Recent developments in pattern mining

Citation for published version (APA):

Calders, T. G. K. (2012). Recent developments in pattern mining. In J-G. Ganascia, P. Lenca, & J-M. Petit (Eds.), *Discovery Science (15th International Conference, DS 2012, Lyon, France, October 29-31, 2012. Proceedings)* (pp. 2-2). (Lecture Notes in Computer Science; Vol. 7569). Springer. https://doi.org/10.1007/978-3-642-33492-4_2

DOI:

[10.1007/978-3-642-33492-4_2](https://doi.org/10.1007/978-3-642-33492-4_2)

Document status and date:

Published: 01/01/2012

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Recent Developments in Pattern Mining

Toon Calders

TU Eindhoven
The Netherlands
`t.calders@tue.nl`

Abstract. Pattern Mining is one of the most researched topics in the data mining community. Literally hundreds of algorithms for efficiently enumerating all frequent itemsets have been proposed. These exhaustive algorithms, however, all suffer from the pattern explosion problem. Depending on the minimal support threshold, even for moderately sized databases, millions of patterns may be generated. Although this problem is by now well recognized in the pattern mining community, it has not yet been solved satisfactorily. In my talk I will give an overview of the different approaches that have been proposed to alleviate this problem. As a first step, constraint-based mining and condensed representations such as the closed itemsets and the non-derivable itemsets were introduced. These methods, however, still produce too many and redundant results. More recently, promising methods based upon the minimal description length principle, information theory, and statistical models have been introduced. We show the respective advantages and disadvantages of these approaches and their connections, and illustrate their usefulness on real life data. After this overview we move from itemsets to more complex patterns, such as sequences and graphs. Even though these extensions seem trivial at first, they turn out to be quite challenging. I will end my talk with an overview of what I consider to be important open questions in this fascinating research area.