Track Similarity-based Typhoon Search Engine for Disaster Preparedness

C.M. Hsieh[a], C.Y. Ho[b], H.K. Kung[b], H.Y. Chan[c], M.H. Tsai[c] and Y.C. Tsai[d]

[a]Department of Economics, National Taiwan University, Taiwan
[b]Department of Geography, National Taiwan University, Taiwan
[c]Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan
[d]School of Big Data Management, Waishuanghsi Campus, Soochow University, Taiwan
E-mail: b06303059@ntu.edu.tw, b06208030@ntu.edu.tw, b06208001@ntu.edu.tw, d10705005@mail.ntust.edu.tw, menghan@mail.ntust.edu.tw, pecutsai@scu.edu.tw

Abstract –

Plenty of variance lies in the development stage of a typhoon, including its location, direction, and surrounding meteorological conditions. Affected by such numerous contributing factors, it is not always easy for nowaday technology to make accurate typhoon forecasts. The experiences from the past typhoons may provide decision-makers with sufficient information to anticipate potential damage and thus develop appropriate strategies. The typhoon track plays a vital role in selecting appropriate historical cases based on the need to grasp the situation during the disaster preparedness phase. This paper summarizes our published journal article, which proposed an algorithm to compare the similarity between the current typhoon forecast track and all the typhoon tracks of the western North Pacific in the past. Based on the forecast track points of the current typhoon, the algorithm suggests a list of historical typhoons with the highest similarity in tracks. Hence, the tracks and the disastrous area of those historical typhoons can be used as crucial alerts and hints for disaster preparedness. Inside the algorithm, the mechanism follows the Recentness Dominance Principle, which is elaborated in our published journal article. The principle states that the more recent the forecast track point is, the higher the weight that it possesses and thus should be emphasized. For implementation, the study develops a user-friendly front-end interface that synchronizes the latest forecast typhoon tracks from six major meteorological institutions automatically, including CWB, HKO, JMA, JTWC, KMA, and NMC. The result comes with a convenient and concise search engine that assists decision-makers in finding similar historical typhoon records.

Keywords –

typhoon track similarity; search engine; user interface design; disaster preparedness; decision support

1 Introduction

Precise response strategies to typhoons rely on accurate forecasts. However, plenty of variance lies in the development stage of a typhoon, including its location, direction, and surrounding meteorological conditions. Affected by such numerous contributing factors, it is not always easy for nowaday technology to make accurate typhoon forecasts, causing the difficulties of developing precise strategies.

The experiences from the past typhoons may provide decision-makers with sufficient information to anticipate potential damage and thus develop appropriate strategies for precaution. The damage of a typhoon, caused by the winds and rainfall, is regarded as a consequence of the interaction between the typhoon and the land. Such interactions are affected by typhoon tracks, making the track a critical factor for disaster reduction of typhoon events. Based on the need to predict and grasp the situation, during the disaster preparedness phase, the typhoon track plays a vital role in selecting appropriate historical cases. Decision-makers may anticipate the impact of an upcoming typhoon by utilizing the historical typhoon records with similar tracks.

The categorization of historical typhoon tracks helps to analyze the patterns of damage caused by different typhoons. For example, the Central Weather Bureau (CWB) of Taiwan has determined nine categories of tracks of typhoons that invades Taiwan [1]. The CWB has also linked the behavior of wind and rainfall of typhoons with different tracks to different regions in Taiwan [1], supporting decision-makers to figure out the potential damage of upcoming typhoons.

This paper summarizes our published journal article [2], which proposes a track similarity algorithm to compare the similarity between an upcoming typhoon forecast track and all the tropical cyclones (TCs) in the western North Pacific (WNP) in the past. A comparison model, which is the core of the algorithm, was first developed in our preliminary work [3], and in the present study, we further explore the details of the model and complete the comprehensive algorithm. We also developed a new information display panel featuring auto-importing forecast data. A literature review of typhoon databases is presented in Section 2, including a discussion of our preliminary work. The algorithm and its detailed discussion are discussed in Section 3. The design and implementation of the new panel are described in Section 4. More information can be found in the journal article [2].

