

ห้องสมุดเจ้าพระยา สำนักงานคณะกรรมการวิจัยแห่งชาติ



E47229

การหาอุปแบบความถี่น้ำฝนของเขื่อนกั้นน้ำจากอุทยานแห่งชาติเขื่อนการ

นายโกเมศ ชัยพันธ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

อธิการบดีของจุฬาลงกรณ์มหาวิทยาลัย

600254148

E47229

การหารูปแบบความถี่สม่ำเสมอเคอันดับแรกจากฐานข้อมูลรายการ

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการการวิจัยแห่งชาติ



นายโกเมศ อัมพวัน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2553

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย



MINING TOP-K REGULAR-FREQUENT ITEMSETS FROM TRANSACTIONAL DATABASE

Mr. Komate Amphawan

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

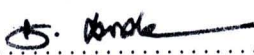
Chulalongkorn University

Academic Year 2010

Copyright of Chulalongkorn University

Thesis Title MINING TOP-K REGULAR-FREQUENT ITEMSETS FROM
TRANSACTIONAL DATABASE
By Mr. Komate Amphawan
Field of Study Computer Engineering
Thesis Advisor Assistant Professor Athasit Surarerks, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of
the Requirements for the Doctoral Degree

.....  Dean of the Faculty of Engineering
(Associate Professor Boonsom Lerdhirunwong, Dr.Ing.)

THESIS COMMITTEE

.....  Chairman
(Associate Professor Wanchai Rivepiboon, Ph.D.)

.....  Thesis Advisor
(Assistant Professor Athasit Surarerks, Ph.D.)

.....  Examiner
(Assistant Professor Pizzanu Kanongchaiyos, Ph.D.)

.....  External Examiner
(Assistant Professor Krisana Chinnasarn, Ph.D.)

.....  External Examiner
(Assistant Professor Arnon Rungsawang, Ph.D.)

โกเมศ อัมพวัน: การหารูปแบบความถี่สม่ำเสมอเคอันดับแรกจากฐานข้อมูลรายการ. (MIN-ING TOP-K REGULAR-FREQUENT ITEMSETS FROM TRANSACTIONAL DATABASE) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผู้ช่วยศาสตราจารย์ อรรถสิทธิ์ สุรฤกษ์, 197 หน้า.

E47229

การหาความสัมพันธ์ภายใต้ค่าสนับสนุนและค่าความเชื่อมั่นที่ผู้ใช้ระบุเป็นงานสำคัญของกลุ่มวิจัยด้านการทำเหมืองข้อมูล อย่างไรก็ตามความถี่ของการเกิดข้อมูลในรูปแบบต่างๆอาจไม่เพียงพอที่จะเป็นเงื่อนไขของการหารูปแบบที่มีความหมาย ดังนั้นลักษณะการเกิดของรูปแบบที่มีความสัมพันธ์กันจึงเป็นประเด็นสำคัญของงานในหลายๆด้าน รูปแบบหนึ่งของการเกิดความสัมพันธ์ของข้อมูลก็คือการเกิดขึ้นอย่างสม่ำเสมอ นั่นคือข้อมูลจะปรากฏขึ้นห่างกันเป็นช่วงๆในระยะห่างที่ผู้ใช้ให้ความสำคัญ ซึ่งในการหาข้อมูลที่มีการเกิดขึ้นในลักษณะดังกล่าวจะต้องมีการกำหนดค่าสนับสนุนเพื่อใช้ในการเลือกกรองความสม่ำเสมอของข้อมูลที่เกิดขึ้น แต่ในทางปฏิบัติแล้วการกำหนดค่าขีดแบ่งที่เหมาะสมนี้ทำได้ยาก เนื่องจากถ้ากำหนดค่าขีดแบ่งนี้น้อยเกินไปจำนวนผลลัพธ์ที่ได้ก็จะมีปริมาณมากและในทางกลับกันถ้ากำหนดค่าขีดแบ่งมากผลลัพธ์ที่ได้ก็จะมีจำนวนน้อยหรืออาจจะไม่พบคำตอบที่สอดคล้องกับค่าขีดแบ่งนี้ ดังนั้นแทนที่จะกำหนดค่าขีดแบ่งนี้การกำหนดจำนวนผลลัพธ์ที่ต้องการจะเป็นการเหมาะสมกว่าจากการสำรวจงานวิจัยที่มีอยู่ในปัจจุบันพบว่าไม่มีวิธีใดที่ให้ผู้ระบุจำนวนการเกิดข้อมูลแบบสม่ำเสมอที่ผู้ใช้ต้องการ ดังนั้นวิทยานิพนธ์นี้จึงนำเสนอการใช้จำนวนผลลัพธ์เป็นเป้าหมายในการค้นหาข้อมูลที่มีการเกิดอย่างสม่ำเสมอ นั่นคือการหาผลลัพธ์ที่มีค่าสนับสนุนสูงที่สุดเคอันดับแรก ซึ่งวิธีที่วิทยานิพนธ์นี้แนะนำมีดังนี้ (1) วิธีการหาคำตอบโดยอ่านข้อมูลจากฐานข้อมูลเพียงครั้งเดียว (2) การใช้การค้นหาแบบเลือกทางดีที่สุดเพื่อลดปริมาณการค้นหา (3) การแบ่งข้อมูลและการประมาณค่าสนับสนุนเพื่อช่วยลดการคำนวณที่ไม่จำเป็นลง (4) การใช้วิธีการแทนข้อมูลแบบใหม่เพื่อลดเวลาและหน่วยความจำที่ใช้ในการประมวลผล ซึ่งจากการวิเคราะห์ประสิทธิภาพในเชิงเวลาและหน่วยความจำของวิธีการที่นำเสนอพบว่า ไม่ว่าจะกำหนดจำนวนของผลลัพธ์มากหรือน้อย ไม่ว่าข้อมูลจะมีการกระจุกหรือกระจายตัว ขั้นตอนวิธีที่นำเสนอนี้ก็สามารถให้ผลลัพธ์ได้อย่างมีประสิทธิภาพ

