Nat. Hazards Earth Syst. Sci., 13, 3211–3220, 2013
www.nat-hazards-earth-syst-sci.net/13/3211/2013/
doi:10.5194/nhess-13-3211-2013

Natural Hazards and Earth System Sciences
Open Access

Novel method for hurricane trajectory prediction based on data mining

X. Dong and D. C. Pi

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Correspondence to: X. Dong (nuaadong_xin@163.com) and D. C. Pi (dc.pi@nuaa.edu.cn)

Received: 26 March 2013 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: 11 September 2013
Revised: – Accepted: 11 November 2013 – Published: 10 December 2013

Abstract. This paper describes a novel method for hurricane trajectory prediction based on data mining (HTPDM) according to the hurricane's motion characteristics. Firstly, all frequent trajectories in the historical hurricane trajectory database are mined by using association analysis technology and their corresponding association rules are generated as motion patterns. Then, the current hurricane trajectories are matched with the motion patterns for predicting. If no association rule is found for matching, a predicted result according to the hurricane current movement trend would be returned. All experiments are conducted with the Atlantic weather Hurricane/Tropical Data from 1900 to 2008. The experimental results show that if the matching failure part is contained, the prediction accuracy is 57.5 %. Whereas, the valve would be to 65 % provided all matches are successful.

1 Introduction

With the rapid development of Word Wide Web (WWW) and wireless communication technologies, mobile communication and mobile computing technologies have been widely used in various fields. Mobile communication equipment, animal migrations, traffic and transportation and clouds cluster tracking are all moving object instances in specific application areas (Morzy, 2007). Some correlative technologies, such as sensor networks, global positioning systems and satellite data services collect and provide a large amount of behavioural data of moving objects, which have brought huge challenges to analyse their inherent regularities. The mobile path prediction has become a hot topic in many research areas.

Hurricanes are tropical cyclones with sustained winds of at least 64 kt (119 km h^{-1}, 74 mph). On an average, more than 5 tropical cyclones become hurricanes in the United States each year causing great human and economic losses (Su et al., 2010). In respect to this fact, the trajectory prediction as the most important measure to reduce losses has become a hot issue in the field of mobile path prediction.

In this paper, we emphasise on the study of a hurricane trajectory prediction method based on data mining. The prediction method we propose gives up the complex modelling process in the traditional objective forecast method. Instead, it identifies the effective motion patterns in the historical trajectory database by using association analysis technology, and then predicts their future trajectories with pattern matching. The overall framework of the hurricane trajectory prediction method is shown as Fig. 1. After data pre-processing, all frequent trajectories from 1900 to 2000 in the historical hurricane trajectory database are mined according to the given minimum support and then generate all corresponding association rules as motion patterns. Secondly, the current hurricane trajectories from 2001 to 2008 are matched with the motion patterns for predicting. If no association rule is returned, the one according to the hurricane current movement trend would be returned. At last, the correctness of this method would be verified.

Instructions

Data preprocessing ①: the data in this stage is for frequent trajectory mining.

Data preprocessing ②: the data in this stage is for pattern matching and correctness verifying. The details are seen in Sect. 7.1.

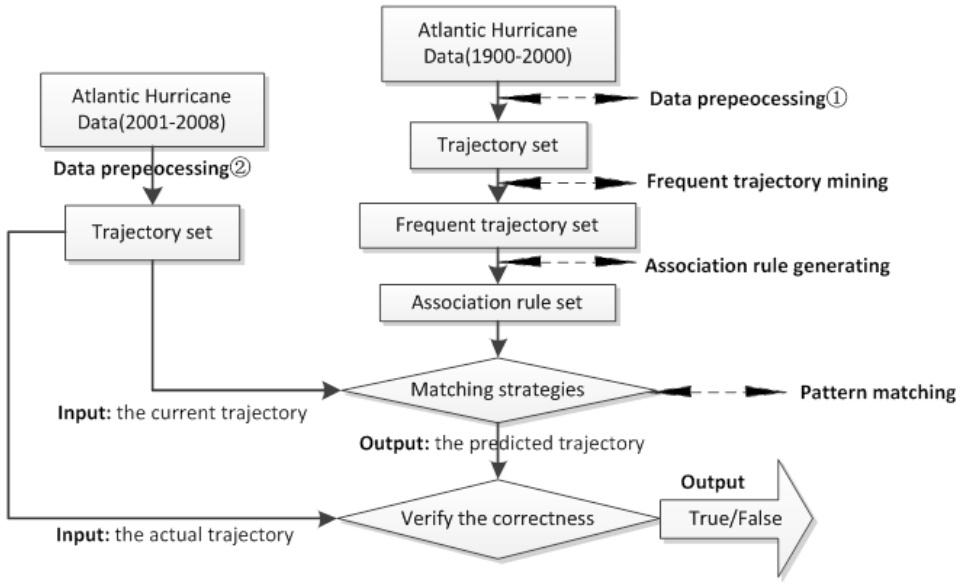

Fig. 1. Prediction system framework.

2 Related work

According to different behavioural characteristics of moving objects, many new ideas for mobile path prediction with data mining are constantly emerging.

The one based on association analysis is concerned by more and more researchers. Zhao et al. (2008) proposed a mobile device location prediction algorithm MPP based on the AprioriAll algorithm, which avoids the state space expansion problem in K order Markov predictor. Long et al. (2009) proposed an effective and simple trajectory prediction algorithm (E^3TP), which mines frequent path model based on FP-Growth algorithm, and introduces speed, one of the most important characteristics into it. Otherwise, FP-Growth requires large storage space. So the effect is not very ideal in the case of a large amount of data. Moreover, E^3TP is appropriate for unordered sequences, not for ordered time series. Kim et al. (2007) proposed a novel method for predicting a future path of an object in an efficient way by analysing past trajectories whose changing pattern is similar to that of a current trajectory of a query object. However, this method is only adapt for the road network, and has many limited conditions. Morzy (2007) mines the database of moving object locations to discover frequent trajectories and movement rules, and then matches the moving object trajectory with the movement rule database to establish a probabilistic model of object location.