2 Literature Review

Many studies about typhoons focus on forecast model performance. Classical prediction techniques are climatological or dynamic with different considerations. For example, Lee et al. [4] developed a climatology model for predicting rainfall at different areas for a specified typhoon center location. On the other hand, some recent studies discuss data-driven approaches to typhoon forecasts. For example, Rüttgers et al. [5] utilized a generative adversarial network (GAN) with satellite images for typhoon track predictions. Although precise forecasts support decision making, the deduction of damage from the distribution of rainfall and wind is not explicit and requires expertise. Also, forecasts cost computation time, hindering immediate reactions to emergencies. Hence, utilizing past typhoon records for developing quicker response strategies may be useful for decision-makers.

The experiences from the past typhoons may provide decision-makers sufficient information to anticipate potential damage and thus develop appropriate strategies for precaution. Wu and Kuo [6] states that significant mesoscale variations in weather are derivative from the interaction of typhoons and the complicated topography in Taiwan. Also, Huang et al. [7] has investigated the relationship between typhoon track, rainfall patterns, and flood peak time, concluding that that preferable rainfall types vary by typhoon tracks. For the preparedness and the emergency response phase of typhoon events in Taiwan, decision-makers seek the status of the disaster, the response strategies, and other relevant data of similar past typhoons for decision making support [2].

Several databases for TCs in the western North Pacific WNP have been developed, including the Taiwan Typhoon Database of CWB [8] and the Digital Typhoon Database of the National Institute of Informatics (NII) of Japan [9]. Such databases feature filtering by time, pressure, wind, rainfall, intensity, etc. However, searching by TC track is not commonly provided. The Taiwan Typhoon Database allows searching by the track categories proposed by the CWB, but the actual category that a forecast track belongs to still requires the users' determination. The task of track matching is less intuitive for users due to the lack of explicit integration with forecast and historical records.

T-search, a real-time historical typhoon search engine, was introduced by the authors to assist relevant personnel to intuitively search and review the historical typhoon information and efficiently develop corresponding response strategies [3]. A track similarity comparison model was also developed in the work. It also provided an intuitive cross-platform user interface for retrieving the historical typhoon records. T-search has been validated by a usability test and a three-year field test coordinated with the government of Taiwan. The real case study has shown that T-search enhances the accessibility to typhoon records. However, the preliminary work did not discuss the usage of each parameter in the comparison model, making the parameter setting without baseless. Also, the large number of typhoon records and track points resulted in long calculation times for the track similarity algorithm. Further, the system only allowed users to input the forecast data manually instead of automatically import, thus reducing the convenience of the system.

3 Methodology

This paper summarizes our published journal article [2], which proposes an algorithm to compare the similarity between an upcoming typhoon track forecast and all the past TC tracks in the WNP. The proposed algorithm is composed of the comparison model developed in the preliminary work [3] and a ranking strategy. In the present study, we decompose the comparison model into a static sector and a dynamic sector, propose a strategy to enhance the efficiency of the computation and set an order to ensure the discrimination of similarity. Also, the Recentness Dominance Principle (RDP) is introduced based on the presumption of the high uncertainty of the typhoon's direction during development [2]. The RDP states that those later user-specific track points are more relevant than the earlier ones, and thus their relative weighting should be increased. We extend elaborates on the preliminary work and provides the definition and interpretation of the time weighting. More information can be found in the journal article [2].

3.1 Comparison Model

The comparison model selects the most similar historical typhoon track for the user-specific track of the forecasted typhoon through track comparison and matching. For each past typhoon, the model calculates the similarity based on points of tracks, valid regions given by the user, time, etc., for the user-specific track points and quantifies scores for comparison. The similarity of the forecasted typhoon and each historical typhoon is calculated with Equations (1) and (2):

$$similarity(i) = \sum_{k=1}^{M}(1+kw)\times\delta_k = \sum_{k=1}^{M}\delta_k + w\sum_{k=1}^{M}k\times\delta_k \quad (1)$$

$$\delta_k = \begin{cases} 1, d_{jk} < R_k \\ 0, otherwise \end{cases} \quad (2)$$

M is the number of the user-specific track points. k is the index of the user-specific track points starting from 1. i is the index of the historical typhoons. j is the index of the closest track point of the historical typhoon i to the user-specific track point k. w is the time weighting of all the user-specific points, which determines the importance of each point. The binary variable δ_k represents whether the historical typhoon track point j falls in the given valid region of the user-specific track point k. The valid regions corresponding to the user-specific track points are designed to restrict the distances between a forecast track point and historical track points, denoted as d_{jk} in Equation (2). At least one point of each historical typhoon track should lie within the valid region set by the user-specified radius, denoted as R_k in Equation (2), or the similarity scores of the typhoons are zero since they are determined to be too far.