ภาควิชาวิศวกรรมคอมพิวเตอร์..... ลายมือชื่อนิสิตโกเมศ อัมพวัน.....
สาขาวิชาวิศวกรรมคอมพิวเตอร์..... ลายมือชื่อ.ที่ปรึกษาวิทยานิพนธ์หลัก
ปีการศึกษา 2553

4871857021: MAJOR COMPUTER ENGINEERING

KEYWORDS: DATA MINING / ASSOCIATION RULES / TOP-K SIGNIFICANT ITEMSETS
/ REGULAR-FREQUENT ITEMSETS / TOP-K REGULAR-FREQUENT ITEMSETS

KOMATE AMPHAWAN : MINING TOP-K REGULAR-FREQUENT ITEMSETS.FROM
TRANSACTIONAL DATABASE. THESIS ADVISOR : Assistant Professor Athasit Surar-
erks, Ph.D., 53 pp.

E47229

Association rule based on support-confident framework is an important task in data mining community. However, the occurrence frequency of a pattern may not be sufficient criterion for mining meaningful patterns. The occurrence behavior can be revealed as an important key in several applications. A pattern is a regular pattern if it regularly occurs in a user-given period (regularity threshold). To mine regular itemsets, a support threshold is used to filter some regular itemsets. However, in practice, it is often difficult for users to provide an appropriate support threshold. Indeed, a too small support threshold could yield a number of regular-frequent itemsets impractically large while a too large threshold could yield very few or no regular-frequent itemsets. Therefore, the use of a support threshold tends to produce a large number of regular-frequent itemsets and it could be better to ask for the number of desired results.

Currently, from the deep survey, there is no existing approach permitting users to specify the number of regular-frequent itemsets to be mined. Therefore, a new approach allowing the users to control the number of results (*i.e.* k regular itemsets with the highest supports) is presented. There are several techniques proposed in this dissertation to mine this kind of itemsets: (i) a single-scan approach, (ii) a best-first search strategy used to prune the search space, (iii) the partitioning and estimation techniques assisting in reducing unnecessary computational costs, and (iv) a new concise representation helping to save runtime and memory. From the performance study, the proposed approaches are efficient and scalable in terms of time and memory for small and large values of desired results on sparse and dense datasets.

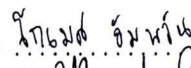
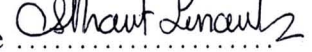
Department: Computer Engineering

Field of Study: Computer Engineering

Academic Year:2010.....

Student's Signature

Advisor's Signature

Acknowledgements

First and foremost, I wish to thank the Higher Education Commission of Thailand who gave me a research scholarship. I would like to express my deep gratitude to my supervisor, Athasit Surarerks, Ph.D., to whom with his guidance, advices and helps me to overcome the difficulties of the process of conducting this dissertation. He patiently and thoroughly guided me for an almost 5-year period from a starting point to reading the chapters of this thesis. I also would like to express my deepest gratitude to my co-supervisor, Philippe Lenca, Ph.D. at Telecom Bretagne, France. His invaluable advice and patience allowed me to undergo and finish the challenging process of my doctoral degree program. I also express my thankfulness to my dissertation committees: Wanchai Rivepiboon, Ph.D., Pizzanu Kanongchaiyos, Ph.D., Krisana Chinnasan, Ph.D. and Arnon Rungsawang, Ph.D., for their advices and guidance to help me focus on my research activities. I am also grateful to Sunisa Rimchareon, Ph.D., and Miss Warisa Sritiratanarak whom always encourage me to do research, give me moral support and help me everything as they can do.

