A Statistical Method for Detecting Move, Stop, and Noise Episodes in Trajectories

Tales P. Nogueira[1], Hervé Martin[1], Rossana M. C. Andrade[2]

[1]Univ. Grenoble Alpes
CNRS, Grenoble INP, LIG
F-38000 Grenoble, France

[2]Group of Computer Networks, Software Engineering and Systems
Federal University of Ceará
Fortaleza, Brazil

tales@great.ufc.br, herve.martin@imag.fr, rossana@great.ufc.br

Abstract. *Detecting stops is an important task in trajectory analysis. Stops can reveal interesting aspects of a moving object behavior such as its daily routine, bottlenecks in traffic jams, or visiting times of touristic places. In order to record those traces, trajectories must be sampled and, in some cases, post-processed. This process from collecting raw data to storing them may vary according to the devices and applications that collect the data. Another important characteristic in many trajectories is the presence of noisy segments, a fact is often ignored by most stop detection methods. In this work, we present a method that exploits gaps in time and space to identify episodes of movement, stop, and periods where some classification is inconclusive, which we define as noise. In addition, our method does not rely on contextual information as opposed to some current methods, which makes our proposal also suitable for trajectories recorded in free space.*

1. Introduction

The ubiquitous presence of trajectory data is constantly growing in our digital lives and we are constantly producing it in many ways. Structuring trajectories into periods of stops and moves has been proved to be a fundamental task (Spaccapietra et al. 2008) in trajectory analysis. In fact, different criteria can be used to segment trajectories (Alewijnse et al. 2014; Buchin et al. 2011), expanding the possibilities of structuring moving object traces beyond the stop-move model. Viewing trajectories as sequences of moves and stops can be the first step towards a more complex model for trajectory analysis.

Trajectories are continuous events in real life. However, they must be treated as discrete events in order to be recorded. Different sampling rates and optimizations can be used in this process that may hinder the ability of stop detection algorithms to correctly detect the different parts of a trajectory.

Detecting occurrence and absence of movement is a fundamental segmentation task that has been vastly explored in the literature (Alvares et al. 2007; de Graaff et al. 2016; Palma et al. 2008; Rocha et al. 2010; Yan et al. 2010). Applications that deal with real-world data also have to deal with noisy measurements which, in some cases, makes it impossible to determine the actual state of the moving object. Although some related works have considered the presence of noise in trajectories, they usually handle this by previously smoothing or by using additional metadata that is not always available.

The characteristics of a recorded trajectory can vary broadly according to a range of factors such as sensor's physical components, sampling rate, post-processing algorithms, environmental conditions. The factors may yield trajectories with different levels of quality even for traces captured by the same device. Therefore, this facet of spatio-temporal data research disfavor the possibility of proposing a universal method for detecting stops and moves as well as trajectory segmentation methods based on other criteria such as speed or direction.

In this context, a method for stop detection should consider how data is recorded and stored in order to have good performance. Thus, it is necessary to make assumptions about collected data before proposing an approach to trajectory segmentation.

We observe that relevant methods for identifying stops rely on the assumption that trajectories are sampled at regular intervals of time. This assumption allows the application of clustering algorithms to identify points near each other and then classify groups of points as stops according to some temporal threshold. However, this assumption may not hold due to a variety of reasons, such as periods of GPS failure, noisy measurements, different sampling strategies, pre-processing procedures, among other factors.

Andrienko et al. (2008) defined how a trajectory can be observed according to various sampling strategies as follows: *time-based*, when positions are recorded at regular intervals of time; *change-based*, when positions are recorded only when the object moves; *location-based*, when the location is collected only if the object approaches a specific location, e.g. near a sensor; *event-based*, when the moving object performs a specific action, e.g. making a call; and various combinations of these methods. While most of the state-of-the-art methods of stop detection deal mainly with the *time-based* recording strategy, it should be noted that some applications may store trajectories following any combination of the above types. Also, applications may make modifications to the captured data in order to eliminate redundant information. In this scenario, algorithms that rely on clustering points located near each other are most likely to fail.

