



Universiteit
Leiden
The Netherlands

Pattern mining for label ranking

Pinho Rebelo de Sá, C.F.

Citation

Pinho Rebelo de Sá, C. F. (2016, December 16). *Pattern mining for label ranking*. Retrieved from <https://hdl.handle.net/1887/44953>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/44953>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/44953> holds various files of this Leiden University dissertation.

Author: Pinho Rebelo de Sá, C.F.

Title: Pattern mining for label ranking

Issue Date: 2016-12-16

Pattern Mining for Label Ranking

by

Cláudio Frederico Pinho Rebelo de Sá

Pattern Mining for Label Ranking

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op vrijdag 16 december 2016
klokke 11.15 uur

door

Cláudio Frederico Pinho Rebelo de Sá

geboren te Porto, Portugal
in 1984

Promotiecommissie

Promotor: prof. dr. J. N. Kok (Universiteit Leiden)
Co-promotor: dr. C. M. Soares (Universidade do Porto)
Co-promotor: dr. A. J. Knobbe (Universiteit Leiden)
Overige leden: prof. dr. T. H. W. Bäck (Universiteit Leiden)
prof. dr. H. J. van den Herik (Universiteit Leiden)
dr. P. Kralj Novak (Jožef Stefan Institute)
dr. M. Atzmüller (Universität Kassel)

Front and back cover patterns by Vecree.com, available under a Creative Commons License (CC BY-SA 2.0).

Printing: Ridderprint BV, the Netherlands

Dedicado à minha mãe,
por sempre acreditar em mim

Contents

1	Introduction	1
1.1	Preference Learning	1
1.2	Label Ranking	3
1.2.1	Definition	4
1.2.2	Evaluation	6
1.2.3	Reduction techniques	8
1.2.4	Direct approaches	9
1.3	Contributions of this thesis	10
1.3.1	Label Ranking Association Rules	11
1.3.2	Discretization	12
1.3.3	Tree-based models	13
1.3.4	Descriptive mining for label ranking	14
1.3.5	Label Ranking Data	15
1.4	Thesis outline	17
2	Preference Rules	19
2.1	Introduction	20
2.2	Association Rule Mining	21
2.2.1	Interest measures	22
2.2.2	APRIORI Algorithm	22
2.2.3	Pruning	23
2.3	Label Ranking	24
2.3.1	Methods	25
2.3.2	Evaluation	27
2.4	Label Ranking Association Rules	28
2.4.1	Interestingness measures in Label Ranking	29
2.4.2	Generation of LRAR	31
2.4.3	Prediction	32
2.4.4	Parameter tuning	34
2.5	Pairwise Association Rules	34

2.6	Experimental Results	36
2.6.1	Datasets	36
2.6.2	Experimental setup	37
2.6.3	Results with LRAR	38
2.6.4	Results with PAR	45
2.7	Conclusions	47
3	Entropy-based discretization methods for ranking data	49
3.1	Introduction	50
3.2	Label Ranking	51
3.2.1	Association Rules for Label Ranking	52
3.2.2	Naive Bayes for Label Ranking	53
3.3	Discretization	54
3.3.1	Entropy-based methods	56
3.4	Discretization for Label Ranking	57
3.4.1	Adapting the concept of entropy for rankings	58
3.5	Experimental Results	63
3.5.1	Sensitivity to the θ_{disc} parameter	64
3.5.2	Results on Artificial Datasets	65
3.5.3	Results on Benchmark Datasets	74
3.6	Conclusions	76
4	Label Ranking Forests	79
4.1	Introduction	80
4.2	Label Ranking	81
4.2.1	Formalization	81
4.2.2	Ranking Trees	82
4.2.3	Entropy Ranking Trees	85
4.3	Random Forests	86
4.3.1	Label Ranking Forests	87
4.4	Empirical Study	88
4.4.1	Experimental setup	88
4.4.2	Results with Label Ranking Trees	89
4.4.3	Results with Label Ranking Forests	92
4.5	Conclusions	93
5	Exceptional Preferences Mining	97
5.1	Introduction	98
5.1.1	Main Contributions	98
5.2	Label Ranking	99
5.3	Subgroup Discovery and Exceptional Model Mining	100

5.3.1	Traversing the Search Space	102
5.4	Exceptional Preferences Mining	102
5.4.1	Preference Matrix	103
5.4.2	Characterizing Exceptional Subgroups	105
5.5	Experiments	108
5.5.1	Datasets	108
5.5.2	Results	110
5.6	Conclusions	113
6	Permutation Tests for Label Ranking	117
6.1	Introduction	118
6.2	Label Ranking	119
6.2.1	IB-PL	120
6.2.2	APRIORI-LR	120
6.2.3	Datasets	120
6.3	Swap Randomization	121
6.4	Validating ranking data with permutation tests	122
6.4.1	Random permutation of rankings	122
6.4.2	Random permutation of labels	123
6.5	Experiments	123
6.5.1	Ranking permutations	125
6.5.2	Labelwise permutations	126
6.6	Conclusions	129
7	Conclusions	131
	Bibliography	148
	Nederlandse Samenvatting	149
	English Summary	153
	Resumo	155
	List of publications	157
	Acknowledgments	159
	Curriculum Vitae	161