I also would like to thank all of my lovely friends (e.g. Pai, Phae, Te, Yring, Rin, Chan, Aui, Krit, Tuk, Kai, Kik, Camp, Oat, Zeng, Wat, Tar, Trung, Kwan, Pum, etc.) and the graduate students in Computer Engineering department (e.g. P' Chai, Koh, Pook, Pla, Woot, Wut, P' Tom, P' Chang, P' Jim, P' Phueng, P' Chit, P'Vic, Tei, Pae, Pun, Dong, Kai, Puh, Pair, Na, Pong, Dear, Oat(Elite), Oat(CG), Petch(Van), Petch(Dek), P'Keng, Tair, Chris, Tuy, Jumbo, Noot, P' Ae, P' Aoe, Matt, P' Mao, P' Aui, P' Sakorn, P' Yui, Petch(Mhee), P' Dae, P' Jung, J' Nhan, P'Fu, June, Pia, Vit, etc.) whose advice and friendship gave me a lot of encouragement for my studying at Chulalongkorn University.

I would like to thank the Department of Computer Engineering of Chulalongkorn University for giving me the opportunity to pursue a Ph.D. degree in computer engineering. I am also thankful to the support staffs in Computer Engineering Department for always being helpful.

Last but not least, I am very grateful to my family: Father, Mother and my Sister for their love, patience, continuous moral support and encouragement. Without all of these I could not have accomplished my doctoral degree.

Contents

	Page
Abstract (Thai)	iv
Abstract (English)	v
Acknowledgements	vi
Contents	vii
List of Tables	x
List of Figures	xi
Chapter	
I Introduction	1
1.1 Objectives of Study	4
1.2 Scopes of Study	5
1.3 Research Methodology	5
1.4 Organization	5
II Related work	7
2.1 Frequent itemsets mining	7
2.2 Top- k significant itemsets mining	10
2.3 Regular-frequent itemsets mining	12
2.4 Benchmark datasets	13
III Mining top-k regular-frequent itemsets	16
3.1 Top- k regular-frequent itemsets mining	16
3.2 Preliminary of MTKPP	17
3.3 MTKPP: Top- k list structure	17
3.4 MTKPP algorithm	18
3.4.1 MTKPP: Top- k list initialization	18
3.4.2 MTKPP: Top- k mining	18
3.5 Example of MTKPP	20
3.6 Performance evaluation	22
3.6.1 Experimental setup	22
3.6.2 Execution time	22
3.6.3 Memory consumption	23
3.6.4 Scalability test	23
3.7 Summary	24

Chapter	Page
IV TKRIMPE: Top-K Regular-frequent Itemsets Mining using database Parti-	
tioning and support Estimation	48
4.1 Preliminary of TKRIMPE	48
4.2 TKRIMPE: Top- <i>k</i> list structure	49
4.3 Database Partitioning	49
4.4 Support Estimation	52
4.5 TKRIMPE algorithm	54
4.5.1 TKRIMPE: Top- <i>k</i> list initialization	54
4.5.2 TKRIMPE: Top- <i>k</i> mining	56
4.6 Example of TKRIMPE	56
4.7 Complexity analysis	59
4.8 Performance Evaluation	60
4.8.1 Experimental setup	60
4.8.2 Advantages of the database partitioning and the support estimation tech-	
niques applied in TKRIMPE	61
4.8.3 Execution time	62
4.8.4 Memory consumption	62
4.8.5 Scalability test	63
4.9 Summary	64
V TKRIMIT: Top-K Regular-frequent Itemsets Mining based on Interval Tidset	
representation	99
5.1 Preliminary of TKRIMIT	99
5.2 Interval Tidset representation	99
5.3 TKRIMIT: Top- <i>k</i> list structure	102
5.4 TKRIMIT algorithm	102
5.4.1 TKRIMIT: Top- <i>k</i> list initialization	103
5.4.2 TKRIMIT: Top- <i>k</i> mining	106
5.5 Example of TKRIMIT	106
5.6 Complexity analysis	108
5.7 Performance evaluation	109
5.7.1 Experimental setup	109

