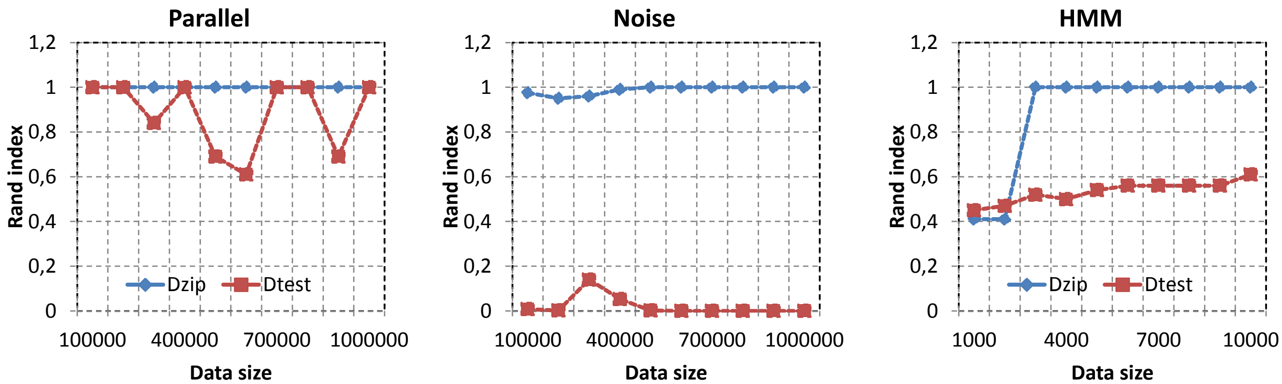

Figure 2: Rand index measures the similarity (higher is better) between the decompositions by the algorithms and the ground-truths. Rand index is equal to 1 if it is a perfect decomposition.

	MasterPortal	Msnbc	Machine	Parallel	Noise	HMM
Dzip	28	38	131387	8	192	1
Dtest	159	500	201385	138	25130	1

Figure 3: Running time in seconds of two algorithms. Dzip is about an order of magnitude faster than Dtest.

was shown to be more effective than the state of the art method based on statistical hypothesis testing. There are various directions to extend the paper for the future works. At the moment, we assume that each independent process must produce a disjoint subset of events. In practice, the case that independent processes produce overlapping subset of events is not rare. Extending the work to this more general case can be considered as an interesting future work.

9. ACKNOWLEDGEMENTS

The work is part of the project *Mining Complex Patterns in Stream* (COMPASS) supported by the *Netherlands Organisation for Scientific Research* (NWO).

10. REFERENCES

[1] Kira Radinsky, Eric Horvitz: Mining the web to predict future events. WSDM 2013: 255-264
[2] Julia Kiseleva, Hoang Thanh Lam, Toon Calders and Mykola Pechenizkiy: Discovery temporal hidden contexts in web sessions for user trail prediction TempWeb workshop at WWW 2013.
[3] Heikki Mannila, Dmitry Rusakov: Decomposition of event sequences into independent components. SDM 2001
[4] Thomas Cover, Joy Thomas: Elements of information theory. Wiley and Son, second edition 2006.
[5] Peter Grünwald: The minimum description length principle. MIT press, 2007.
[6] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Journal of Neural Network 2000.
[7] Michael Mampaey, Jilles Vreeken: Summarizing categorical data by clustering attributes. Data Min. Knowl. Discov. 2013
[8] Rudi Cilibrasi, Paul M. B. Vitányi: Clustering by compression. IEEE Transactions on Information Theory 2005
[9] Eamonn J. Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana, Li Wei, Sang-Hee Lee, John Handley: Compression-based data mining of sequential data. Data Min. Knowl. Discov. 2007
[10] Jilles Vreeken, Matthijs van Leeuwen, Arno Siebes: Krimp: mining itemsets that compress. Data Min. Knowl. Discov. 2011
[11] Hoang Thanh Lam, Fabian Moerchen, Dmitriy Fradkin, Toon Calders: Mining Compressing Sequential Patterns. SDM 2012: 319-330
[12] Hardy, G. H. (1992). Divergent Series. Providence: American Mathematical Society. ISBN 978-0-8218-2649-2.
[13] W. M. Rand (1971). Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66 (336).
[14] van der Aalst, W. M. P., Weijters, T. and Maruster, L. (2004). Workflow Mining: Discovering Process Models from Event Logs.. IEEE Trans. Knowl. Data Eng., 16, 1128-1142.