Hurricane is a kind of strong and deep tropical cyclone generated in the Eastern Atlantic and the North Pacific region, also known as typhoon, cyclone. Hurricane movement is regarded as disengaged motion, which is generally accompanied by strong wind and heavy rain. For a long time, the formation of hurricane has been concerned by many researchers in various disciplines (Rozanova, 2004). Much progress has been made over the last decade in the understanding of physical processes and the quality of operational prediction of hurricane (Chan, 2005; Weber, 2005). The present prediction methods mainly include the numerical prediction, the objective forecast and the comprehensive forecast, where the objective forecast method based on statistical dynamics, due to its higher precision, has been more and more popular in recent years. Currently, the objective forecast method is commonly used in the hurricane trajectory prediction:

1. Persistence and Climatology Method (i.e. PC method). It is one of the most widely used forecast methods, which has the advantages of simple calculation and high prediction precision.

2. Integrated Forecast Method. It is a comprehensive forecast method, which is integrated by many techniques' calculations, such as stepwise regression, multiple regression, stepwise multilevel discriminant, analysis of variance, fitting error analysis, autoregressive, etc. This method has a so short running time and can forecast many items.

3. Probability Forecast Method. It includes some techniques, such as REEP, discrete likelihood estimation, uncertain consumption, Bayesian, Markov chain experience statistics, etc.

However, the objective forecast method is so complex, because it takes many factors into account, such as phase transitions, vertical advection, and boundary layer effects, etc. The

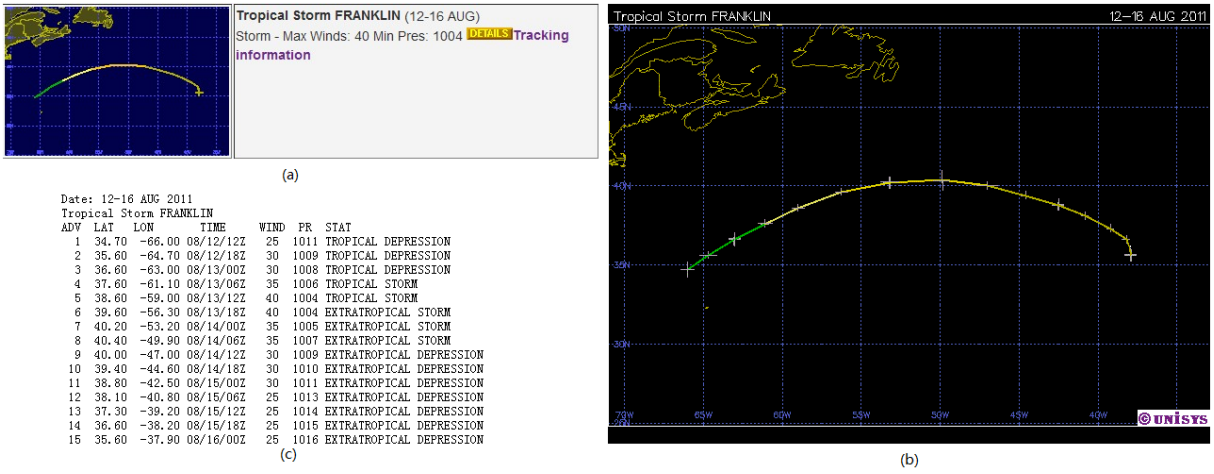

Fig. 2. A certain hurricane. **(a)** the summary information; **(b)** the schematic diagram of its trajectory; **(c)** the trajectory information.

assumptions for the understanding of the underlying physical process, to simplify the complexity of the original system, often ignore some important properties (Rozanova et al., 2010). For example, due to using multiple regression analysis, PC method should select predictors from many factors according to the F value to construct prediction model. However, these selected predictors inevitably affect the prediction model itself.

Therefore, some researchers begin to use data mining technology to predict hurricanes trajectories. Zou et al. (2008) established a model, which is developed based on the similarities of key points on typhoon tracks to forecast typhoon tracks using historical data. The centres of active typhoons are compared with historical records of similar typhoons. The typhoon tracks are weighted based on their similarity. The centre positions are then quickly forecasted based the similarity weights. Kim et al. (2011) introduced the feasibility of a straightforward metric to incorporate the entire shapes of all tracks into the fuzzy c-means clustering method. This method is suitable for the data where cluster boundaries are ambiguous. Kim et al. (2012) proposed a seasonal tropical cyclone forecast model based on the tracking mode. This model combines fuzzy C-means clustering method with statistical dynamics. In addition, some concepts, such as grey theory, neural network, etc. are also introduced into the hurricane weather forecast.

Others are committed to examine the problem of forecasting the intensification of hurricanes using data mining techniques, and have got some achievements. Su et al. (2010) proposed a new hurricane intensity prediction model, WFL-EMM, which is based on the data mining techniques of feature weight learning (WFL) and Extensible Markov Model (EMM). Chatzidimitriou et al. (2005) formulated tropical storm intensification prediction as a supervised data mining problem; the objective being to produce accurate early warnings with respect to changes in wind speed of a particular storm. They examined two alternative approaches to discover classification rules on current hurricane data: particle swarm optimisation and class association rules.

3 Region discretization

The raw hurricane trajectory data is the Atlantic weather Hurricane/Tropical Data (1900–2008), got from the website http://weather.unisys.com/hurricane/. The data from 1900 to 2000 is used to mine the motion patterns, and the rest is for predicting and verifying.