Chapter	Page
5.7.2 Compactness of using interval tidset representation	110
5.7.3 Execution time	110
5.7.4 Memory consumption	111
5.7.5 Scalability test	112
5.8 Summary	112
VI H-TKRIMP: Hybrid representation on Top-K Regular-frequent Itemsets	
Mining based on database Partitioning	143
6.1 Preliminary of H-TKRIMP	143
6.2 H-TKRIMP: Top- <i>k</i> list structure	143
6.3 Database Partitioning	144
6.4 Hybrid representation	145
6.5 Calculation of Regularity and Support	147
6.6 H-TKRIMP algorithm	149
6.6.1 H-TKRIMP: Top- <i>k</i> initialization	150
6.6.2 H-TKRIMP: Top- <i>k</i> mining	151
6.7 Example of H-TKRIMP	154
6.8 Complexity analysis	156
6.9 Performance evaluation	157
6.9.1 Experimental setup	157
6.9.2 Execution time	158
6.9.3 Memory consumption	159
6.9.4 Scalability test	160
6.10Summary	161
VII Conclusion	185
7.1 Summary of Dissertation	185
7.2 Discussion	187
References	189
Biography	197

List of Tables

Table	Page
2.1 Horizontal representation	8
2.2 Vertical Tidset representation	9
2.3 Datasets classification from (Flouvat et al., 2010)	15
2.4 Database characteristics	15
3.1 A transactional database as a running example of MTKPP	20
4.1 A transactional database as a running example of TKRIMPE	50
5.1 A transactional database as a running example of TKRIMIT	102
6.1 A transactional database as a running example of H-TKRIMP	144

List of Figures

Figure	Page
3.1 MTKPP: Top- k list with hash table	18
3.2 Top- k list initialization	21
3.3 Top- k regular-frequent itemsets	21
3.4 Runtime of MTKPP on <i>accidents</i> ($\sigma_r = 1\%$)	25
3.5 Runtime of MTKPP on <i>accidents</i> ($\sigma_r = 2\%$)	25
3.6 Runtime of MTKPP on <i>accidents</i> ($\sigma_r = 3\%$)	26
3.7 Runtime of MTKPP on <i>chess</i> ($\sigma_r = 2\%$)	26
3.8 Runtime of MTKPP on <i>chess</i> ($\sigma_r = 4\%$)	27
3.9 Runtime of MTKPP on <i>chess</i> ($\sigma_r = 6\%$)	27
3.10 Runtime of MTKPP on <i>connect</i> ($\sigma_r = 1\%$)	28
3.11 Runtime of MTKPP on <i>connect</i> ($\sigma_r = 2\%$)	28
3.12 Runtime of MTKPP on <i>connect</i> ($\sigma_r = 3\%$)	29
3.13 Runtime of MTKPP on <i>mushroom</i> ($\sigma_r = 4\%$)	29
3.14 Runtime of MTKPP on <i>mushroom</i> ($\sigma_r = 6\%$)	30
3.15 Runtime of MTKPP on <i>mushroom</i> ($\sigma_r = 8\%$)	30
3.16 Runtime of MTKPP on <i>pumsb</i> ($\sigma_r = 2\%$)	31
3.17 Runtime of MTKPP on <i>pumsb</i> ($\sigma_r = 4\%$)	31
3.18 Runtime of MTKPP on <i>pumsb</i> ($\sigma_r = 6\%$)	32
3.19 Runtime of MTKPP on <i>pumsb*</i> ($\sigma_r = 1\%$)	32
3.20 Runtime of MTKPP on <i>pumsb*</i> ($\sigma_r = 2\%$)	33
3.21 Runtime of MTKPP on <i>pumsb*</i> ($\sigma_r = 3\%$)	33
3.22 Runtime of MTKPP on <i>BMS-POS</i> ($\sigma_r = 1\%$)	34
3.23 Runtime of MTKPP on <i>BMS-POS</i> ($\sigma_r = 2\%$)	34
3.24 Runtime of MTKPP on <i>BMS-POS</i> ($\sigma_r = 3\%$)	35
3.25 Runtime of MTKPP on <i>retail</i> ($\sigma_r = 6\%$)	35
3.26 Runtime of MTKPP on <i>retail</i> ($\sigma_r = 8\%$)	36
3.27 Runtime of MTKPP on <i>retail</i> ($\sigma_r = 10\%$)	36
3.28 Runtime of MTKPP on <i>T10I4D100K</i> ($\sigma_r = 4\%$)	37
3.29 Runtime of MTKPP on <i>T10I4D100K</i> ($\sigma_r = 6\%$)	37
3.30 Runtime of MTKPP on <i>T10I4D100K</i> ($\sigma_r = 8\%$)	38
3.31 Runtime of MTKPP on <i>T20I6D100K</i> ($\sigma_r = 2\%$)	38
3.32 Runtime of MTKPP on <i>T20I6D100K</i> ($\sigma_r = 4\%$)	39
3.33 Runtime of MTKPP on <i>T20I6D100K</i> ($\sigma_r = 6\%$)	39