In this paper, we describe a way of creating episodes based on the detection of stops and moves during a single trajectory. The main assumption of our method is that, for a given trajectory, points may not sampled at the same frequency along the path. In other words, we consider the existence of a post-processing filtering phase that discards redundant nearby points or stops recording points when the object is not moving, a fact that can be observed in many applications. Also, we consider that the sampling rate is approximately constant when the object is moving, i.e. new points are recorded at near equally spaced intervals of time.

Another important difference in our method is that the notion of stop in other works is usually related to the identification of Regions of Interest, allowing the classification of a point as a stop even when there is some movement. In our case, we aim at identifying locations where an actual stop happened. Moreover, our proposal does not need external data (e.g. polygons of adjacent geographic features) or additional sensor data (e.g. GPS accuracy information). This characteristic of our proposal can be appealing for applications that deal with trajectories recorded in free space.

The remainder of this paper is organized as follows: Section 2 present relevant work devoted to detecting stops in trajectories. Section 3 describes characteristics of the dataset considered in this research. Section 4 explains the Outlier Labeling Rule, which is the base of our method. Section 5 present the details of our contribution, the MSN (Move-Stop-Noise) algorithm, which is compared to other important methods in Section 6. Section 7 encloses our conclusions and perspectives of future work.

2. Related Work

We can observe that many state-of-the-art stop detection methods rely on some assumptions about the gathered data and, in some cases, additional external data. The SMoT method (Alvares et al. 2007) classifies as stops the trajectory points that intersect "candidate stops", i.e. a previously defined set of polygons, each one associated to a minimum time duration. A major weakness of this approach is the need for manually selecting candidate stop polygons as well as minimal time durations needed to consider each region as a stop. Putting a hard threshold on the duration of stop may cause the algorithm to miss important stops that have a time duration close to the threshold.

The SMoT method was later extended by the SMoT+ algorithm (Moreno et al. 2014). SMoT+ is able to identify stops in different levels of granularity (e.g. a shop inside a mall which is located in a town). SMoT+ presents the same drawbacks of SMoT as their parameters are very similar. The concept of Interesting Sites (IS) is similar to SMoT's candidate stops. Additionally, there is an additional parameter representing a hierarchy of containments among the sites.

The PIE algorithm (de Graaff et al. 2016) uses the underlying geography polygons, but it also considers reductions in speed, changes in direction and the accuracy of each GPS point. Whereas speed and direction can be easily computed from trajectory points, the availability of accuracy data, while very important to assess signal quality, is not commonly stored by most applications. This factor imposes an important obstacle to use this method with trajectories captured by third-party applications.

Palma et al. (2008) proposed the CB-SMoT algorithm. CB-SMoT considers that a moving object's speed decreases significantly when an interesting place is being visited (therefore, it is a stop). However, they also assume is that the recording device keeps storing points even when the object is stopped, thus stops are characterized as regions with a greater spatial density of points. Both SMoT and CB-SMoT were reused by Moreno et al. (2010) to identify stops and infer behavior of moving objects.

Yan et al. (2010) proposed a model and computing platform for abstracting trajectories at different levels of abstraction. In the first layer of their computing platform, trajectories are smoothed and outliers are identified by velocity thresholds according to domain knowledge (e.g. car, human, bicycle etc.) In the Trajectory Structure Layer, the identification of stops is done by determining a speed threshold based on the type of moving object and a function that takes into account the moving object's average speed and the average speed of other moving objects. For calculating the latter, the space is divided into a grid and an average speed is associated to each cell. Differently from our proposal, the authors have used the non-robust average speed measure, which may difficult a correct identification of stops if there is a large range of speeds in a single trajectory. Second, while there was an effort to dynamically set speed thresholds, this has not been done to