Figure 2 shows the raw information of a certain hurricane. Unfortunately, the domain of position coordinates is continuous and the granularity level of raw data is very low. Therefore, any pattern discovered from raw data cannot be generalised. To overcome this problem we choose to transform original paths of moving objects into trajectories expressed on a coarser level (Morzy, 2007).

The moving region can be divided into many square areas with same size. Each trajectory is converted to an area sequence.

For example, Fig. 3 shows an original trajectory of a moving object, expressed as $\{(-66, 34.7), (-64.7, 35.6), \ldots, (-38.2, 36.6), (-37.9, 35.6)\}$, as shown on the right side in Fig. 3. After region division, as in Fig. 4, the original moving region is divided by dotted line into several sub-regions, the trajectory above can be expressed as $\{(-13, 6), (-12, 7), (-11, 7), (-10, 8), (-9, 8), (-8, 7), (-7, 7)\}$. If the trajectory t_1 is the sub-trajectory (continuous sequence) of the trajectory t_2, then t_2 contains t_1, denoted $t_1 \subseteq t_2$.

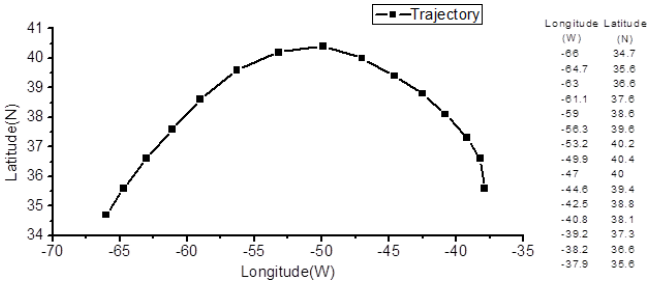

Fig. 3. An original path of a moving object.

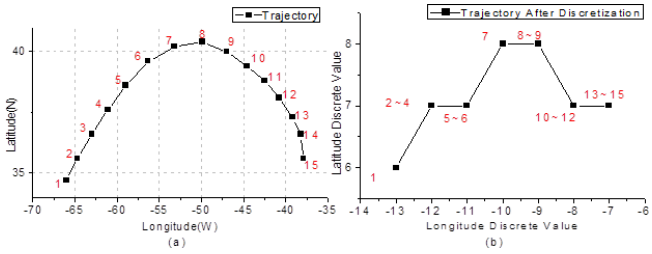

Fig. 4. The trajectory after discretization.

4 Frequent trajectory mining

The pseudo code for the HTPDM algorithm is given in the annex for a transaction database TD, a support threshold of minsup and a confidence threshold of minconf. Frequent trajectory mining as the first step is to mine all frequent trajectories based on the Apriori algorithm, according to the user-defined minimum support threshold.

The Apriori algorithm (Agrawal and Srikant, 1994) is a classical algorithm for association rules mining. The name of the algorithm comes after a prior knowledge about frequent itemsets was used. The prior knowledge is that any non-empty subset of a frequent itemset is also frequent. Apriori algorithm uses a level-wised and iterative approach, it first generates the candidates then test them to delete the non-frequent item sets. Most of previous studies adopted an Apriori-like candidates generation-and-test approach. The hurricane trajectory is different from the traditional item sets, but a time series that change with time (Qin and Shi, 2006). The frequent trajectory mining problem from single series can be viewed as the mining problem of sequential patterns. Before introducing the algorithm, we first give some related concepts.

Trajectory length. The length of the trajectory t_1 is the number of elements <lat, lon> in the trajectory sequence, denoted $|t_1|$ or length (t_1). We refer to a trajectory of length x as *x sequence*. When x equals one, the trajectory is called unit trajectory, i.e. *1 sequence*.

Adjacent unit trajectories. Let $t_1 = \{P_1\}, t_2 = \{P_2\}$ be unit trajectories. If the square regions they represent share an edge, or at least one of the sequences $\{P_1, P_2\}$ and $\{P_2, P_1\}$ is the sub-trajectory of a certain trajectory, t_1 and t_2 would be said to be adjacent. t_1 and t_2 can be merged into a 2-sequence trajectory, denoted $t_1 \wedge t_2 = \{P_1, P_2\}$ or $t_2 \wedge t_1 = \{P_2, P_1\}$.

Trajectory connection. Let $t_1 = \left\{P_1^1, P_2^1, P_3^1, \ldots P_k^1\right\}, t_2 = \left\{P_1^2, P_2^2, P_3^2, \ldots P_k^2\right\}$ be two non-unit trajectories with same length. If the later $k-1$ items of t_1 and the former $k-1$ items of t_2 are identical, i.e. $\left\{P_2^1, P_3^1, \ldots P_k^1\right\} = \left\{P_1^2, P_2^2, \ldots P_{k-1}^2\right\}$, t_1 and t_2 can be connected. The connection of t_1 and t_2 is denoted $t_1||t_2$, i.e. $t_1||t_2 = \left\{P_1^1, P_2^1, P_3^1, \ldots P_k^1, P_k^2\right\}$.

Support of trajectory. Given a database of trajectories $\mathrm{TD} = \{T_1, T_2, \ldots, T_n\}$. The support of a trajectory t_i is the percentage of trajectories in TD that support the trajectory t_i.

$$\text{support}\ (t_i) = \frac{\sum_{k=1}^{n} \{\text{count}(t_i \subseteq T_k)\}}{|\mathrm{TD}|}, \quad (1)$$

where $|\mathrm{TD}|$ is the trajectory number in TD, and count $(t_i \subseteq T_k)$ expresses the number of T_k *containing* t_i. For example, if $T_1 = \{a, b, c, a, b\}, t_1 = \{a, b\}$, then count $(t_1 \subseteq T_1) = 2$.

Frequent trajectory set.

A trajectory t is frequent if its support exceeds user-defined threshold of minimum support (a real numbers between 0 and 1), denoted minsup. The set of all k frequent trajectories is denoted F_k. The collection of all frequent trajectories is called frequent trajectory set, denoted *FreTraSet*.