Figure	Page
3.34 Runtime of MTKPP on <i>T40I10D100K</i> ($\sigma_r = 2\%$)	40
3.35 Runtime of MTKPP on <i>T40I10D100K</i> ($\sigma_r = 4\%$)	40
3.36 Runtime of MTKPP on <i>T40I10D100K</i> ($\sigma_r = 6\%$)	41
3.37 Memory usage of MTKPP on <i>accidents</i>	41
3.38 Memory usage of MTKPP on <i>chess</i>	42
3.39 Memory usage of MTKPP on <i>connect</i>	42
3.40 Memory usage of MTKPP on <i>mushroom</i>	43
3.41 Memory usage of MTKPP on <i>pumsb</i>	43
3.42 Memory usage of MTKPP on <i>pumsb*</i>	44
3.43 Memory usage of MTKPP on <i>BMS-POS</i>	44
3.44 Memory usage of MTKPP on <i>retail</i>	45
3.45 Memory usage of MTKPP on <i>T10I4D100K</i>	45
3.46 Memory usage of MTKPP on <i>T20I6D100K</i>	46
3.47 Memory usage of MTKPP on <i>T40I10D100K</i>	46
3.48 Scalability of MTKPP ($k : 500, \sigma_r = 6$)	47
3.49 Scalability of MTKPP ($k : 10,000, \sigma_r = 6$)	47
4.1 TKRIMPE: Top- k list with a hash table	49
4.2 Top- k list initialization	58
4.3 Top- k frequent itemsets	58
4.4 The number of early terminated itemsets on <i>accidents</i> dataset	65
4.5 The number of early terminated itemsets on <i>chess</i> dataset	65
4.6 The number of early terminated itemsets on <i>connect</i> dataset	66
4.7 The number of early terminated itemsets on <i>mushroom</i> dataset	66
4.8 The number of early terminated itemsets on <i>pumsb</i> dataset	67
4.9 The number of early terminated itemsets on <i>pumsb*</i> dataset	67
4.10 The number of early terminated itemsets on <i>BMS-POS</i> dataset	68
4.11 The number of early terminated itemsets on <i>retail</i> dataset	68
4.12 The number of early terminated itemsets on <i>T10I4D100K</i> dataset	69
4.13 The number of early terminated itemsets on <i>T20I6D100K</i> dataset	69
4.14 The number of early terminated itemsets on <i>T40I10D100K</i> dataset	70
4.15 The number of non-regarded tids during intersection process on <i>accidents</i> dataset	70
4.16 The number of non-regarded tids during intersection process on <i>chess</i> dataset	71
4.17 The number of non-regarded tids during intersection process on <i>connect</i> dataset	71
4.18 The number of non-regarded tids during intersection process on <i>mushroom</i> dataset	72