For discussing the roles of k and w, Equation (1) is divided into the static sector and the dynamic sector. The static sector refers to the constant part constant part ($\sum_{k=1}^{M}\delta_k$), while the dynamic sector refers to the second component ($w\sum_{k=1}^{M}k\times\delta_k$). In the static sector, only the valid regions of the given track points are of concern, indicating that how many given track points of a forecast typhoon a historical typhoon passes. The dynamic sector includes time as a factor in the model by multiplying the time series (k) by the time weighting (w). The adjustment to the given time weighting emphasizes the importance of time of the track points.

Both the static sector and the dynamic sector impact the result of similarity comparison. An example is shown in Figure 1, with one typhoon passing through the last three input points with the score of $3+12w$, and the other passes through the first three input points and the last input point with the score of $4+11w$. In this case, the first typhoon has a higher static sector than the second, but if the time weighting w is greater than one, the second typhoon eventually obtains higher similarity. Hence, it is essential to discuss the usage of time weighting for a comprehensive explanation.

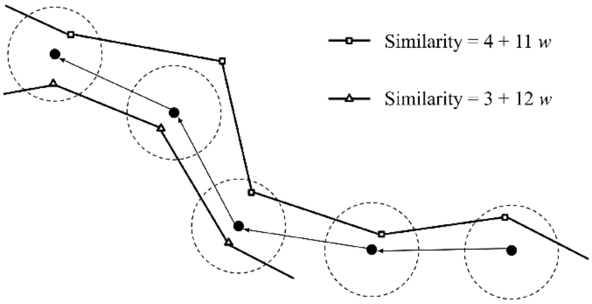


Figure 1. An example of two typhoon tracks for the similarity comparison demonstrating the impact of the dynamic sector [2].

3.2 Recentness Dominance Principle

The Recentness Dominance Principle states that the more recent the forecast track point is, the higher the weight that it possesses and thus should be emphasized [2]. Since a TC direction is highly uncertain during its developing stage, we tend to emphasize the importance of later track points by giving higher weightings to them. Following the RDP, the time weighting should be changed when more track points are given. The similarity score of a historical typhoon may be distorted in some circumstances if the time weighting is not adequately set following the RDP or does not correlate with the number of given track points.

For ensuring the RDP, we introduce the global recentness dominance time weighting (gRDW) and the individual recentness dominance time weighting (iRDW). They are obtained by the following approach (Equations (3) and (4)):

$$\hat{w} = \max\{w_i : i = 1..|P|\} \quad (3)$$

$$w_i = \frac{s_{i,b} - s_{i,a}}{d_{i,a} - d_{i,b}} \forall\left(s_{i,a} + d_{i,a}w, s_{i,b} + d_{i,b}w\right) \quad (4)$$

$$P \equiv (s_a + d_a w, s_b + d_b w) \forall a, b \in \mathbb{N}(s_a > s_b \wedge d_a < d_b) \quad (5)$$

The value of gRDW (denoted as $\hat{w}$ in Equation (3)) determines the number of the indeterministic score pairs (denoted as P in Equations (3) and (4)), which are defined in Equation (5). A properly specified value of gRDW eliminates the indeterministic score pair that contravenes RDP, and the similarity scores can then be calculated. In this study, the model equalizes the score of each score pair to obtain iRDW (denoted as w_i in

Equations (3) and (4)) for simplicity. To emphasize the recentness, the model takes the maximum of the set of iRDW as gRDW.