Priori principle. If a trajectory is frequent, then all its sub-trajectories must also be frequent. Conversely, if a trajectory is a non-frequent, then all its sub-trajectories also must be non-frequent.

Frequent trajectory mining is the first step of the HTPDM algorithm we proposed. Firstly, it scans the database TD for the first time for calculating all supports of the unit trajectories, and selects 1-frequent trajectory set F_1 through the comparison with minsup. Then it generates candidate frequent trajectory sets of length k C_k from frequent trajectory sets of length $k-1$ F_{k-1}, and prunes the candidates which have an infrequent sub pattern. After that, it scans the database TD to determine frequent trajectory set F_k among the candidates.

5 Association rule generating

After frequent trajectory mining, the corresponding association rules can be generated according to the user-defined minimum confidence threshold, as the motion patterns stored in the database. Frequent trajectories are transformed into movement rules. A movement rule is an expression of the form $h \Rightarrow t - h$, where t is a frequent trajectory, and h is the rule's premise; $t - h$ is the rule's conclusion. With respect to the movement rules, there are some concepts to introduce.

Support Of Movement Rule. The support of $h \Rightarrow t-h$ is defined as the support of trajectory t.

$$\text{support}(h \Rightarrow t-h) = \frac{\sum_{k=1}^{n} \{\text{count}(t \subseteq T_k)\}}{|\text{TD}|}. \quad (2)$$

Confidence Of Movement Rule. The confidence of $h \Rightarrow t-h$ is the conditional probability of $t-h$ given h.

$$\text{confidence}(h \Rightarrow t-h) = P(t-h|h) = \frac{\text{support}(t)}{\text{support}(h)}. \quad (3)$$

Association rule generating is the second step of the HTPDM algorithm we proposed. The objective is to generate all corresponding association rules, which confidence exceeds user-defined threshold of minimum confidence.

A k-frequent trajectory can generate $k-1$ association rules $(h_x \rightarrow t-h_x)$. For example, let a trajectory t be $\{a,b,c,d\}$, it can be decomposed into three rules: $\{a\} \rightarrow \{b,c,d\}, \{a,b\} \rightarrow \{c,d\}, \{a,b,c\} \rightarrow \{d\}$. Take each rule $h_x \rightarrow t-h_x$ into account, if its conf $(=\sup(t)/\sup(h_x))$ is greater than minconf, this rule would be stored in the database.

6 Pattern matching for predicting

This part is the last step of the HTPDM algorithm. It firstly matches the hurricane trajectories, which are used to predict after pre-processing, with each rule generated in the previous step and chooses the optimal one according to the evaluation function. If any association rule can not be found, a predicted point depend on the movement trend would be returned. Then the hurricane actual future trajectory and the predicted one by pattern matching would be compared to determine whether the predicted result is correct by computing the centre points' distance. Some related concepts are as follows.

Matching Length. Let $t_1 = \{P_1^1, P_2^1, P_3^1, \ldots P_n^1\}, t_2 = \{P_1^2, P_2^2, P_3^2, \ldots P_k^2\}$ be two trajectories. We say that the matching length of t_1 and t_2 is c, if there exists a positive integer c, so that $P_n^1 = P_k^2 \wedge P_{n-1}^1 = P_{k-1}^2 \wedge \ldots \wedge P_{n-c+1}^1 = P_{k-c+1}^2 (c < n, c < k)$. Otherwise, the matching length is 0.

Evaluation Function. The evaluation function's value reflects the matching degree of current trajectory with motion patterns quantitatively. The higher the evaluation function's value is, the higher the matching degree would be. The impact factors are the rule's confidence and the matching length. Evaluation function is defined as follows.

$$f = \text{conf} \times (1 - e^{-l_{\text{match}}}), \quad (4)$$

where conf is the rule's confidence, and l_{match} is the matching length of the current trajectory with the rule's premise.

According to Eq. (4), we can calculate its value in different cases. Seen in Table 1, we find that the influence of l_{match} is more and more obvious as the increase of the confidence. Under the condition of the same evaluation function value, i.e. two different rules' conf and l_{match} are the same, we would select the one, which conclusion's length is longer, for pattern matching.

Table 1. Evaluation function's value.

conf	l_{match}			
	1	2	3	4
	$(1-e^{-l_{\text{match}}})$			
	0.63212	0.86466	0.95021	0.98168
0.3	0.18964	0.25940	0.28506	0.29451
0.4	0.25285	0.34587	0.38008	0.39267
0.5	0.31606	0.43233	0.47511	0.49084
0.6	0.37927	0.51880	0.57013	0.58901
0.7	0.44248	0.60527	0.66515	0.68718
0.8	0.50570	0.69173	0.76017	0.78535
0.9	0.56891	0.77820	0.85519	0.88352
1	0.63212	0.86466	0.95021	0.98168

If no association rule is found for matching, a predicted point depend on the movement trend would be returned. For example, let $t = \{P_1, \ldots\ldots, P_{k-1}, P_k\}\{(k > 2)$ be a current trajectory, where $P_{k-1} =< \text{lat}_{k-1}, \text{lon}_{k-1} >, P_k =< \text{lat}_k, \text{lon}_k >$, then the predicted trajectory $t_p = \{2 \times \text{lat}_k - \text{lat}_{k-1}, 2 \times \text{lon}_k - \text{lon}_{k-1}\}$.

Standard For Correct Prediction Select the minimum m, according to the actual future trajectory's length and the predicted trajectory's length with pattern matching. Cut out the former m items of the two trajectories, and get their centre points. If the distance of the two points is not greater than 1, we would say that the prediction is correct.