Figure	Page
4.19 The number of non-regarded tids during intersection process on <i>pumsb</i> dataset	72
4.20 The number of non-regarded tids during intersection process on <i>pumsb*</i> dataset	73
4.21 The number of non-regarded tids during intersection process on <i>BMS-POS</i> dataset	73
4.22 The number of non-regarded tids during intersection process on <i>retail</i> dataset	74
4.23 The number of non-regarded tids during intersection process on <i>T10I4D100K</i> dataset	74
4.24 The number of non-regarded tids during intersection process on <i>T20I6D100K</i> dataset	75
4.25 The number of non-regarded tids during intersection process on <i>T40I10D100K</i> dataset	75
4.26 Runtime of TKRIMPE on <i>accidents</i> ($\sigma_r = 1\%$)	76
4.27 Runtime of TKRIMPE on <i>accidents</i> ($\sigma_r = 2\%$)	76
4.28 Runtime of TKRIMPE on <i>accidents</i> ($\sigma_r = 3\%$)	77
4.29 Runtime of TKRIMPE on <i>chess</i> ($\sigma_r = 2\%$)	77
4.30 Runtime of TKRIMPE on <i>chess</i> ($\sigma_r = 4\%$)	78
4.31 Runtime of TKRIMPE on <i>chess</i> ($\sigma_r = 6\%$)	78
4.32 Runtime of TKRIMPE on <i>connect</i> ($\sigma_r = 1\%$)	79
4.33 Runtime of TKRIMPE on <i>connect</i> ($\sigma_r = 2\%$)	79
4.34 Runtime of TKRIMPE on <i>connect</i> ($\sigma_r = 3\%$)	80
4.35 Runtime of TKRIMPE on <i>mushroom</i> ($\sigma_r = 4\%$)	80
4.36 Runtime of TKRIMPE on <i>mushroom</i> ($\sigma_r = 6\%$)	81
4.37 Runtime of TKRIMPE on <i>mushroom</i> ($\sigma_r = 8\%$)	81
4.38 Runtime of TKRIMPE on <i>pumsb</i> ($\sigma_r = 2\%$)	82
4.39 Runtime of TKRIMPE on <i>pumsb</i> ($\sigma_r = 4\%$)	82
4.40 Runtime of TKRIMPE on <i>pumsb</i> ($\sigma_r = 6\%$)	83
4.41 Runtime of TKRIMPE on <i>pumsb*</i> ($\sigma_r = 1\%$)	83
4.42 Runtime of TKRIMPE on <i>pumsb*</i> ($\sigma_r = 2\%$)	84
4.43 Runtime of TKRIMPE on <i>pumsb*</i> ($\sigma_r = 3\%$)	84
4.44 Runtime of TKRIMPE on <i>BMS-POS</i> ($\sigma_r = 1\%$)	85
4.45 Runtime of TKRIMPE on <i>BMS-POS</i> ($\sigma_r = 2\%$)	85
4.46 Runtime of TKRIMPE on <i>BMS-POS</i> ($\sigma_r = 3\%$)	86
4.47 Runtime of TKRIMPE on <i>retail</i> ($\sigma_r = 6\%$)	86
4.48 Runtime of TKRIMPE on <i>retail</i> ($\sigma_r = 8\%$)	87
4.49 Runtime of TKRIMPE on <i>retail</i> ($\sigma_r = 10\%$)	87
4.50 Runtime of TKRIMPE on <i>T10I4D100K</i> ($\sigma_r = 4\%$)	88
4.51 Runtime of TKRIMPE on <i>T10I4D100K</i> ($\sigma_r = 6\%$)	88
4.52 Runtime of TKRIMPE on <i>T10I4D100K</i> ($\sigma_r = 8\%$)	89

Figure	Page
4.53 Runtime of TKRIMPE on <i>T20I6D100K</i> ($\sigma_r = 2\%$)	89
4.54 Runtime of TKRIMPE on <i>T20I6D100K</i> ($\sigma_r = 4\%$)	90
4.55 Runtime of TKRIMPE on <i>T20I6D100K</i> ($\sigma_r = 6\%$)	90
4.56 Runtime of TKRIMPE on <i>T40I10D100K</i> ($\sigma_r = 2\%$)	91
4.57 Runtime of TKRIMPE on <i>T40I10D100K</i> ($\sigma_r = 4\%$)	91
4.58 Runtime of TKRIMPE on <i>T40I10D100K</i> ($\sigma_r = 6\%$)	92
4.59 Memory usage of TKRIMPE on <i>accidents</i>	92
4.60 Memory usage of TKRIMPE on <i>chess</i>	93
4.61 Memory usage of TKRIMPE on <i>connect</i>	93
4.62 Memory usage of TKRIMPE on <i>mushroom</i>	94
4.63 Memory usage of TKRIMPE on <i>pumsb</i>	94
4.64 Memory usage of TKRIMPE on <i>pumsb*</i>	95
4.65 Memory usage of TKRIMPE on <i>BMS-POS</i>	95
4.66 Memory usage of TKRIMPE on <i>retail</i>	96
4.67 Memory usage of TKRIMPE on <i>T10I4D100K</i>	96
4.68 Memory usage of TKRIMPE on <i>T20I6D100K</i>	97
4.69 Memory usage of TKRIMPE on <i>T40I10D100K</i>	97
4.70 Scalability of TKRIMPE ($k : 500, \sigma_r = 6$)	98
4.71 Scalability of TKRIMPE ($k : 10,000, \sigma_r = 6$)	98
5.1 TKRIMIT: Top- k list structure with hash table	102
5.2 Top- k list initialization	107
5.3 Top- k during mining process	108
5.4 The number of reduced tids from TKRIMIT on <i>accidents</i> datasets	114
5.5 The number of reduced tids from TKRIMIT on <i>chess</i> datasets	114
5.6 The number of reduced tids from TKRIMIT on <i>connect</i> datasets	115
5.7 The number of reduced tids from TKRIMIT on <i>mushroom</i> datasets	115
5.8 The number of reduced tids from TKRIMIT on <i>pumsb</i> datasets	116
5.9 The number of reduced tids from TKRIMIT on <i>pumsb*</i> datasets	116
5.10 The number of reduced tids from TKRIMIT on <i>BMS-POS</i> datasets	117
5.11 The number of reduced tids from TKRIMIT on <i>retail</i> datasets	117
5.12 The number of reduced tids from TKRIMIT on <i>T10I4D100K</i> datasets	118
5.13 The number of reduced tids from TKRIMIT on <i>T20I6D100K</i> datasets	118
5.14 The number of reduced tids from TKRIMIT on <i>T40I10D100K</i> datasets	119
5.15 Runtime of TKRIMIT on <i>accidents</i> ($\sigma_r = 1\%$)	119

