

Issues of Computer Mouse Data for Related Applications

Peter KRÁTKY*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
peter.kratky@stuba.sk*

Mouse data representing cursor movement or button clicks are often employed in applications such as user interface evaluation methods, implicit feedback gathering methods and recently even mouse-based biometric systems. Especially, the last mentioned application requires high quality data in order to get desired performance. This paper provides an overview of issues and recommendations for researchers who want to study and work with computer mouse data.

Mouse events provided by the physical device are mouse *movement*, mouse *button up/down* and mouse *wheel scrolling*. Most basically, each event is described by time in milliseconds and coordinates (X and Y) of the cursor on the screen. The first issue that should be considered by a researcher is log size. Each event is valuable piece of data and should not be missed. Especially, mouse movements can be very frequent (approximately each 8 milliseconds). Thus, we analyzed how large a storage space is necessary. Based on four experiments, we summarized average number of events per minute in an e-learning system, tourist portal, e-shop and a game with values 342, 496, 673 and 2770, respectively. Consequently, we assume that every minute of user interaction with web page requires approximately 20 kilobytes of disc space.

Another issue of raw data is the precision of mouse movement coordinates. Positions can occur very often, but the coordinates are rounded to integers. One solution could be decreasing number of events per second. However, fine-grained information is lost in this case. So, we propose smoothing of the recorded path and aligning positions to this new path. Another benefit from this approach is that uneven spaces between points could be covered with new points what solves some problems in processing stage. To smooth the path we calculate an approximation curve fitting the measured points. Various interpolation methods, such as natural cubic splines or Bezier curves, use measured points as control points which are crossed by the calculated curve. A bit different method is B-spline which does not cross the control points, what is desired feature as those are assumed to be imprecise.

* Supervisor: Daniela Chudá, Institute of Informatics and Software Engineering

Raw data need to be translated into something meaningful. There are a lot of features that could characterize user's behavior, such as movement velocity, curvature, etc. Especially mouse movement data hold very small amount of information rather than sequence of such points do. A window of N movement events could be used to calculate values of features. Existing (research) implementations mostly aggregate events into meaningful actions in order to represent intent of users. Usually, *strokes* or *curves* are actions composed of mouse movement event sequence separated by a pause or click [10, 8]. In [1] there are four actions recommended to study – point and click, movement only, drag and drop, silence (no activity).

Aggregated data often include weird values that should be omitted. One should be careful when discarding values that not easily distinguished what is real and what is not real. The values of mouse features follow asymmetric distributions, thus proper outlier detection method should be used. For example, Tukey's method used in boxplots is suitable, which discards values out of the range 3 IQRs (inter quartile range) below the first quartile and 3 IQRs above the third quartile [9].

When conducting this kind of research which includes monitoring users, it is quite obvious that the users might feel uncomfortable. We conducted a short survey in which two questions about the privacy issues were asked: 1) *Imagine that you anonymously browse a web page (e.g. e-shop). It records all cursor movement for further analysis. Would you worry about your privacy?* 2.) *Would you leave such a web page if you knew it records your cursor movement?* There were 78 participants of different nations, different age groups, both males and females in the study. Completed questionnaires show that 63 participants would worry about their privacy and 58 would leave the page if they knew their cursor is being monitored. We could definitely conclude that this is a sensitive topic for common web users who could feel kind of distress of something unknown that has a potential to be misused.

Amended version was published in Proc. of the 11th Student Research Conference in Informatics and Information Technologies (IIT.SRC 2015), STU Bratislava, 60-61.

Acknowledgement: This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG 1/0646/15.

References

- [1] Ahmed, A.A.E., Traore, I.: A New Biometric Technology Based on Mouse Dynamics. In: *IEEE Transactions on Dependable and Secure Computing*, (2007), pp. 165–179.
- [2] Schulz, D.A.: Mouse curve biometrics. In: *2006 Biometrics Symposium: Special Session on Research at The Biometric Consortium Conference*, (2006), pp. 1–6.
- [3] Seo, S.: A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets. M.S. University of Pittsburgh, (2006).
- [4] Zheng, N., Paloski, A., Wang, H.: An efficient user verification system via mouse movements. In: *Proceedings of the 18th ACM conference on Computer and communications security - CCS '11*, (2011), pp. 139–150.