Periodic Pattern Discovery Algorithm of Migratory Bird

Zhengzheng Wei [1, 2, a], Zengzhen Shao [1, 2, b] and Dong Chen [1, 2, c]

[1]School of Information Science and Engineering, Shandong Normal University, Jinan, 250014, China

[2]Shandong Provincial Logistics Optimization and Predictive Engineering Technology Research Center, Jinan, 250014, China

[a]wzzsdnu@163.com; [b]shaozengzhen@163.com; [c] haohaoxuexilema@163.com;

Abstract

Migratory birds possess a close relationship with humans and demonstrate a great impact on human's life. Bird migration has attracted an increasing attention. Previous research, however, has encountered many problems, such as ineffective data processing and analysis methods. In this paper, periodic pattern discovery algorithm is illustrated based on data-mining technology in accordance with the spatio-temporal characteristics of bird-watching data. Then periodic pattern discovery algorithm are applied to discover bird migration pattern and their periodic migration routes.

Keywords

Migratory birds, bird migration, periodic pattern.

1. Introduction

All groups of organisms birds are special in that a large number of species migrate annually between their breeding and non-breeding areas[1]. The migration of birds has a great impact on the environment and production life of human beings. Studying the migration of birds and **discov**ering periodic pattern can help people prevent the spread of epidemics and maintain species diversity. Meanwhile , human activities will also affect the migration and habitat selection of birds [2]. Therefore the study of migration process and migration habitat is crucial for people to protect the birds and natural environment and maintain species diversity.

Various methods are adopted and developed to carry out research on migratory birds from different aspects by domestic and foreign researchers in order to understand the migration patterns of birds. Among these methods, the fixed-point investigation is the earliest bird migration research method. The most common and most popular method to study the migration of birds is bird-banding[2]. It can be easily implemented and widely applied. But its monitoring cycle is long and the data recycling is quite complicated[3]. It is also unable to produce noticeable effects in a short time. Next method is the satellite positioning method. Its accurately collected data can achieve continuous tracking to the individual bird. However it is not suitable for small birds with high cost and difficult popularization as well as limited amount of data. There are also some other methods including radar monitoring, sensitive geographical location and others. But they have low precision, difficulty in popularizing and limited data and other issues. In addition, the usage and analysis of the collected migration data of birds has also attracted the attention of researchers at home and abroad. It is analyzed by the early bird data that only through the track point marked in GIS by biologists or the distribution points gotten by artificial statistics can migratory lands and migration routes be acquired [4]. In 2004, the Japanese scientist Shimazaki proposed to deal with bird flight data through using the method of ISODATA clustering[5]. In this method the migratory state of birds is determined in accordance with their flight speed and further the migratory location is obtained. Nonetheless the migratory routes of birds still need to be manually marked and further the spatial location information is unable to be processed with this method. In 2010, Zhou Yuanchun and others found the birds gathering land by using the density-based hierarchical clustering algorithm to cluster bird GPS, the association of the aggregation rules and further migratory routes of birds by virtue of the Apriori algorithm or GSP algorithm[6]. However some defects that the changes of the habitats and migratory routes of birds cannot be found still exist due to small amount of data, less number of birds and short time span of the data.

In 2012 Li Xueyan and others established a bird-watching database using China Birding Report and displayed the changes of bird distribution of recent years by GIS[7]. But it only relies on artificial statistical methods to mark bird discovery sites in the GIS, neither using the "quantitative observation" to analyze the data deeply, nor dealing with the problems of repeated sampling and uneven distribution for the bird-watching data.

From the above there are still a lot of problems that need to be solved in the above research which are embodied in: 1) The amount of data for analysis and study is relatively small. 2) The time-space relationship as well as migration cycle about migratory birds are rarely involved in. 3) The utilization and processing of birds data demonstrates much limitations and the hidden knowledge and periodic pattern in the data fails to be dug out[2].