Figure	Page
5.16 Runtime of TKRIMIT on <i>accidents</i> ($\sigma_r = 2\%$)	120
5.17 Runtime of TKRIMIT on <i>accidents</i> ($\sigma_r = 3\%$)	120
5.18 Runtime of TKRIMIT on <i>chess</i> ($\sigma_r = 2\%$)	121
5.19 Runtime of TKRIMIT on <i>chess</i> ($\sigma_r = 4\%$)	121
5.20 Runtime of TKRIMIT on <i>chess</i> ($\sigma_r = 6\%$)	122
5.21 Runtime of TKRIMIT on <i>connect</i> ($\sigma_r = 1\%$)	122
5.22 Runtime of TKRIMIT on <i>connect</i> ($\sigma_r = 2\%$)	123
5.23 Runtime of TKRIMIT on <i>connect</i> ($\sigma_r = 3\%$)	123
5.24 Runtime of TKRIMIT on <i>mushroom</i> ($\sigma_r = 4\%$)	124
5.25 Runtime of TKRIMIT on <i>mushroom</i> ($\sigma_r = 6\%$)	124
5.26 Runtime of TKRIMIT on <i>mushroom</i> ($\sigma_r = 8\%$)	125
5.27 Runtime of TKRIMIT on <i>pumsb</i> ($\sigma_r = 2\%$)	125
5.28 Runtime of TKRIMIT on <i>pumsb</i> ($\sigma_r = 4\%$)	126
5.29 Runtime of TKRIMIT on <i>pumsb</i> ($\sigma_r = 6\%$)	126
5.30 Runtime of TKRIMIT on <i>pumsb*</i> ($\sigma_r = 1\%$)	127
5.31 Runtime of TKRIMIT on <i>pumsb*</i> ($\sigma_r = 2\%$)	127
5.32 Runtime of TKRIMIT on <i>pumsb*</i> ($\sigma_r = 3\%$)	128
5.33 Runtime of TKRIMIT on <i>BMS-POS</i> ($\sigma_r = 1\%$)	128
5.34 Runtime of TKRIMIT on <i>BMS-POS</i> ($\sigma_r = 2\%$)	129
5.35 Runtime of TKRIMIT on <i>BMS-POS</i> ($\sigma_r = 3\%$)	129
5.36 Runtime of TKRIMIT on <i>retail</i> ($\sigma_r = 6\%$)	130
5.37 Runtime of TKRIMIT on <i>retail</i> ($\sigma_r = 8\%$)	130
5.38 Runtime of TKRIMIT on <i>retail</i> ($\sigma_r = 10\%$)	131
5.39 Runtime of TKRIMIT on <i>T10I4D100K</i> ($\sigma_r = 4\%$)	131
5.40 Runtime of TKRIMIT on <i>T10I4D100K</i> ($\sigma_r = 6\%$)	132
5.41 Runtime of TKRIMIT on <i>T10I4D100K</i> ($\sigma_r = 8\%$)	132
5.42 Runtime of TKRIMIT on <i>T20I6D100K</i> ($\sigma_r = 2\%$)	133
5.43 Runtime of TKRIMIT on <i>T20I6D100K</i> ($\sigma_r = 4\%$)	133
5.44 Runtime of TKRIMIT on <i>T20I6D100K</i> ($\sigma_r = 6\%$)	134
5.45 Runtime of TKRIMIT on <i>T40I10D100K</i> ($\sigma_r = 2\%$)	134
5.46 Runtime of TKRIMIT on <i>T40I10D100K</i> ($\sigma_r = 4\%$)	135
5.47 Runtime of TKRIMIT on <i>T40I10D100K</i> ($\sigma_r = 6\%$)	135
5.48 Memory usage of TKRIMIT on <i>accidents</i>	136
5.49 Memory usage of TKRIMIT on <i>chess</i>	136

