A combinatorial view on BWT variants

Marinella Sciortino

University of Palermo, Italy,
`marinella.sciortino@community.unipa.it`

Abstract. The Burrows-Wheeler Transform (BWT) is a reversible transformation that was introduced in 1994 in the field of Data Compression and it has also become a fundamental tool for self-indexing data structures. It is a context-dependent transformation that produces a permutation of the input text that is likely to have runs of equal letters (clusters) longer than the ones in the original text. Combinatorial properties of BWT and its extension to a multisets of strings [5] have been used in several other applications, especially in Bioinformatics [7,4]. In [6] a complexity measure that counts the number of equal-letter runs produced by the BWT is introduced, by exploring how the number of clusters of the BWT output varies depending on the number of clusters of the input.

Over the years other variants of the BWT have been proposed, without however obtaining transformations entirely capable of playing the same variegated role as the BWT. Very recently, a whole new class of transformations, called *local ordering trasformation*, has been introduced [3]. They have all the same prerogatives as BWT, i.e. they can be computed and inverted in linear time, they produce provably highly compressible output, and they support linear time pattern search directly on the transformed text. Such a class is a special case of a more general family of transformations based on context adaptive alphabet orderings. This more general class includes also the Alternating BWT (ABWT), another invertible transformation recently introduced in connection with a generalization of Lyndon words. Algorithmic and combinatorial issues on ABWT have been investigated in [1,2].

In the talk some combinatorial aspects of *BWT* are focused. Moreover, an overview of the above mentioned transformations is presented, by focusing on some distinctive algorithmic and combinatorial features that could represent effective tools to investigate and handle the structure of the input text.

References

1. R. Giancarlo, G. Manzini, A. Restivo, G. Rosone, and M. Sciortino. The Alternating BWT: an algorithmic perspective. *Theoret. Comput. Sci.* In Press.
2. R. Giancarlo, G. Manzini, A. Restivo, G. Rosone, and M. Sciortino. Block Sorting-Based Transformations on Words: Beyond the Magic BWT. In *DLT*, pages 1–17. Springer International Publishing, 2018.
3. R. Giancarlo, G. Manzini, G. Rosone, and M. Sciortino. A New Class of Searchable and Provably Highly Compressible String Transformations. In *CPM 2019*, volume 12 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1–12, 2019.

4. F. A. Louza, G. P. Telles, S. Gog, and L. Zhao. Computing Burrows-Wheeler Similarity Distributions for String Collections. In *SPIRE 2018*, volume 11147 of *Lecture Notes in Computer Science*, pages 285–296. Springer, 2018.
5. S. Mantaci, A. Restivo, G. Rosone, and M. Sciortino. An extension of the Burrows-Wheeler Transform. *Theor. Comput. Sci.*, 387(3):298–312, 2007.
6. S. Mantaci, A. Restivo, G. Rosone, M. Sciortino, and L. Versari. Measuring the clustering effect of BWT via RLE. *Theor. Comput. Sci.*, 698:79–87, 2017.
7. N. Prezza, N. Pisanti, M. Sciortino, and G. Rosone. SNPs detection by eBWT positional clustering. *Algorithms for Molecular Biology*, 14(1):3:1–3:13, 2019.