In this paper, A feasible, efficient and general method is sought to solve the above problems, to discovery bird migration pattern and their periodic migration routes.

2. Data and Data preprocessing

2.1 Data source

As the important supplementary 2.2 information of traditional bird distributions, Chinese bird-watching data is comprehensive and reflects Chinese bird watching achievements accurately. These data comes from three aspects. The first is from the network such as Bird Report (www.birdreport.cn)[8] and China Bird Watching Network (www.chinabirdnet.org)[9] etc. The second comes from ornithological books and literature such as "China Bird Report 2003-2007" [10], "China Coastal Waterbird Census Report 2005–2011" [11] and "A Checklist and Distribution of the birds in Shandong" [12] etc. The third is provided by many ornithologists led by Prof. Sai Daojian. Total 189350 bird watching records have been verified by ornithologists that insure the authority of these data.

2.2 Data characteristics

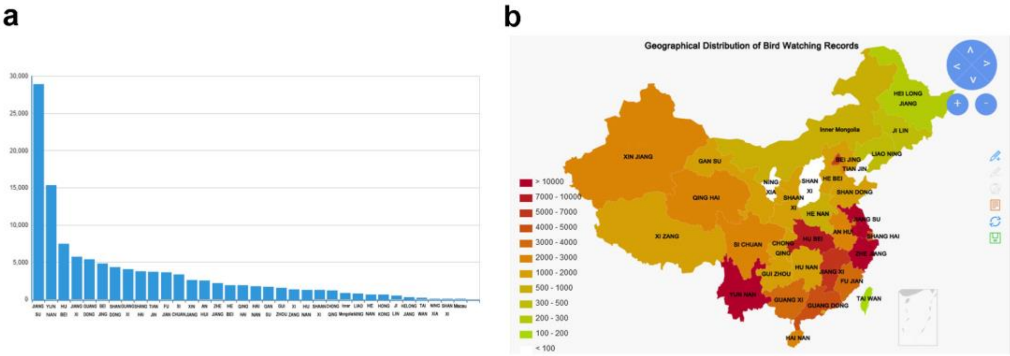
Chinese bird-watching data records spatio-temporal information including species, date, location, number and observer in detail as shown in Fig. 1. The following "Number" records the number of birds observed in each bird watching.

BirdsID	BirdsEngName	BirdsLatinName	Address	Number	ActionTime	Observer
0257	Indian Jungle Nightjar	Caprimulgus indicus	Chongming Dongtan NR,Shanghai	6	2003/9/20	NR
0547	Black-crowned Night Hero	Nycticorax nycticorax	Futian NR, Shenzhen,Guangdong	300	2003/1/26	HK
0343	Common Redshank	Tringa totanus	Huahu, Zoige,Sichuan	30	2006/5/5	SY
0345	Common Greenshank	Tringa nebularia	Yellow River Delta NR, Dongying,Shandong	58	2006/4/15	SK
0096.1	Common Teal	Anas crecca	East Dongting Hu, Yueyang,Hunan	980	2006/2/12	LJY
0614	Brown Shrike	Lanius cristatus	Tianjin Normal University,Xiqing District, Tianjin	12	2015/5/20	NK
0518	Little Grebe	Tachybaptus ruficollis	Ma'anshan Forest Park,Hongshan District,Wuhan,Hubei	10	2015/6/14	HWJ
0542	Great Egret	Casmerodius albus	East Dongting Hu NR, Yueyang,Hunan	2	2015/12/12	XS
0349	Wood Sandpiper	Tringa glareola	Miyun Reservoir,Miyun,Beijing	12	2012/4/22	TT
0067	Whooper Swan	Cygnus cygnus	Yellow River Delta NR,Shandong	1000	2005/3/6	CHY
0068	Tundra Swan	Cygnus columbianus	Poyang Hu NR,Jiangxi	3497	2005/2/16	WWFC
0069	Swan Goose	Anser cygnoides	Shengjinhu NR,Anhui	24211	2005/2/19	WWFC
0090	Spot-billed Duck	Anas poecilorhyncha	Jiaocheng, Ningde,Fujian	150	2003/10/2	XY
0092	Northern Shoveler	Anas clypeata	Futian NR, Shenzhen,Guangdong	1000	2003/1/5	YY
0357	Great Knot	Calidris tenuirostris	Yalu Jiang NR, Dandong,Liaoning	1700	2006/5/13	BQ
0085	Gadwall	Anas strepera	Yellow River Wetland NR, Mengjin,Henan	108	2005/2/10	DQ