7 Experiment

All experiments were conducted on a PC equipped with Pentium T2390 CPU, 1G RAM, and a SATA hard drive running under Windows XP SP2 Home Edition. Algorithms and the front-end application were implemented in C# and run within Microsoft .NET 2.0 platform. The experimental data is stored in Microsoft SQL Server 2000.

1. Data preprocessing ①:

   The data in this stage is from 1900 to 2000.

   a. Set the region size to 5×5, making the most of the processed trajectory points with single coordinate value within 1 unit jump;

   b. Process the raw trajectories as explained in Sect. 3. A complete trajectory's storage type is string, such as $\{1,2;2,3;3,4\ldots\}$;

Table 2. The datum in raw_data table.

Field	year	id	num	name	traj	*t*_length	flag
Data	…	…	……	……	……	……	……
	1961	7	498	FRANCES	3, 11; 3, 12; 3, 13; 4, 13; 4, 14; 5, 14; 6, 13; 7, 12; 8, 13; 9, 12; 9, 11	11	0
	1961	8	499	GERDA	3, 15; 4, 15; 5, 15; 5, 14; 6, 14; 7, 13; 8, 13; 8, 12; 8, 11; 8, 10; 8, 9	11	0
	1961	9	500	HATTIE	2, 16; 3, 16; 3, 17; 3, 18	4	0
	1961	10	501	JENNY	3, 12; 4, 12; 4, 11; 5, 11; 5, 10; 5, 9; 5, 8; 5, 9; 5, 10; 6, 10; 6, 9	11	0
	1961	11	502	INGA	4, 18; 4, 19; 4, 18; 3, 18	4	0
	1962	1	503	ALMA	5, 15; 5, 16; 6, 15; 7, 15; 7, 14; 8, 14; 8, 13; 8, 12; 7, 12; 7, 13; 8, 12; 9, 11	12	0
	…	…	……	……	……	……	……

Table 3. The datum in experiment_data table.

Field	id	former_traj	next_traj
Data	…	……	……
	121	3, 9; 4, 9; 4, 10; 5, 10; 5, 11; 6, 11; 6, 10; 7, 10; 7, 9; 7, 8; 8, 8; 8, 7; 9, 7	10, 6; 11, 6; 11, 5; 12, 4; 12, 3; 12, 2; 12, 0
	122	3, 9; 4, 9; 4, 10; 5, 10; 5, 11; 6, 11; 6, 10; 7, 10; 7, 9; 7, 8; 8, 8; 8, 7; 9, 7; 10, 6; 11, 6	11, 5; 12, 4; 12, 3; 12, 2; 12, 0
	123	5, 13; 6, 13; 6, 12; 6, 11; 6, 10; 6, 9; 6, 8; 7, 8	7, 7; 8, 6; 8, 5; 9, 5
	124	5, 13; 6, 13; 6, 12; 6, 11; 6, 10; 6, 9; 6, 8; 7, 8; 7, 7	8, 6; 8, 5; 9, 5
	125	5, 15; 6, 15; 6, 14; 7, 14; 8, 13; 8, 12; 9, 11; 9, 10; 9, 9; 10, 8; 10, 7; 10, 6; 10, 5	10, 4; 11, 3; 11, 2; 12, 1; 12, 0; 13, 0; 13, 1
	126	5, 15; 6, 15; 6, 14; 7, 14; 8, 13; 8, 12; 9, 11; 9, 10; 9, 9; 10, 8; 10, 7; 10, 6; 10, 5; 10, 4; 11, 3	11, 2; 12, 1; 12, 0; 13, 0; 13, 1
	…	……	……

c. Deposit each hurricane's information (including *name, ID number, serial number, year, trajectory, trajectory length, flag*) in the *raw_data* table. Preprocessing result is shown in Table. 2.

2. Data preprocessing ②:

   The data in this stage is from 2001 to 2008.

   a. Preprocess the trajectories as mentioned above;

   b. Cut each trajectory into the head and the tail section. The head section is used for pattern matching and choosing the optimum association rule. The tail section is used for correctness verification;

   c. Deposit each trajectory's information into the *experiment_data* table. Preprocessing result is shown in Table 3.

Depend on the HTPDM algorithm, the results of frequent trajectory mining and association rules generating are shown in Fig. 5. The predicted results are shown in Fig. 6.

Take for example the record with the id of 126 in Table 3. The current trajectory for predicting is $\{5, 15; 6, 15; 6, 14; \ldots\ldots; 10, 4; 11, 3\}$, and the future trajectory for verifying is expressed as $\{11, 2; 12, 1; 12, 0; 13, 0; 13, 1\}$. We get its predicted trajectory by HTPDM algorithm, and the result is $\{11, 2; 12, 2; 12, 1; 13, 1\}$. According to the standard for correct prediction mentioned in Sect. 6, the two centre points are (12,0.75) and (12,1.5). Their distance is 0.75, lower than 1, so we think that the predicted result is true. Details are seen in Fig. 7.

8 Analysis

Throughout the process of prediction, minsup (minimum support) and minconf (minimum confidence) are the most important parameters for impacting prediction accuracy. minsup directly control the frequent trajectory mining, and minconf indirectly control the generation of association rules under the premise of frequent trajectories.