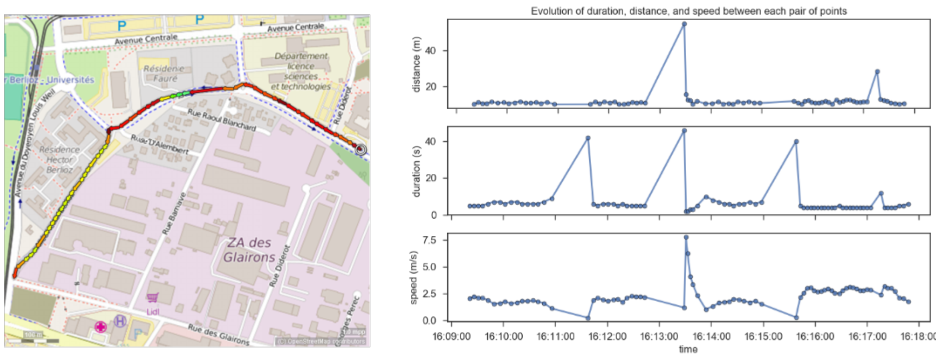


Figure 1. An example trajectory and its speed, distance, and duration time series.

the stop duration, which is still defined as an absolute metric value (e.g. 15 minutes).

Nogueira et al. 2014 proposed a statistical method for detecting candidate stops. Its only parameter is a minimum speed for a point to be considered as a stop. However, they did not consider noisy trajectories segments. Moreover, non robust statistic measures were used as they have relied in standard deviations from the mean, a metric that can be easily broken by large outliers.

3. Exploratory Data Analysis

A useful task when analyzing a dataset is verifying the correlation strength among its variables. The output of this analysis usually highlights the relationships of variables that tend to better explain the dataset variability. We have used the Spearman correlation because it is more resistant to outliers as it diminishes the importance of extreme scores by first ranking the two variables and then correlating the ranks instead of the actual values.

Figure 1 shows an illustrative trajectory that was recorded in a controlled manner in order to have two stops of a few seconds and two periods of noise that have been simulated by turning off the smartphone's GPS for a few seconds.

For each pair of sequential points in the trajectory of Figure 1, we have calculated its speed, distance and duration. We can observe some interesting characteristics based on previous knowledge about this particular trajectory. The trajectory starts with distances of about 10 meters between points, durations of 5 to 7 seconds, and a fairly constant speed until there is a peak of 42 seconds in the duration between points series. At the same time, we can observe that the speed drops to a value near to zero while the distance remains unchanged. This characterizes a stop taking into consideration the characteristics of this dataset. Some seconds later, another peak in duration is noticeable at the same time of a peak in the distance between points that are not followed by a decrease in speed. This characterizes a period of noise. In the remaining of the trajectory, another stop and another noise period can be noticed with these same characteristics.

Table 1 shows the mean Spearman correlation among movement attributes of 2226 trajectories, which were collected from a widely used third-party mobile application for tracking sport activities. Walking and running activities were selected. These trajectories range from 2 to 42 kilometers, they are located in Grenoble (France) and Barcelona

Table 1. Median of Spearman correlations among attributes of 2226 trajectories

	duration	distance	speed	acceleration
duration	1	-	-	-
distance	0.16	1	-	-
speed	-0.86	0.29	1	-
acceleration	0.34	-0.03	-0.36	1

(Spain), and they have been recorded with Android smartphones. From this data, we can observe that the pairing between speed and duration is the one that presents the strongest correlation. In this case, a strong negative correlation indicates that when the values of duration increase, the values of speed tend to decrease and vice-versa, which is what one can expect given the previously explained assumptions about the data.

From this exploratory data analysis, we can conclude that there is a negative correlation between the values of speed and duration that characterizes a stop. For the noisy cases, there is no pair of variables that helps the classification. Thus, we make use of the assumption that points are recorded at near constant distance intervals most of the time.

4. Outlier Labeling Rule

Based on the exploratory data study, we can approach the classification of moves, stops and noise as an outlier detection problem. In order to identify outliers in time series, we use the modified z-score proposed by Iglewicz and Hoaglin (1993). The usage of this method is motivated by the poor performance of other popular measures like the standard deviation and the mean in the presence of outliers.