The dynamic sector of the model constructs a gRDW that ensures the dominance of recentness. To discuss the effect of setting different values of time weighting, we relax RDP by allowing users to set the preferred time weighting. Seven special weight groups are divided, giving different interpretations of the time weighting. One of the weight group, when w equals gRDW, is the critical value to ensure RDP. The remaining six groups are:

- Over gRDW: $w \in (gRDW, \infty)$. In this case, the outcome of the similarity model remains identical to the outcome's time weighting set as gRDW, showing that exaggerating RDP generates the same result.
- Positive related time weighting: $w \in (0, gRDW)$. In this case, the time weighting determines whether the impact of the dynamic sector can defeat the static sector, providing a trade-off between the two sectors. The closer values are to gRDW, the more influential the recentness.
- Static sector: $w = 0$, only presenting the static sector's score. It eliminates the dynamic sector, presenting the neat summation of the historical typhoon track passing through the valid regions of the user-specific track points.
- Negative related time weighting: $w \in (-1, 0)$. The recentness is proportionally decreased in this case, making older typhoons more important than recent typhoons.
- Repulsion force weighting: $w = -1$. In this case, the comparison model reduces to $(\sum_{k=1}^{M}(1-k)$. Mathematically, the similarity score must not be greater than zero, referring to select historical typhoons who pass through none ($M = 0$) or only the first ($k = 1$) of the given points.
- Reverse similarity: $w \in (-\infty, -1)$. In this case, the model simply eliminates all the historical typhoons that pass through the user-specific track points.

Setting time weightings with the six groups may not provide proper search results since they are obtained by relaxing the RDP. The actual usage of non-RDP time weighting is not yet clarified; however, the discussion of the behavior of search results with time weightings of different groups potentially provides insight for further usage.

3.3 Ranking Strategy

For the distinction of similarity between the historical typhoon tracks and the user-specific track, the ranking strategy is in five steps as follows:

1. Total similarity score;
2. Dynamic sector;
3. Static sector;
4. Month disparity;
5. Year gap.

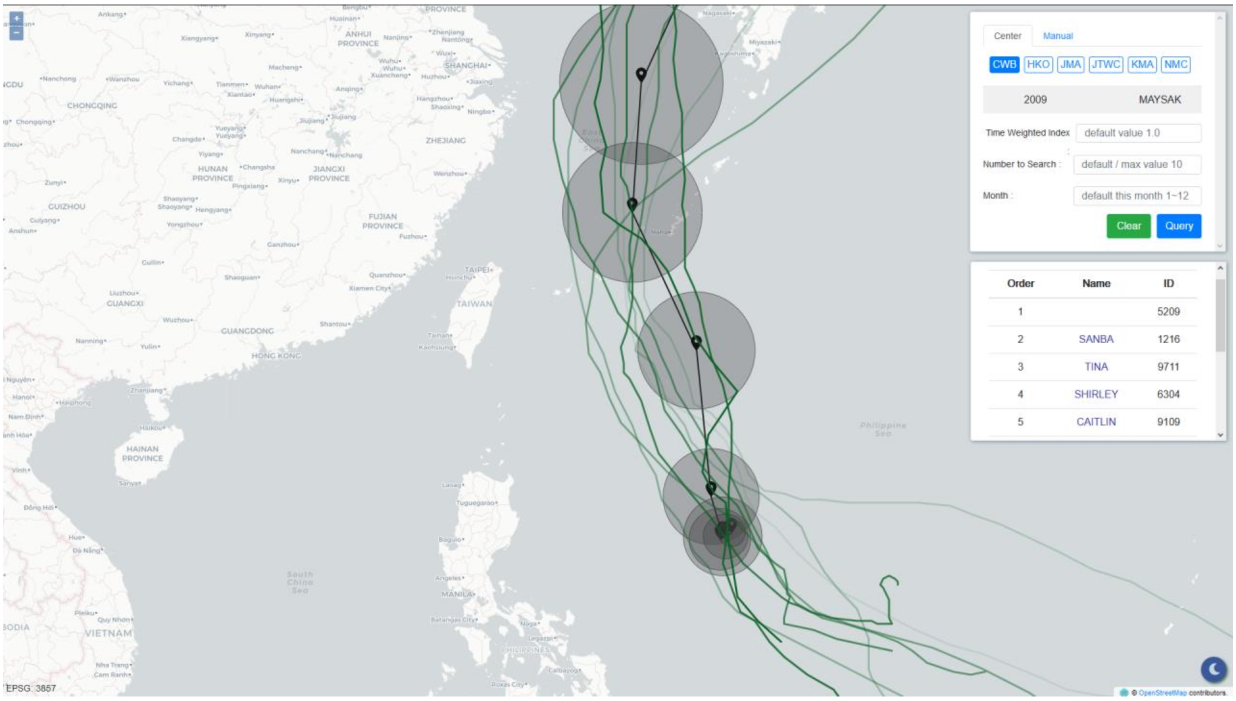
First, the model sorts the past typhoons by the computed similarity scores with the user-specific track using Equation (1). The similarity score considers RDP by utilizing the time weighting assigned to each given track points with higher values for later points, indicating that the score is an index of general similarity. When the scores of some historical typhoons are equally matched, the dynamic sectors are compared first, and then the static sectors are compared, which reaffirms that RDP does matter. If the ranks still make no difference, the algorithm keeps comparing by the disparity of months, and the gap between the present year and the year of each historical typhoon. Finally, the comparison result should be clear.