8.1 Minconf and the prediction accuracy

Figure 8 presents the prediction accuracy with respect to the varying value of the minconf threshold for a set value of minsup = 0.003. The calculation methods for the prediction accuracy are shown as follows:

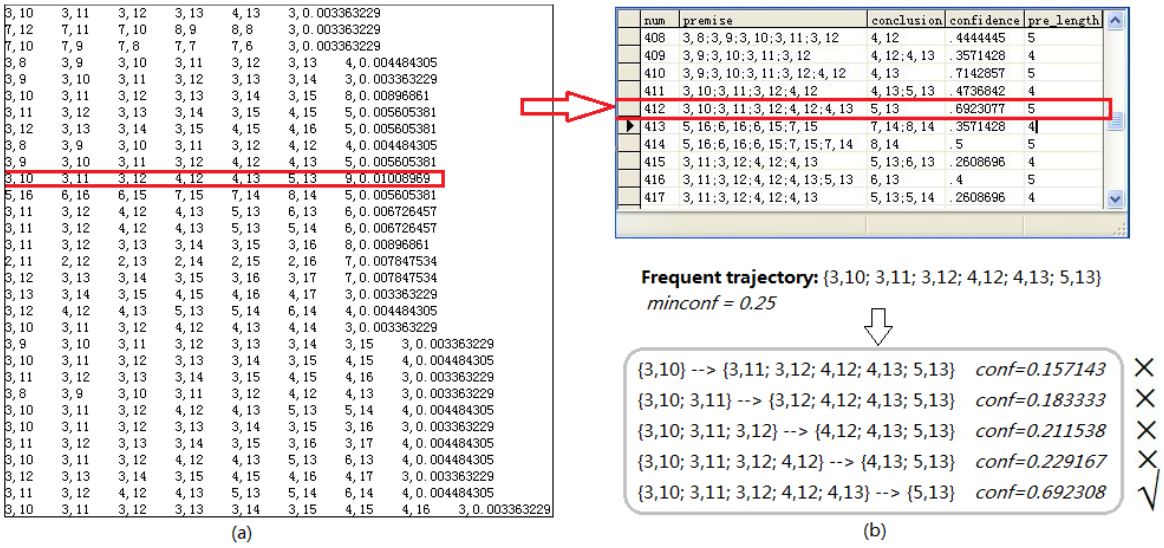

Fig. 5. Mining motion patterns. **(a)** frequent trajectory mining; **(b)** association rules generating.

```
The current trajectory: 3,13;2,13;3,13;3,12;4,12;4,11;5,11
The actual future trajectory:6,10;6,9;7,9;7,8
The predicted result: 6,11
The rule's number for pattern matching: 57, matching length: 1
TRUE!

The current trajectory: 3,13;2,13;3,13;3,12;4,12;4,11;5,11;6,10
The actual future trajectory:6,9;7,9;7,8
Can not find any association rule, a predicted point depend on the movement tren
d would be returned: The predicted result: 7,9
TRUE!

The current trajectory: 2,16;3,16;3,15;4,15
The actual future trajectory:4,16;5,17
The predicted result: 4,16
The rule's number for pattern matching: 449, matching length: 2
TRUE!

The total number: 214, the correct number: 123, the error number: 91
```

Fig. 6. Predicted results (Console output).

1.

$$\text{correct\_rate1} = \frac{T\text{num(PatternMatching\_True)}}{T\text{num(PatternMatching\_All)}}$$

Tnum(PatternMatching_True): the number of the correct predicted results by pattern matching; Tnum(PatternMatching_ALL): the number of the total predicted results by pattern matching.

2.

$$\text{correct\_rate2} = \frac{T\text{num(PatternMatching\_True)}}{T\text{num(All)}}$$

Tnum(All): the total trajectory number.

The first correct rate reflects the accuracy of data mining technology for itself. With the increase of minconf, remove the last abnormal point, the correct rate overall presents a rising trend, and always maintain above 60 %. Because minconf is larger, the credibility of the generated association rules is higher; the error rate would be smaller. When minconf continues to increase, because the number of the generated association rules is less, the accuracy would present an unstable state, and would appear abnormal points as shown in Fig. 8.

The second correct rate reflects the efficiency of data mining technology for the prediction system. With the increase of minconf, the correct rate is stable at the beginning and remains above 45 %. When minconf continue to increase from 0.4, however, the correct rate decreases rapidly to about 10 %. This is because, for the whole prediction system, if the system can not find the matching pattern, it would return a result according to the movement trend, and the accuracy of this speculation is fairly low.

8.2 Minsup and the prediction accuracy

Figure 8 shows that when minconf = 0.25, the prediction accuracy is the best. Therefore, we let minconf value to be 0.25.

Figure 9 shows the prediction accuracy with respect to the varying value of the minsup threshold for a set value of minconf = 0.25. With the increase of minsup threshold, two correct rates fluctuate by small degrees partly, but decrease generally. Under the condition of the same minconf value, the increase of minsup leads to the decrease of the frequent trajectory number, as well as the number of association rules. All these reasons result in the decrease of prediction accuracy.

Through repeated experiments, we find that when minsup set to 0.003 and minconf set to 0.25, prediction accuracy is the best. The experimental results show that in the 214 trajectories, the number of the trajectories, which match the motion patterns successfully, is 160, including 104 correct predicted

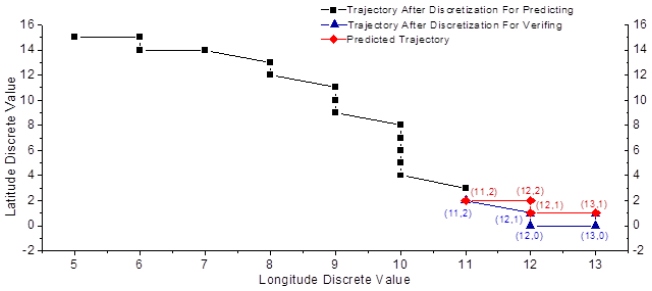

Fig. 7. The predicted result of the trajectory with the id number of 126. The black part is the current trajectory for predicting; the blue part is the actual future trajectory for verifying; The red part is the predicted trajectory by data mining.