An indicator of the robustness of a statistic is its breakdown point, i.e. the maximum proportion of outlier data points that can be added to a dataset before the statistic gives a wrong result. The *mean* has a breakdown point of 0% because if just one value of a given series is set to infinity, its mean goes to infinity. On the other hand, the *median* has a high breakdown point because the median value of a series is only affected if more than 50% of the data is set to infinity.

Another estimator that is easily modified in the presence of outliers is the standard deviation, as it takes into consideration the squared distance from the mean for each value. According to Huber and Ronchetti (2009), the most useful ancillary estimate of scale is the MAD (see Equation 1), which is the median of absolute distances from a series' median. The constant scale factor 1.4826 makes the MAD unbiased at the normal distribution (Rousseeuw and Hubert 2011).

$$MAD = 1.4826 \times median(|Y_i - \tilde{Y}|) \tag{1}$$

Iglewicz and Hoaglin (1993) recommend using the modified z-score shown in Equation 2 where each element of a series is subtracted from the median ($\tilde{x}$), multiplied by a factor to make the MAD consistent at the normal distribution (0.6745). As a recommendation from the authors, points having modified z-scores with an absolute value greater than 3.5 have a high probability of being outliers (NIST/SEMATECH 2012). Another advantage of using the MAD statistic is the fact that it is also adequate for application in populations that do not fit perfectly a Gaussian distribution (Gorard 2005), which is the case for real world GPS track datasets.

$$M_i = \frac{0.6745(x_i - \tilde{x})}{\text{MAD}} \tag{2}$$

5. The MSN algorithm

Our statistical method for stop, move, and noise (MSN) detection builds upon the previously explained theoretical background.

Considering a trajectory $\tau = \{(s_1, t_1), (s_2, t_2), ...(s_n, t_n)\}$, where each position $s_i = (lat, lon)$ is a pair of latitude and longitude coordinates, and each time instant t_i is represented by a timestamp, for each pair of points $(s_i, t_i), (s_{i+1}, t_{i+1})$, we compute its distance, duration, speed, and turning angle. Then, we store these values in their respective time series S_τ, T_τ, V_τ, A_τ. For A_τ, a turning angle consists on the angle formed by three neighboring points.

From the above time series, we can formulate an algorithm for determining which instants of the trajectory are likely to be stops, moves or undefined states, considered as noise in this work (see Algorithm 1). The algorithm's input are the initial calculated time series besides the thresholds ϵ_s, ϵ_t, and ϵ_v representing the modified z-score limits for distance, duration, and speed. Additionally, a minimum turning angle parameter (θ) can be used to improve the noise detection following the intuition that it is improbable for a moving object to take successive turns with small angles, and a random uniform jitter (ρ) to avoid the MAD breakdown point.

As the trajectory sampling rate is assumed to be nearly constant while the object is moving and locations are not recorded while the object is stopped, the problem can be summarized as searching for outliers into time series as they are expected to have relevant gaps in time that characterize periods of stop or noise.

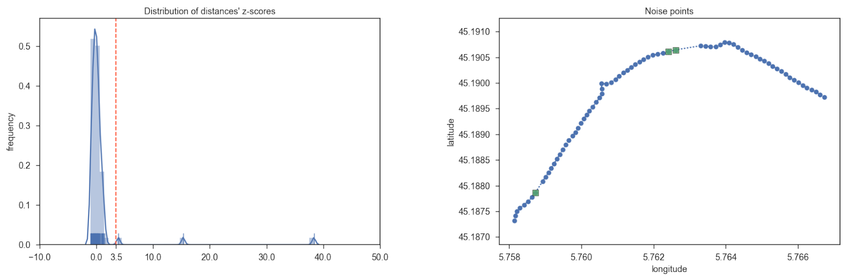
In order to better explain the MSN algorithm, we consider the example trajectory of Figure 1 with the following parameters: $\epsilon_s = \epsilon_v = 3.5$, $\epsilon_t = 5.0$, $\theta = 45$. It is important to notice that we have used the recommended threshold of 3.5 for both distance (ϵ_s) and speed (ϵ_v) parameters. However, for the duration threshold (ϵ_t), we have achieved better results when we increased it to 5.0 as some slow walking segments were being misclassified as stops.

