# An Improved Sequential Data Mining Model for Hurricane Trajectory Prediction

Jasmin Bissonnette, Caden Marofke, Taylor Cox
Department of Computer Science
University of Manitoba
Winnipeg, Manitoba
Email: {bissonnj, marofkec, coxt3}@myumanitoba.ca

***Abstract*—**

***Index Terms*—Hurricane Trajectory, Apriori, Prediction**

## I. Introduction

## II. Related Work

## III. Experimental Setup

*1) Distance Formulas:* Accurate distance measurement is a critical task in hurricane trajectory prediction. Pattern matching strategies in the proposed hurricane trajectory prediction model depend on coherent distance measurements between real-world spatial points. Consider the following prediction-correctness evaulation method proposed by Dong et al (1). To determine whether a trajectory prediction is correct, the distance between the first p points of the predicted and actual trajectories is compared against a defined maximum, where p is the minimum length of the two trajectories. If all p pairs of points fit within the maximum distance, the trajectory prediction is considered correct.

Two different distance measurement formulas are evaluated in this work. This work compares the Euclidean distance formula against the Haversine distance formula. The Euclidean distance formula (as used in (1)) is designed to determine the unit distance between two N dimensional points in a straight line. In the proposed model, units correspond to degrees of latitude and longtitude. In contrast, the Haversine formula is designed to determine the kilometers between two latitude-longtitude points over a large sphere (such as the earth). For this reason, the Haversine formula is anticipated to provide a more accurate definition of distance between recorded hurricane points. When used in this work, the Euclidean distance formula is evaluated with two dimensions, correspoding respectively to latitude and longtitude. The two formulas, denoted Deucl and Dhavr are defined as follows:

To compare the prediction accuracy between the two proposed formulas, a sample execution of the hurricane trajectory prediction model is executed with identical variables, except for the difference in distance evaluation. The distance formulas are interchanged in the prediction correctness checking step of the model, described previously in this section. When using the Euclidean formula, the distance threshold is 1. This corresponds to the distance threshold proposed in (1). When the Haversine formula is employed, the maximal distance is set to 450km. This value corresponds to slightly less than approximately one average hurricane-diameter. Since the average hurricane diameter is between 333 and 670km (2), two points are considered to be within the distance threshold if they are less than one average hurricane-diameter apart.

Between the Euclidean and Haversine distance formulas, the Haversine formula delivers a increased correctness rate of 39.3%, corresponding to 72.7% compared to 33.4% correctness when the Euclidean formula was employed. The realistic nature of the Haversine formula in conjunction with its corresponding maximal distance threshold is able to provide a higher rate of accurate predictions compared to the Euclidean distance formula, which is not designed for non-euclidean coordinate systems such as coordinate systems over spheres.

Table of parameters:

- Weighted Training data: No No - Training Data: 1950-2000 1950-2000 - Testing Data: 2001-2015 2001-2015 - Minsup: 30 30 - Minconf: 0.25 0.25 - Region Method: DC DC - Fitness Function: [1] [1] - Rule Fit Flexible: No No - Correctness Method: [1] [1] - Distance Formula: Euclidean Haversine - Distance Threshold: 1 (unit) 450 (km) ———————————————— - Correctness ratio 33.4% 72.7%

(DC: Disretized Coordinates)

## IV. Experimental Results

## V. Conclusions and Future Work

## References

[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]