Fig. 1 Chinese Bird-Watching Data

2.3 Data Statistics

The 239350 Chinese bird-watching records involve 24 Orders, 100 Families, and 1230 Species accounting for 85.7% of China's existing bird species [13]. The range of these records covers 34 provinces, municipalities and autonomous regions, including Hong Kong, Macao and Taiwan approximately 46 years ranging from 1970 to the present. Jiangsu Province and Yunnan Province enjoy the two maximum bird-watching records with 289761 and 14363 records respectively. Macao has the fewest records, only 101, as shown in Fig. 2a. The distribution of bird-watching records in each province is showed in Fig. 2b.



a Columnar chart for the number of records. b National distribution map of bird-watching records.

Fig. 2 The Number of Records of Each Province

The number of bird-watching records is more in the east than that in the west, and more in the south than that in the north. The proportions of bird-watching records also vary with years. 2015 enjoys the highest proportion of bird-watching records reaching 33.57%. Then is the 2014 and 2004 accounting for 10.52% and 10.49% respectively. The proportion of other years is less than 10%. The number of records from 2001 to 2016 accounts for more than half of the total. In Fig. 3

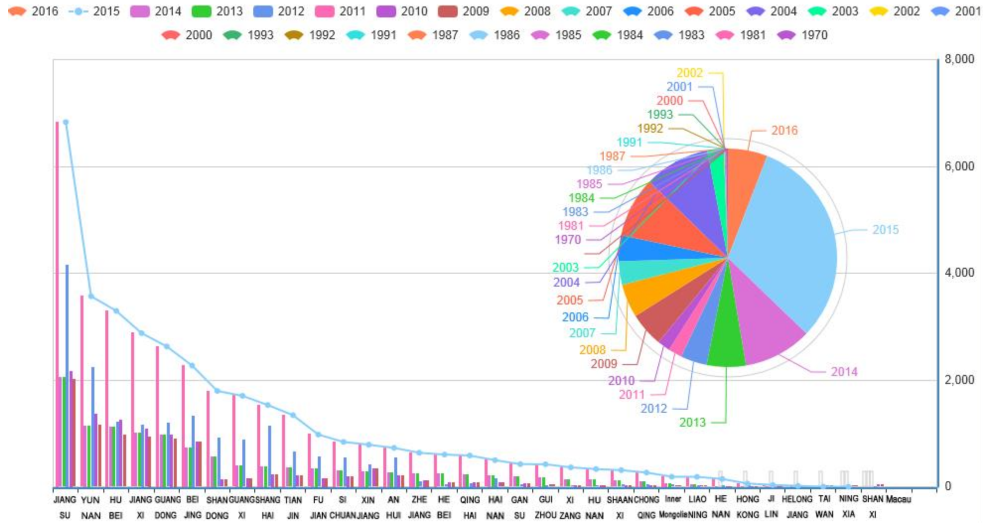


Fig. 3 The number of records of years (1970-2016)

Thus, it can be inferred that the bird watching records in China are distributed unevenly in species, space and time. There are many reasons for this situation. First bird watchers have an uneven geographical distribution. Second some species are highly concerned (such as Alcedo) and are easily observed (such as Passer montanus), which easily causes repeated sampling [7]. Third the enthusiasm of bird watchers is also an important factor. Some bird watchers can provide 50 records per year. However there are a considerable number of watchers who just provide less than 1 record per year. Fourth bird watching activities started relatively late in China and the number of activities and population of bird watching not experienced rapid growth until in recent years, which cause uneven distribution of the records in time.