The first part of MSN identifies potential noisy points. This classification, shown in lines 2-8, identifies points with relatively long distances. In the example (Figure 5), three points are identified in this case. They have distances of about 17, 28, and 55 meters, while the median distance of all pairs of sequential trajectory points is 11 meters.

The second step of noise classification consists in verifying the turning angles (lines 9-15). We account for the fact that a single sharp angle in a trajectory may represent a movement of "turning back", while two consecutive sharp angles is less likely to happen and can be considered as a potential noisy segment. This case is not present in the example trajectory as there is no group of points as vertices of angles of less than 45 degrees.

Algorithm 1 Move-Stop-Noise classification algorithm

```
procedure MoveStopNoise(S_τ, T_τ, V_τ, A_τ, ε_s, ε_t, ε_v, θ, ρ)
    distance_outliers = [ ]
    M_s = ModifiedZScore(S_τ, MAD_s, s̃)                    ▷ Equation 2
    for i = 0 to length(M_s) do:
        if M_s[i] > ε_s then                                ▷ Identifying long distances
            Append i to distance_outliers
        end if
    end for
    direction_outliers = [ ]
    for i = 0 to length(A_τ) do:
        if A_τ[i] < θ and A_τ[i + 1] < θ then              ▷ Identifying sharp turning angles
            Append i and i + 1 to direction_outliers
            i++
        end if
    end for
    noise_indexes = distance_outliers + direction_outliers
    clean_indexes = τ - τ[noise_indexes]
    τ = τ[clean_indexes]                                    ▷ Removing noisy points
    T_τ = T_τ + ρ                                           ▷ Adding small random uniform noise
    duration_outliers = [ ]
    M_t = ModifiedZScore(T_τ, MAD_t, t̃)
    for i = 0 to length(M_t) do:
        if M_t[i] > ε_t then                                ▷ Identifying long durations
            Append i to duration_outliers
        end if
    end for
    V_τ = ln V_τ                                            ▷ Natural log of speed
    speed_outliers = [ ]
    M_v = ModifiedZScore(V_τ, MAD_v, ṽ)
    for i = 0 to length(M_v) do:
        if M_v[i] < −ε_v then                               ▷ Identifying slow speeds
            Append i to speed_outliers
        end if
    end for
    stop_indexes = duration_outliers ⋂ speed_outliers
    move_indexes = clean_indexes − stop_indexes
    return move_indexes, stop_indexes, noise_indexes
end procedure
```



(a) (b)

Figure 2. The density plot of distances in (a) and the three trajectory points with long distances in (b)

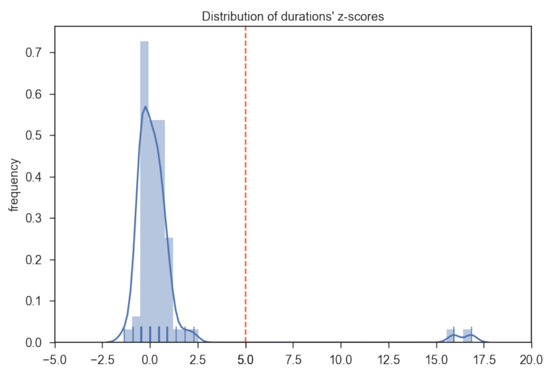


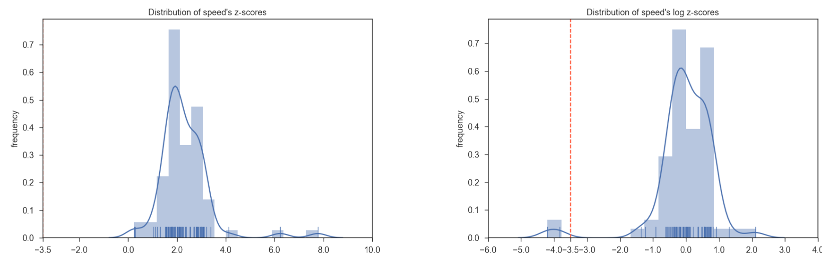
Figure 3. Distribution plot of slightly "jittered" duration between points with two outliers.