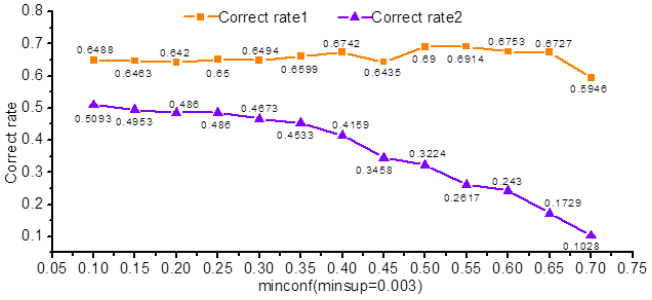

Fig. 8. Minconf and the correct rate when minsup = 0.003.

trajectories and 56 incorrect predicted trajectories. The accuracy is 65 %. The other 54 trajectories fail to match the patterns, including 19 correct predicted trajectories and 35 incorrect predicted trajectories. The accuracy is 35.2 %. In the whole 214 trajectories, the number of the correct predicted trajectories is 123 and the number of the incorrect ones is 91, the accuracy of the whole prediction system is 57.5 %.

9 Conclusions

This paper proposes a novel method for hurricane trajectory prediction based on data mining by integrating association analysis technology and using the real American Atlantic hurricane data. This method is different from the traditional dynamics modelling forecast affected by multiple factors. Firstly, all frequent trajectories in the historical hurricane trajectory database are mined by using association analysis technology and their corresponding association rules are generated as motion patterns. Then, the current hurricane trajectories are matched with the motion patterns for predicting. If no association rule is found for matching, a predicted result according to the hurricane current movement trend would be returned. The experiments show that the prediction accuracy is ideal and satisfactory.

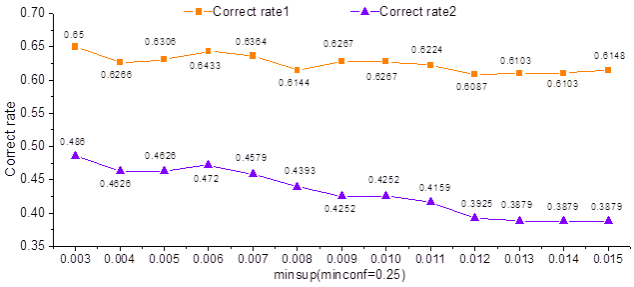

Fig. 9. Minsup and the correct rate when minconf = 0.25.

Our future work includes:

1. Replacing uniform moving regions with differently sized areas that divide the movement area based on the density and congestion of moving objects;
2. Including spatial information into the movement rules;
3. Developing more effective matching strategies and finding more useful evaluation functions.

Appendix A

The pseudo-code of the HTPDM algorithm

Algorithm Hurricane Trajectory Prediction Based On Data Mining

Input: (1) Moving trajectory database *TD*, minimum support threshold *minsup*, minimum confidence threshold *minconf*
Output: (1) Predicted trajectory t_p
(2) True/False
Algorithm:
//Step1: Mining frequent trajectory set FreTraSet
01: Scan the table T_1 containing the all history trajectories;
02: $C_1 =$ { all different trajectory points };
03: $F_1 = \{ c \in C_1 | c.sup \geq minsup \}$;
04: $FreTraSet = F_1$;
05: $max\_length =$ the maximum trajectory length in T_1;
06: FOR ($k = 2; k \leq max\_lentgh; k++$)
07: $C_k =$ get_candidate (F_{k-1});
08: IF (C_k is null)
09: Break;
10: FOR all trapsactions $t \in T_1$;
11: $flag = 1$;
12: IF ($t.flag == 0$)
13: FOR all trajectories $c \in C_k$;
14: IF ($c \in t$)
15: $flag = 0$;
16: Record the number of the occurrences of c, and it to the $c.count$;

17: END IF;
18: END FOR;
19: END IF;
20: IF (*flag* == 1)
21: Modify this trajectory's *t.flag* to 1;
22: END IF;
23: END FOR;
24: FOR all trajectories $c \in C_k$;
25: $c.sup = c.count / |D|$;
26: END FOR;
27: $F_k = \{c \in C_k \mid c.sup \geq minsup\}$;
28: IF (F_k is null)
29: Break;
30: $FreTraSet = FreTraSet \cup F_k$;
31: END FOR;
//Step2: Generating the table T_3 containing all association rules
32: FOR ($k = 2$; *k<FreTraSet.count* + 1; i++)
33: FOR all trajectories $t \in F_k$;
34: $h_1 =$ gettail (t);
35: CALL rules_generation (t, h_1);
36: END FOR;
37: END FOR;
//Step3: Predicting and verifying the results
38: FOR each trajectory t for predicting in table T_2 containing all experimental data
39: $f = 0$;
40: Id_r = -1;
41: FOR all rules $r \in T_3$;
42: Calculate the evaluation function's value f_r;
43: IF ($f_r > f$)
44: $f = f_r$;
45: $Id_r =$ r.num;
46: Record the conclusion's length of the new rule;
47: $t_p = r.conclusion$;
48: ELSE IF ($f_r = f$)
49: Id_r = the ID number of the one which has a longer conclusion;
50: Record the conclusion's length of the new rule;
51: $t_p = r.conclusion$;
52: END IF;
53: END FOR;
54: IF ($Id_r == -1$)
55: t_p = a predicted point depend on the movement trend;
56: END IF;
57: $m =$ min (t_p*_length*, t_n*_length*) >5? 5: min (t_p*_length*, t_n*_length*);
58: Capture their former m items respectively, and calculate their centre points: p_{pm}, p_{nm};
59: IF ($p_{pm} - p_{nm} > 1$)
60: RETURN (t_p, *FALSE*);
61: ELSE RETURN (t_p, *TURE*);
62: END IF;
63: END FOR;

Alg. 1 Algorithm HTPDM

A1 Instructions

1. Line 07. The function *get_candidate(F_k)*. If the elements in F_{k-1} are all unit trajectories, all adjacent ones would be merged into 2-sequence trajectories according to the concept of "Adjacent Unit Trajectories" in Sect. 4. Otherwise, all trajectories in F_{k-1}, which can be connected, would be connected to k sequence trajectories according to the concept of "Trajectory Connection" in Sect. 4.