Despite the above flaws, the advantages of bird-watching data are obvious such as substantial data, long-term span, and high accuracy. These advantages can provide great convenience for the analysis and mining of bird data.

2.4 Data processing

Bird migration is a relatively long and complex process which reflects in spatial and temporal changes directly. In this paper Chinese bird-watching data is used as the data source to explore the law of change on time and space during the migration of birds. In order to explore the periodic patterns of migratory birds, three steps of data processing be taking:

1) The temporal and spatial distance of bird-watching data is calculated between the points to discover implicit duplicate data;

2) The special points instead of the duplicated bird-watching data are used to produce a standardized data set;

3) Standardized bird-watching data set is clustered by the density-based clustering algorithm to obtain the high density area of the migratory activity that is used as a habitat during the migration.

3. Algorithm

3.1 Algorithm construction

Migratory birds migrate periodically between their non-breeding and breeding grounds. Therefore there is a time correlation between migratory birds and their habitats. The temporal pattern of migratory birds hope to be found in each habitat. Then migration routes with the temporal pattern will be generated. Migratory bird ej is regarded as event ej.

First a habitat of dl (ej) is selected randomly and a continuous timeline is divided into discrete points using a certain time granularity. Second the timestamps of all points in dl (ej) will be preserved. These timestamps are treated as a time series of ej. Finally the time series of ej is dealt: 1) detecting possible periods; 2) identifying the validity of each possible period; 3) using the valid periods to generate the periodic patterns.

Definition Periodic pattern: Periodically recurring patterns of ej can be expressed as: Per_Patternsl(ej) = <dl(ej), Tl(ej), Rep> where dl(ej)∈D(ej), l∈[1,L] and L is the number of habitats. Rep is the times of the periodical repetition. Tl(ej)=(lengthl, ta, tb, ej) where lengthl is the period length, ta is the starting time and tb is the ending time. Here, ta ≤ tb and lengthl > 0.

The basic steps of mining for periodic patterns are as follows:

Step1: Using the appropriate time granularity to divide the timeline. The continuous time is divided into discrete points;

Step2: Creating the HES (Habitats and Events Schedule): The timestamps of all points in dl (ej) will be preserved. These timestamps can be considered as the time series of ej;

Step3: Setting parameters. The length of the minimum period is Min_length, the maximum period Max_length. The values of Min_length and Max_length are set by the user. A period that needs to repeat itself at least a certain number of times is called min_rep, a value set by the user. Count_length is used to accumulate the counts for each period length and its initial value is 0. Rep is to accumulate the number of consecutive matches for a period length. The count of the time series of an event is Count_Time.

Step4: Scanning the Habitats and Events Schedule for an event ej, creating a sliding window and comparing the duration between any two time instants (tα and tβ). Length=tα-tβ, α∈[1,Count_Time], β∈[α+1,Count_Time];

Step5: If Min_length ≤ length ≤ Max_length, then Count_length = Count_length +1; if not, a sliding window is created from tα+1;

Step6: If Count_length ≥ min_rep, then the length will be a possible period; possible periods: Pot_length(ej) ={pot_length1, pot_length2,…,pot_lengthλ,…,pot_lengthγ }, λ ∈ [1,γ], γ is the number of the possible periods;

Step7: This process (Step4- Step6) won’t be repeated until all possible periods of ej have been detected.

Step8: Creating a PPS (Possible Period Schedule). In this table, Pot_length is used to record all possible periods of ej, Time'∈Time.

Step9: A Hash List is created. For each possible period pot_lengthλ($\lambda \in [1,\gamma]$) of an event ej in PPS the modulus mod = tα'% pot_lengthλ, tα' ∈Time' is computed. A Hash List is built for pot_lengthλ.