Once potential noise is identified, the second part of our method consists in labeling potential stops. Before, the noise points are removed for the further analysis.

Lines 19-26 contains the code designed to identify long duration gaps. We have observed that the time series of duration between points may contain repeated values in more than 50% of the data. In these cases, the MAD is equal to zero (Equation 1), which causes a division by zero in the modified z-score (Equation 2). To avoid this, we add a small amount of random uniform noise to the duration series (line 19). The value to be added is randomly selected from the interval [-ρ, ρ). As a default, ρ is set to 0.5.

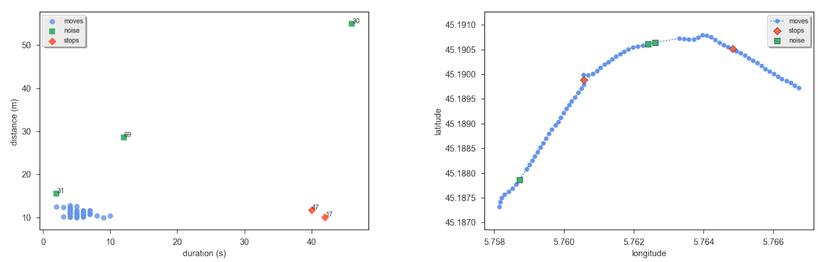
Then, the modified z-score is applied to find duration gaps. However, we have set the modified z-score threshold to 5 in order to avoid false positives. Figure 3 shows the distribution of durations for the example trajectory. Two long durations with 40 and 42 seconds are identified, while the median duration for the trajectory is 5 seconds.

The complement of stop identification (lines 27-34) concerns the analysis of speed time series. The Outlier Labeling Rule presented in Section 4 should be applied to approximately normally distributed datasets. However, for the trajectories considered in this



(a) Original speed (b) Natural logarithm transformation

Figure 4. Difference of speed data before and after natural logarithmic transformation. Also, the outlier threshold is shown in (b)



(a) (b)

Figure 5. Final classification of the MSN algorithm for the example trajectory. (a) Outliers identified as noise and stops. Normal points are considered as periods of movement. (b) The outliers marked in the trajectory geometry.

work, speed data has demonstrated to be positively skewed in general. In order to normalize speed data, the natural logarithm was applied to restore symmetry. Figure 4 shows the importance of this transformation to finding slow speed outliers. In Figure 4(b), it is possible to see that two points have speeds relatively slower ($0.23m/s$ and $0.29m/s$) while the median speed value for the example trajectory is $2.1m/s$.

Finally, we classify points that present both slow speed and long durations as stops. Figure 5 shows all points with their classification as either move, stop, or noise. According to our algorithm, points located at the lower right corner are stops.

6. Evaluation

Due to the assumptions about trajectory sampling strategies, it is difficult to make a comparison with related work by running all algorithms with the same set of trajectories. In this evaluation, we focus on analyzing the theoretical performance of other stop classification methods and we highlight the main differences from our work.

Table 2 shows a general comparison of the main algorithms for stop detection in the literature. As advantages of our method, we can point out the independence of external data, the usage of characteristics that can be completely extracted from the trajectory