4 Implementation

This study re-designs and develops a user-friendly interface that automatically synchronizes the latest forecast typhoon tracks from six major meteorological institutions of TCs in the WNP. The interface is implemented in the form of a web application, consisting of a front-end panel that displays the information and a back-end server that handles the data retrieving and the similarity comparison.

All required typhoon data are collected from the meteorological institutions. For the historical typhoons, the system utilizes the historical best track data of TCs in the WNP provided by the Japan Meteorological Agency (JMA), the regional specialized meteorological center of the WNP, and the South China Sea [10], as the references of similarity calculation. The TC data in the jurisdiction of the JMA since 1951 are retrieved from the JMA website. The system parses the raw data and extracts the approximately 1700 typhoons with the contained approximately 65,000 coordinates.

Meanwhile, typhoon track forecasts, the input of the system, are retrieved from several meteorological institutions that provide forecasts of present typhoons in the WNP. In this study, the real-time forecasts are obtained from the Integrated Multi-Agency Tropical Cyclone Forecast, which provides comprehensive information about current typhoon forecasts from several meteorological institutions, including the CWB, the Hong Kong Observatory (HKO), the JMA, the Joint Typhoon Warning Center (JTWC) of the US, the Korea Meteorological Administration (KMA), and the National Meteorological Center (NMC) of China [11].



2009 MAYSAK

Time Weighted Index: default value 1.0

Number to Search : default / max value 10

Month : default this month 1~12

Order	Name	ID
1		5209
2	SANBA	1216
3	TINA	9711
4	SHIRLEY	6304
5	CAITLIN	9109

Figure 2. The information display panel and the query result using Typhoon Maysak (2020).

The information display panel contains a map view built with OpenLayers libraries and a control panel for inputting track points. The system provides two input modes:

- Center auto-import, which offers real-time feeds and forecasts of current typhoons provided by the meteorology institutions;
- Manual input, which allows users to complete comparisons between any desired tracks and historical typhoon tracks by either keying in coordinates or clicking on the map view.

The integration of multi-institution forecasts allows the decision-maker to choose the exact institution as the forecast data provider considering the need. By auto-importing the forecast typhoon track, the time cost for manual input is reduced, and the accuracy of the track data is also improved.

Figure 2 illustrates the overview of the information display panel and the result using Maysak (2020) as an example. The control panel on the right lists all active TCs in the WNP and available forecasts from different meteorological institutions for users to select. In the case of Figure 2, the user selects the forecast data provided by CWB. After auto-import, the map view illustrates the forecast typhoon track points and all the valid regions. Next, the user activates the query. The system starts calculating, finds the most similar historical typhoons, and demonstrates the tracks of the selected historical typhoons on the map view. The control panel also lists the selected historical typhoons and directly links to the Taiwan Typhoon Database of CWB. With the support of the panel, it is convenient for decision-makers to review the historical typhoon records efficiently. The damages caused by typhoons can thus be forecast, and the response strategies such as evacuations and pump pre-allocations can be executed on time. More information can be found in the journal article [2].