Step10: If the remaining list satisfies the following two conditions: a. the mod_rep length of the linked list is larger than that of the min_rep; b. the values of the linked list are an arithmetic sequence and the common difference is pot_length λ, then pot_length λ is a valid period of the bird activity. The arithmetic sequence is a valid sequence. It must repeat the number of times Rep = mod_rep-1. The starting value of the arithmetic sequence is the pot_lengthλ and the last value is the ending time.

Step11: Output all valid periods and the periodic patterns of ej.

Example. Taking e1 as an example:

1) The HES is created to recognize possible periods from time series just as Table1 shows. Min_length=2, Max_length=7, min_rep=2, Count_Time =13. A sliding window is set in Fig.4 and PPS (in Table2) shows e1.

2) The possible period for e1 is pot_lengthλ=5 with min_rep=3. A valid period is discovered by applying Hash List as shown in Fig.5 and Table 2. In this picture, {0, 1, 2, 3, 4} are the remainders when pot_lengthλ is divided by tα', tα'={2, 3, 7, 8, 13, 18, 20, 25}, tα' ∈Time'. The remainder is 3, mod_rep =4 > min_rep and the common difference of the arithmetic sequence is 5. Hence the valid periodicity of e1 is 5, its time position is Time' = {3, 8, 13, 18} and Rep=3.

3) Generating the periodic patterns: Per_Patternsl(e1) = <dl(e1), Tl(e1), 3>; Tl(e1)=(5, 3, 18, e1).

Table1. Habitats and Events Schedule

Event *E*	Habitats	Time
e_1	$d_l(e_1)$	1, 2, 3, 4, 7, 8, 10, 11, 13, 18, 20, 22, 25

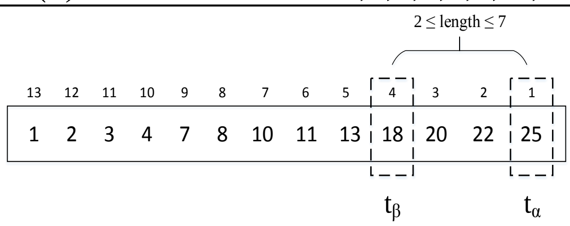


Fig.4 Sliding Window of HES

Table2. Possible Period Schedule

Event *E*	Pot_length	Time'
e_1	2	1, 2, 3, 4, 7, 8, 10, 11, 13, 18, 20, 22
	3	1, 4, 7, 10, 13, 22, 25
	4	3, 4, 7, 8, 18, 22
	5	2, 3, 7, 8, 13, 18, 20, 25
	6	4, 10, 18, 22

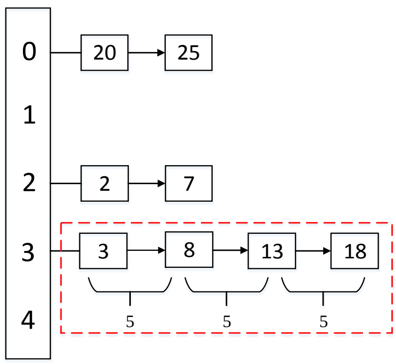


Fig. 5 Hash List of PPS (Pot_length=5)

3.2 Time Complexity Of Algorithm

In the mining of periodic patterns of migratory birds, the SMCA algorithm is improved. This algorithm only needs to scan the database once to find possible periods; the length of the database is n. Then, this algorithm needs to scan the PPS table γ times, where γ is the number of possible periods; In the worst case, there are n data to be scanned every time, γ<<n. So, its time complexity is O(γn+n) << O(n2). This shows that improved SMCA algorithm can be completed in a relatively short time.

4. Experiment

First the time series is divided according to a suitable time granularity (from 1970 to 2016). But years without bird-watching records are ignored. Then a 10-day time granularity is set. The timeline is divided into discrete time points as the unit of 10 days with a collection of years CY={2015, 2014, 2013, 2012, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1993, 1992, 1991, 1987, 1986, 1985, 1984, 1983, 1981, 1970} as in Table3.

