilchul Provincial Marine Park	0.100
3	Cheonjiyeon Waterfall, Saeyeongyo office	*0.07*

Table 5: Fall Season

Rank	Location	Probability
1	Seongsan-eup, Sinyang seopjikoji	0.0696
2	The culture street of pine park	0.04893
3	Cheonjiyeon Waterfall	*0.04846*

Table 6: Winter Season

Rank	Location	Probability
1	Seongsan-eup, Sinyang seopjikoji	0.15
2	Seopjikoji	0.12
3	Jusangjeolli	*0.093*

It should be noted that some places are commonly recommended for every season. Nevertheless, they have different ranks for different seasons, but they made it to the top-ranked for every season. These places are the global recommended spots irrespective of the season. In other words, these spots are not affected by the season, and the number of tourists visits these places remains unaffected.

5. Conclusion and Future Work

In this paper, we worked around a real travel data and applied the Markov Chain model to find the long-term behaviour of the data for each season. The data is pre-processed and converted season-wise splits. Markov chain is then applied for every season, and the result is recorded. It was found that some locations are repeating for every season. These are globally popular irrespective of each season. Some locations are popular for a specific season, and the same location is considered less popular for different locations. The future work of this research can be the correlation between the locations' popularity with external factors like distance and type of travelling spots.

Acknowledgments

This paper was performed for the Development of Radar Payload Technologies for Compact Satellite in Korea Aerospace Research Institute, funded by the Ministry of Science and ICT, and this research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2014-0-00743) supervised by the IITP (Institute for Information & communications Technology Promotion), Any correspondence related to this paper should be addressed to DoHyeun Kim.

References

[1] Hu YC. Predicting foreign tourists for the tourism industry using soft computing-based Grey–Markov models. Sustainability. 2017 Jul 13;9(7):1228.
[2] Mao S, Tu E, Zhang G, Rachmawati L, Rajabally E, Huang GB. An automatic identification system (AIS) database for maritime trajectory prediction and data mining. InProceedings of ELM-2016 2018 (pp. 241-257). Springer, Cham.
[3] Gonzalez H, Han J, Li X, Myslinska M, Sondag JP. Adaptive fastest path computation on a road network: a traffic mining approach. InProceedings of the 33rd international conference on Very large data bases 2007 Sep 23 (pp. 794-805). VLDB Endowment.
[4] Haldankar I, Tiwari M, Usha G, Aruna S. An Intelligent Framework for Road Safety and Driver Behavioral Change Detection System Using Machine Intelligence. InComputational Vision and Bio Inspired Computing 2018 (pp. 913-924). Springer, Cham.
[5] Li L, Miao J, Fang S, Yang IH, Wang J, inventors; Baidu USA LLC, assignee. Method and system to predict vehicle traffic behavior for autonomous vehicles to make driving decisions. United States patent application US 15/359,466. 2018 May 24.
[6] Giannotti F, Nanni M, Pinelli F, Pedreschi D. Trajectory pattern mining. InProceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining 2007 Aug 12 (pp. 330-339). ACM.
[7] Collins L. An introduction to Markov chain analysis. Geo Abstracts Ltd..
[8] Morcous G. Performance prediction of bridge deck systems using Markov chains. Journal of performance of Constructed Facilities. 2006 May;20(2):146-55.
[9] Lounis Z. Reliability-based life prediction of aging concrete bridge decks. Life prediction and aging management of concrete structures. 2000:229-38.
[10] Morcous G, Rivard H. Computer assistance in managing the maintenance of low-slope roofs. Journal of computing in civil engineering. 2003 Oct;17(4):230-42.
[11] Golabi K, Shepard R. Pontis: A system for maintenance optimization and improvement of US bridge networks. Interfaces. 1997 Feb;27(1):71-88.
[12] Thompson PD, Small EP, Johnson M, Marshall AR. The Pontis bridge management system. Structural engineering international. 1998 Nov 1;8(4):303-8.
[13] Cesare MA, Santamarina C, Turkstra C, Vanmarcke EH. Modeling bridge deterioration with Markov chains. Journal of Transportation Engineering. 1992 Nov;118(6):820-33