2. Line 10–23. Every time the tableT_1 is scanned, the transaction's flag, which does not contain any trajectory in C_k, would be set to 1. The transactions in T_1, which flag is 1, would not be considered in the next scanning. It would reduce the complexity of the algorithm. The theory basis is that the one does not contain any trajectory in C_k, would not likely to contain any frequent trajectory with length greater than k (Priori Principle in Sect. 4).

3. Line 32–37. A k frequent trajectory can generate k-1 association rules ($h_x \rightarrow t - h_x$). For example, let a trajectory t be $\{a, b, c, d\}$, it can be decomposed into three rules: $\{a\} \rightarrow \{b, c, d\}, \{a, b\} \rightarrow \{c, d\}, \{a, b, c\} \rightarrow \{d\}$. Take each rule $h_x \rightarrow t - h_x$ into account, if its conf (=sup(t)/sup(h_x)) is greater than minconf, this rule would be stored in the table T_3. *rules_generation(t,h_x)* is a recursive procedure, the algorithm is described as follow.

01: Generate a rule, the tail (the premise h_x), the head (the conclusion $t - h_x$);
02: *rule.conf = sup (t)/sup (h_x)*;
03: IF (*rule.conf* $\geq$ *minconf*);
04: Put it into the table T_3;
05: END IF;
06: h_{x+1} = getnexttail (t, h_x);
07: CALL rules_generation (t, h_{x+1});

Supplementary material related to this article is available online at http://www.nat-hazards-earth-syst-sci.net/13/3211/2013/nhess-13-3211-2013-supplement.pdf.

Acknowledgements. This paper is supported by the Fundamental Research Funds for the Central Universities (NZ2013306), Qing Lan Project, the 333 project of Jiangsu Province, the Technology Foundation of China (JSJC2013605C009).

Edited by: R. Lasaponara
Reviewed by: three anonymous referees

References

Agrawal, R. and Srikant, R.: Fast algorithms for mining association rules, in: Proc. of the 20th Int'l Conf on Very Large DataBases (VLDB'94), edited by: Bocca, J., M. and Zaniolo, C., Santiago, Morgan Kaufmann, 487–499, 1994.

Chan, J. C. L.: The physics of tropical cyclone motion, Ann. Rev. Fluid Mech., 37, 99–128, 2005.

Chatzidimitriou, K. and Sutton, A.: Alternative Data Mining Techniques for Predicting TropicalCyclone Intensification, American Association for Artificial Intelligence, 200, edited by: Chan, J. C. L., The physics of tropical cyclone motion, Ann. Rev. Fluid Mech., 37, 99–128, 2005.

Kim, H.-S., Kim, J.-H., Ho, C.-H., and Chu, P.-S.: Pattern Classification of Typhoon Tracks Using the Fuzzy c-Means Clustering Method, J. Climate, 24, 488–508, 2011.

Kim, H.-S., Ho, C.-H., Kim, J.-H., and Chu, P.-S.: Track-Pattern-Based Model for Seasonal Prediction of Tropical Cyclone Activity in the Western North Pacific, J. Climate, 25, 4660–4678, 2012.

Kim, S.-W., Won, J.-I., Kim, J.-D., Shin, M., Lee, J., and Kim, H.: Path prediction of moving objects on road networks through analyzing past trajectories, in: KES'07/WIRN'07 Proceedings of the 11th international conference, KES 2007 and XVII Italian workshop on neural networks conference on Knowledge-based intelligent information and engineering systems: Part I, edited by: Apolloni, B., Heidelberg, Springer-Verlag Berlin, 379–389, 2007.

Long, T., Qiao, S., Tang, C., Liu, L., Li, T., and Wu, J.: E3TP: A novel trajectory prediction algorithm in moving objects databases[A], in: PAISI '09 Proceedings of the Pacific Asia Workshop on Intelligence and Security Informatics, edited by: Chen, H., Heidelberg, Springer-Verlag Berlin, 76–88, 2009.

Morzy, M.: Mining frequent trajectories of moving objects for location prediction, in: MLDM '07 Proceedings of the 5th international conference on Machine Learning and Data Mining in Pattern Recognition, edited by: Perner, P., Heidelberg, Springer-Verlag Berlin, 667–680, 2007.

Qin, L. X. and Shi, Z. Z.: Efficiently mining association rules from time series, Int. J. Inf. Technol., 12, 30–38, 2006.

Rozanova, O. S.: Note on the typhoon eye trajectory, Regular and Chaotic Dynamics, 9, 129–142, 2004.

Rozanova, O. S., Yu, J.-L., and Hu, C.-K.: Typhoon eye trajectory based on a mathematical model: Comparing with observational data, Nonlinear Analysis: Real World Applications, 11, 1847–1861, 2010.

Su, Y., Chelluboina, S., Hahsler, M., and Dunham, M. H.: A New Data Mining Model for Hurricane Intensity Prediction[A], in: Data Mining Workshops (ICDMW), 2010 IEEE International Conference, 98–105, doi:10.1109/ICDMW.2010.158, 2010.

Weber, H. C.: Probabilistic prediction of tropical cyclones, Part I: Position, Mon. Weather Rev., 133, 1840–1852, 2005.

Zhao, Y., Liu, Y.-H., Yu, X.-G., Wei, D., Shan, C.-W., and Zhao, Y.: Method for mobile path prediction based on pattern mining and matching, Journal of Jilin University (Engineering and Technology Edition), 38, 1125–1130, 2008.

Zou, L., Ren, A.-Z., Xu, F., and Zhang, X.: Typhoon track forecasting based on GIS spatial analyses, Journal of Tsinghua University (Science and Technology), 48, 2036–2040, 2008.