Table3. Time granularity mapping table

Time	Data Points
2015/12/31-2015/12/22	1
2015/12/21-2015/12/12	2
2015/12/11-2015/12/02	3
2015/12/02-2015/11/23	4
2015/11/22-2015/11/13	5
2015/11/12-2015/11/03	6
……	
1970/02/25-1970/02/16	945
1970/02/15-1970/02/06	946
1970/02/05-1970/01/27	947
1970/01/26-1970/01/17	948
1970/01/16-1970/01/07	949
1970/01/06-1970/01/01	950

Second taking advantage of the conclusions of the second experiment the habitats for migratory birds are selected to mine periodic patterns.

With ej= Hirundo rustica d3 is set up as the habitats. Regional events schedule of positions (in Jiangsu Province) is set up for d3 (not considering the situation of individual bird) as is shown in Table4.

Table 4. Habitats and Events Schedule (Hirundo rustica, d3)

Event *E*	Habitats	*Time*(*Data Points*)
Hirundo rustica	d_3	9, 10, 11, 12, 13, 14, 15, 16, 25, 26, 27, 28, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 60, 61, 62, 63, 64, 67, 78, 80, 81, 82, 84, 85, 86, 87, 88, 89, 90, 94, 96, 97, 98, 99, 100, 119……

Setting Min_length = 10, Max_length = 108 and min_rep = 4 Hash List is applied to discover valid periodicities for d3 in Fig. 15c. The valid periodicity is {36, 72}. Its periodic pattern is:

Per_Patterns3(Hirundo rustica)=< d3,T3,23>, T3(Hirundo rustica)=(36, 906, 29);

Per_Patterns3(Hirundo rustica)=< d3,T3,21>, T3(Hirundo rustica)=(36, 892, 9);

Per_Patterns3(Hirundo rustica)=< d3,T3,7>, T3(Hirundo rustica)=(72, 673, 169);

The conclusion is formulated after outputting the periodic pattern that:

In d3 (in Jiangsu Province) Hirundo rustica usually arrives in mid-March (occasionally in late February) and leaves in early October each year.

All of the periodic patterns are calculated for D (Hirundo rustica) and its migratory route is then generated as shown in Fig. 6.

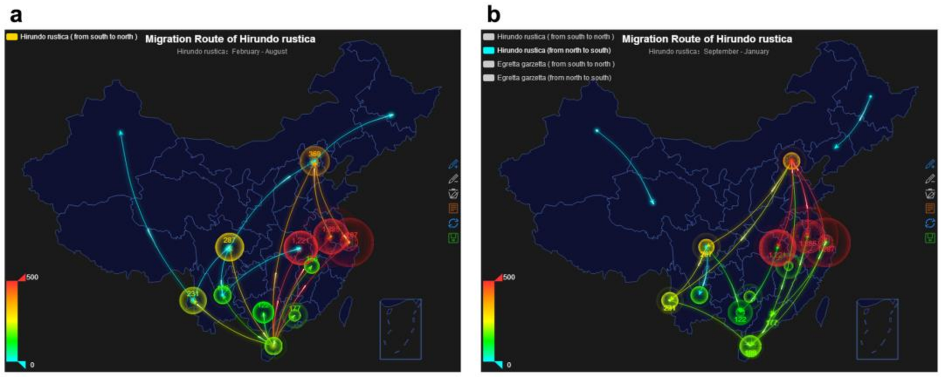


Fig. 6 Migration route (Hirundo rustica).

a Migration from South to North. b Migration from North to South.

For ej= Egretta garzetta Habitats and Events Schedule for d4 (in Shanghai) are set up in the Fig.16c (not considering the situation of individual birds) as shown in Table5.