5 Conclusions

This study has improved the preliminary work and developed a new information display panel featuring auto-importing forecast data, coming with a convenient and concise search engine that assists decision-makers in finding similar historical typhoon records. In this paper, summarized from our published journal article [2], we have discussed the details of the comparison model and completed the ranking strategy. In addition, we have stated the time weightings as a vital role of controlling the impact of time of each track point for the similarity algorithm. The use of RDP has been proposed regarding the effect of the number of track points. Furthermore, the effect of varying time weightings potentially provides insight for decision-makers to modify the time weighting for different purposes based on the need or their profession. We have also developed a new information display panel, featuring both the manual-input method and the center auto-import method to forecast track data, reducing the required time to import data and the possibility of making mistakes in the exact locations of the track points. Several forecasters are additionally integrated into the system, providing various choices for the decision-makers. More information can be found in the article [2].

In future work, the algorithm's insufficiency of only considering track points as input should be solved. Since there are numerous factors that affect the damage of a typhoon, taking other factors such as the intensity of typhoons into consideration may enhance the search engine to get a better search result that meets up to the current scenario. Also, the algorithm can be adapted to utilize ensemble forecasting to comprehensively demonstrate forecasting uncertainty and support the decision-makers to evaluate the risk of different circumstances.

Acknowledgment

We thank Shih-Chung Kang, Yih-Chi Tan, Jihn-sung Lai, James Yichu Chen, Cheng-Hsuan Yang, and Tzong-Hann Wu of National Taiwan University; and Wei-Chyuan Shieh and Chun-Nan Chen of the Meteorological Application and Development Foundation for their contribution to the preliminary work.

References

[1] Central Weather Bureau. Faq for typhoon. On-line: https://www.cwb.gov.tw/V8/E/K/Encyclopedia/typhoon/. Accessed: 01-06-2020.

[2] Tsai M.H., Chan H.Y., Hsieh C.M., Ho C.Y.,. Kung H.K, Tsai Y.C., and Cho I.C. Historical typhoon search engine based on track similarity. *International Journal of Environmental Research and Public Health*, 16(24):4879, 2019. doi:10.3390/ijerph16244879.

[3] Tsai M.H., Chen J.Y. and Kang S.C. T-search: A cross-platform searcher for historical typhoon events. *Journal of the Chinese Institute of Civil and Hydraulic Engineering*, 31(4):351–360, 2019.

[4] Lee C.S., Huang L.R., Shen H.S. and Wang S.T. A climatology model for forecasting typhoon rainfall in taiwan. Natural Hazards, 37(1):87–105, 2006. doi:10.1007/s11069-005-4658-8.

[5] Rüttgers M., Lee S., Jeon S. and You D. Prediction of a typhoon track using a generative adversarial network and satellite images. *Scientific Reports*, 9(1):6057, 2019

[6] Wu C.C. and Kuo Y.H. Typhoons affecting taiwan: Current understanding and future challenges. *Bulletin of the American Meteorological Society*, 80(1):67–80, 1999.

[7] Huang J.C., Yu C.K., Lee J.Y., Cheng L.W., Lee T.Y. and Kao S.J. Linking typhoon tracks and spatial rainfall patterns for improving flood lead time predictions over a mesoscale mountainous watershed. *Water Resources Research*, 48(9), 2012. doi:10.1029/2011WR011508.

[8] Central Weather Bureau. Typhoon database. Online: https: //rdc28.cwb.gov.tw/TDB/. Accessed: 22-072019.

[9] National Institute of Informatics. Digital typhoon: Typhoon images and information. Online: http:// agora.ex.nii.ac.jp/digital-typhoon/. Accessed: 02-10-2019.

[10] Japan Meteorological Agency. Rsmc tokyo - typhoon center. Online: https: //www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/RSMC_HP.htm. Accessed: 01-10-2019.

[11] Typhoon2000.com. Integrated multi-agency tropical cyclone forecast. Online: http://www. typhoon2000.ph/multi/. Accessed: 01-10-2019.