Table 2. Comparison of stop detection algorithms

	Parameters	Noise Handling	Spatial Filter Support	External Data Independence
SMoT (Alvares et al. 2007)	Polygons, minimum stop duration for each polygon	No	No	No
CB-SMoT (Palma et al. 2008)	Polygons, area, minimum stop duration	No	No	No
DB-SMoT (Rocha et al. 2010)	Minimum direction change (degrees), minimum duration (hours), maximum tolerance (number of points)	No	No	Yes
Velocity-based trajectory structure (Yan et al. 2010)	Minimum stop duration, object speed threshold coefficient, cell speed threshold coefficient	No	Yes	No
CandidateStops (Nogueira et al. 2014)	Minimum speed (m/s)	No	Yes	Yes
SMoT+ (Moreno et al. 2014)	Polygons, minimum duration for each polygon, sites hierarchy	No	No	No
PIE (de Graaff et al. 2016)	Polygons, maximum inaccuracy (meters), minimum staypoint distance (meters), minimum staypoint time (seconds), minimum direction change (degrees), maximum projection distance (meters)	Yes	No	No
MSN (this work)	Distance outlier threshold, duration outlier threshold, speed outlier threshold, minimum direction change (degrees)	Yes	Yes	Yes

points, the robustness of statistic methods involved, and the handling of noise.

By not relying on the polygons of the underlying geography, our method is adequate to trajectories that are not in a constrained space, being able to identify stops also in free space. Also, apart from the minimum turning angle in degrees, an important aspect of MSN is that the other threshold parameters are not based on any metric quantities, e.g. distance in meters or duration in seconds.

A drawback of MSN is the fact that it relies on the comparison of data points relatively to the rest of the dataset. Therefore, in order to identify a large time gap correctly, it is necessary to the majority of other time gaps to have a short duration, which is not surprising because we base our method on outlier detection for approximately normally distributed data. However, if the trajectory contains a large quantity of noise episodes, the method may fail in recognizing stops. This can be avoided by a preprocessing step to assess the trajectory's level of noise before applying the MSN algorithm. Then, it could be possible to alleviate the noise by some smoothing method, e.g. interpolation.

The MSN method can be implemented in $\mathcal{O}(n+m)$ considering a raw trajectory with n points before noise removal and m points after the noise identification step. In the worst case, $n = m$ (no points are discarded in the noise classification phase). Thus, the algorithm's complexity is $\mathcal{O}(2n)$. It is important to notice that, for the sake of clarity, we have not shown the most concise and efficient implementation of MSN in Algorithm 1, but it could be easily summarized into a single *for* loop.

7. Conclusion

We have proposed in this paper a new algorithm for detecting episodes of movement, stop, and noise in trajectories called MSN. This method is tailored for trajectories that have been sampled at irregular intervals of time or have been preprocessed to eliminate redundant points at near locations. This particular characteristic of some datasets violates a basic assumption made by state-of-the-art methods, which rely on clustering nearby points, and have motivated our work to fill this gap.

The MSN method has also been designed to be independent of external data (e.g. the underlying geographic features), which renders it as a viable option for trajectories recorded in free-space or lacking contextual data. Moreover, the main parameters of MSN are expressed in no particular system of measurement, i.e. there is no need for defining hard thresholds such as specifying that each stop has to have a duration equal or greater than 10 seconds, for instance. Conversely, the parameters used in our method are informed as absolute numbers as proposed by a robust outlier detection method that can be adapted if needed by the application. This is an important aspect that our work offer for advancing the spatiotemporal analysis field in the area of stop detection methods.

It can be envisaged as future work the application of other algorithms, notably supervised learning ones, as the algorithm proposed in this paper takes advantage only of statistical properties of individual trajectories. Training data is important in order to apply a supervised approach. This implies a manually annotated trajectory dataset with known labels. Therefore, a tool to annotate trajectories with stops and moves can be an interesting development. This may improve results by specializing the algorithms for heterogeneous scenarios where different devices capture positional data using their own sampling strategies and post-processing procedures. Moreover, a labeled dataset would be useful for evaluating the efficiency and accuracy of MSN.

Acknowledgements

The authors would like to thank the French Ministry of Higher Education and Research (Ministère de l'Enseignement Supérieur et de la Recherche de la France – MESR) and the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq). Rossana M. C. Andrade has a researcher scholarship "DT Level 2" sponsored by CNPq.

References

Alewijnse, S. P. A. et al. (2014). "A framework for trajectory segmentation by stable criteria". In: *International Conference on Advances in Geographic Information Systems*. SIGSPATIAL'14. New York, New York, USA: ACM Press, pp. 351–360. DOI: 10.1145/2666310.2666415.