Table 5. Habitats and Events Schedule (Egretta garzetta, d4)

Event *E*	Habitats	*Time*(*Data Points*)
Egretta garzetta	d_4	7, 8, 9, 10, 11, 12, 13, 14, 15, 24, 26, 27, 28, 44, 45, 46, 47, 48, 49, 50, 51, 61, 62, 63, 64, 67, 78, 82, 84, 85, 86, 87, 88, 89, 94,95, 96, 97, 98, 99,100,101,102, 113, 114, 115, 116 ……

Setting Min_length = 10, Max_length = 108 and min_rep = 4 Hash List is applied to discover valid periodicities for d4 in Fig. 16c. The valid periodicity is {37, 70, 107}. Its periodic pattern is:

Per_Patterns4 (Egretta garzetta)=< d4,T4,22>, T4 (Egretta garzetta)=(37, 873, 33);

Per_Patterns4 (Egretta garzetta)=< d4,T4,21>, T4 (Egretta garzetta)=(37, 812, 7);

Per_Patterns4 (Egretta garzetta)=< d4,T4,9>, T4 (Egretta garzetta)=(70, 736, 105);

Per_Patterns4 (Egretta garzetta)=< d4,T4,4>, T4 (Egretta garzetta)=(107, 665, 220);

After outputting the periodic pattern, it is concluded that:

In d4 (in Shanghai) Egretta garzetta usually arrives at the end of March and leaves in late October. There is also a situation in which they may arrive in April and leave at the end of November or early December.

All of the periodic patterns are calculated for D (Egretta garzetta) and its migratory route is then generated as shown in Fig. 7.

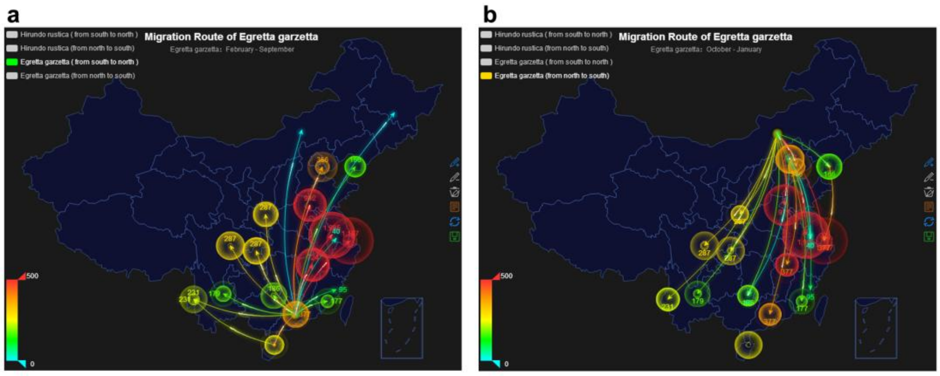


Fig.7 Migration route (Egretta garzetta) a Migration from South to North b Migration from North to South

5. Conclusion

In this paper,the method of data mining is used to analyze bird migration, and further to find its periodical patter which predicts the migration routes. Taking Hirundo rustica and Egretta garzetta as examples, maps and GIS demonstrate the feasibility of the algorithm. The migration routes and habitats of the birds derived by this work are compared with that of the authoritative ornithological literature which shows more accurate and real results and verifies the accuracy of the algorithm.

5.1 Social ang scientific value

In Experiment, it not only calculates the times at which migratory birds reach all regions but also finds the migratory routes across the country. The correspondence between time and space of bird migration can be shown visually from the results. Furthermore these isolated habitats of migratory birds are connected by time series, which makes the migration route clearer. This study is of practical significance: it is reported that various highly pathogenic avian influenza has broken out in wild birds since 2005[14]. The migrations of wild birds have spread the virus to different countries and regions, which poses serious challenges to avian influenza prevention and control [15]. The results can be used to predict the spreading route of the epidemic caused by migratory birds which can provide theoretical support for epidemic prevention and control. Further this study could assist airports to avoid flying in the peak of bird migrations in order to reduce or avoid the occurrences of bird strikes. Finally the study may provide a better guidance for bird watching activities.

5.2 Limitations and shortcomings

1) Bird-watching data cannot track individual bird. Therefore it is not enough to just rely on bird-watching data itself to verify the accuracy of the results. 2) There are spatial and temporal discontinuities in bird-watching data. Hence it is difficult to solely rely on such data to analyze changes of bird migration over the years. Moreover this may insert a negative impact on the predicted results of bird migration. 3) The effects of other factors on bird migration such as climate, the environment and human activities, have not been considered. 4) The uneven distribution of sampling still exists which prevents deeper data analysis and mining activities.

5.3 Future directions

First bird-watching data will be collected continuously. Additional migratory bird data such as satellite-tracking data and bird-banding data will be gradually introduced to supplement further mining and analyses. Second meteorological data will be added to our study. The authors will further study the effects of climate and environment on bird migration.

Acnowledgements

This work was supported by National Natural Science Foundation of China (Nos. 70771094 and 71490722), China Postdoctoral Science Foundation (No. 2016M592697), and Key Science and Technology Project of Shandong Province in China (No. 2014GGH201022).

References

[1] Bock WJ. Presidential Address: Three Centuries of International Ornithology. Acta Zoologica Sinica. 2004, 50(6): 880-912

[2] Wang Y, Zhang ZW, Zheng GM, Li JQ, Xu JL, Ma ZJ. Ornithological Research: Review of the Past Twenty Years and Future Perspectives in China. Biodiversity Science. 2012, 20: 119-137

[3] Ma ZJ. The Research Methods and Study Advances of Bird Migration. Bulletin of Biology. 2009, 44:5-9

[4] Bousquet O, Luxburg UV, Rätsch G. Advanced Lectures on Machine Learning. Springer-Verlag, 2003.

[5] Shimazaki H, Tamura M, Higuchi H. Migration Routes and Important Stopover Sites of Endangered Oriental White Storks (Ciconia boyciana) as Revealed by Satellite Tracking. Revista Enfermería Del Trabajo. 2004, 3:39-43

[6] Zhou YC, Tang MJ, Cui P. Research and Implementation of Data Mining Algorithm for Bird Migratory Behavior in Qinghai Lake. E-Science Technology and Application. 2010, 1:38-50
[7] Li XY, Liang LY, Gong P, Liu Y, Liang FF. Revealing Bird Distribution Changes of Bird Watching in China. Chinese Science Bulletin. 2012, 57: 2956-2963.
[8] Bird Report. http://www.birdreport.cn/. 2016. Accessed 23 Apr 2016.
[9] Bird Watching Network of China. http://www.chinabirdnet.org/indexc.html. 2016. Accessed 1 Apr 2016.
[10] China Ornithological Society. China Bird Report 2007 (in Chinese). Beijing: China Ornithological Society, 2008
[11] China Coastal Waterbird Census Team. China Coastal Waterbird Census Report (in Chinese). Hong Kong: Bird Watching Society, 2011
[12] Sai DJ, Sun YG. A Checklist and Distribution of the birds in Shandong. Beijing: Science Press, 2013
[13] Liu Y, Wei Q, Dong L. Updated New Bird Records in China Recently. Chinese Journal of Zoology. 2013, 48:750-758
[14] Si YL, Boer WFD and Gong P. Different Environmental Factors of Highly Pathogenic Avian Influenza H5N1 Breaking Out in the Poultry and Wild Birds. Plos One. 2013, 8.
[15] Cheng J, Wang P, Xia J, ShaYiLan GL, Lu GL. Ecological Environment and Migration Route of Wild Birds in Xinjiang and Their Effects on Occurrence and Epidemic of Avian Influenza. Progress in Veterinary Medicine. 2013, 34:108-111.