Alvares, L. O. et al. (2007). "A model for enriching trajectories with semantic geographical information". In: *International Conference on Advances in Geographic Information Systems*. SIGSPATIAL'07. New York, New York, USA: ACM Press. DOI: 10.1145/1341012.1341041.

Andrienko, N. et al. (2008). "Basic Concepts of Movement Data". In: *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. Ed. by Fosca Giannotti and Dino Pedreschi. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 15–38. DOI: 10.1007/978-3-540-75177-9_2.

Buchin, M. et al. (2011). "Segmenting trajectories: A framework and algorithms using spatiotemporal criteria". In: *Journal of Spatial Information Science* 3, pp. 33–63. DOI: 10.5311/JOSIS.2011.3.66.

de Graaff, V, R. A. de By, and M. van Keulen (2016). "Automated Semantic Trajectory Annotation with Indoor Point-of-interest Visits in Urban Areas". In: *31st Annual ACM Symposium on Applied Computing*. SAC'16. Pisa, Italy: ACM, pp. 552–559. DOI: 10.1145/2851613.2851709.

Gorard, S. (2005). "Revisiting a 90-year-old Debate: The Advantages of the mean deviation". In: *British Journal of Educational Studies* 53.4, pp. 417–430. DOI: 10.1111/j.1467-8527.2005.00304.x.

Huber, P. J. and E. M. Ronchetti (2009). *Robust Statistics*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc., p. 380. DOI: 10.1002/9780470434697.

Iglewicz, B. and D. C. Hoaglin (1993). "Volume 16: How to Detect and Handle Outliers". In: *The ASQC Basic References in Quality Control: Statistical Techniques*. Ed. by Edward F. Mykytka. ASQC/Quality Press, p. 87.

Moreno, B. et al. (2010). "Looking Inside the Stops of Trajectories of Moving Objects". In: *XI Brazilian Symposium on Geoinformatics*. GEOINFO'10. Campos do Jordão, São Paulo, Brazil, pp. 9–20.

Moreno, F. et al. (2014). "SMOT+: Extending the SMOT algorithm for discovering stops in nested sites". In: *Computing and Informatics* 33.2, pp. 327–342.

NIST/SEMATECH (2012). *NIST/SEMATECH e-Handbook of Statistical Methods*. URL: http://www.itl.nist.gov/div898/handbook/ (visited on 11/10/2017).

Nogueira, T. P., R. B. Braga, and H. Martin (2014). "An Ontology-Based Approach to Represent Trajectory Characteristics". In: *5th International Conference on Computing for Geospatial Research and Application*. COM.Geo'14. Washington, DC, USA, pp. 102–107. DOI: 10.1109/COM.Geo.2014.22.

Palma, A. T. et al. (2008). "A Clustering-based Approach for Discovering Interesting Places in Trajectories". In: *23rd Annual ACM Symposium on Applied Computing*. SAC'08. Fortaleza, Brazil: ACM, pp. 863–868. DOI: 10.1145/1363686.1363886.

Rocha, J. A. M. R. et al. (2010). "DB-SMoT: A direction-based spatio-temporal clustering method". In: *5th IEEE International Conference Intelligent Systems*. IS'10. IEEE, pp. 114–119. DOI: 10.1109/IS.2010.5548396.

Rousseeuw, P. J. and M. Hubert (2011). "Robust statistics for outlier detection". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1.1, pp. 73–79. DOI: 10.1002/widm.2.

Spaccapietra, S. et al. (2008). "A conceptual view on trajectories". In: *Data & Knowledge Engineering* 65.1, pp. 126–146. ISSN: 0169023X. DOI: 10.1016/j.datak.2007.10.008.

Yan, Z. et al. (2010). "A Hybrid Model and Computing Platform for Spatio-semantic Trajectories". In: *7th Extended Semantic Web Conference*. ESWC'10. Springer Berlin Heidelberg, pp. 60–75. DOI: 10.1007/978-3-642-13486-9_5.