Figure	Page
5.50 Memory usage of TKRIMIT on <i>connect</i>	137
5.51 Memory usage of TKRIMIT on <i>mushroom</i>	137
5.52 Memory usage of TKRIMIT on <i>pumsb</i>	138
5.53 Memory usage of TKRIMIT on <i>pumsb*</i>	138
5.54 Memory usage of TKRIMIT on <i>BMS-POS</i>	139
5.55 Memory usage of TKRIMIT on <i>retail</i>	139
5.56 Memory usage of TKRIMIT on <i>T10I4D100K</i>	140
5.57 Memory usage of TKRIMIT on <i>T20I6D100K</i>	140
5.58 Memory usage of TKRIMIT on <i>T40I10D100K</i>	141
5.59 Scalability of TKRIMIT ($k : 500, \sigma_r = 6$)	141
5.60 Scalability of TKRIMIT ($k : 10,000, \sigma_r = 6$)	142
6.1 H-TKRIMP: Top- k list structure with hash table	144
6.2 Top- k list initialization	154
6.3 Top- k during mining process	155
6.4 Runtime of H-TKRIMP on <i>accidents</i> ($\sigma_r = 1\%$)	162
6.5 Runtime of H-TKRIMP on <i>accidents</i> ($\sigma_r = 2\%$)	162
6.6 Runtime of H-TKRIMP on <i>accidents</i> ($\sigma_r = 3\%$)	163
6.7 Runtime of H-TKRIMP on <i>chess</i> ($\sigma_r = 2\%$)	163
6.8 Runtime of H-TKRIMP on <i>chess</i> ($\sigma_r = 4\%$)	164
6.9 Runtime of H-TKRIMP on <i>chess</i> ($\sigma_r = 6\%$)	164
6.10 Runtime of H-TKRIMP on <i>connect</i> ($\sigma_r = 1\%$)	165
6.11 Runtime of H-TKRIMP on <i>connect</i> ($\sigma_r = 2\%$)	165
6.12 Runtime of H-TKRIMP on <i>connect</i> ($\sigma_r = 3\%$)	166
6.13 Runtime of H-TKRIMP on <i>mushroom</i> ($\sigma_r = 4\%$)	166
6.14 Runtime of H-TKRIMP on <i>mushroom</i> ($\sigma_r = 6\%$)	167
6.15 Runtime of H-TKRIMP on <i>mushroom</i> ($\sigma_r = 8\%$)	167
6.16 Runtime of H-TKRIMP on <i>pumsb</i> ($\sigma_r = 2\%$)	168
6.17 Runtime of H-TKRIMP on <i>pumsb</i> ($\sigma_r = 4\%$)	168
6.18 Runtime of H-TKRIMP on <i>pumsb</i> ($\sigma_r = 6\%$)	169
6.19 Runtime of H-TKRIMP on <i>pumsb*</i> ($\sigma_r = 1\%$)	169
6.20 Runtime of H-TKRIMP on <i>pumsb*</i> ($\sigma_r = 2\%$)	170
6.21 Runtime of H-TKRIMP on <i>pumsb*</i> ($\sigma_r = 3\%$)	170
6.22 Runtime of H-TKRIMP on <i>BMS-POS</i> ($\sigma_r = 1\%$)	171
6.23 Runtime of H-TKRIMP on <i>BMS-POS</i> ($\sigma_r = 2\%$)	171

Figure	Page
6.24 Runtime of H-TKRIMP on <i>BMS-POS</i> ($\sigma_r = 3\%$)	172
6.25 Runtime of H-TKRIMP on <i>retail</i> ($\sigma_r = 6\%$)	172
6.26 Runtime of H-TKRIMP on <i>retail</i> ($\sigma_r = 8\%$)	173
6.27 Runtime of H-TKRIMP on <i>retail</i> ($\sigma_r = 10\%$)	173
6.28 Runtime of H-TKRIMP on <i>T10I4D100K</i> ($\sigma_r = 4\%$)	174
6.29 Runtime of H-TKRIMP on <i>T10I4D100K</i> ($\sigma_r = 6\%$)	174
6.30 Runtime of H-TKRIMP on <i>T10I4D100K</i> ($\sigma_r = 8\%$)	175
6.31 Runtime of H-TKRIMP on <i>T20I6D100K</i> ($\sigma_r = 2\%$)	175
6.32 Runtime of H-TKRIMP on <i>T20I6D100K</i> ($\sigma_r = 4\%$)	176
6.33 Runtime of H-TKRIMP on <i>T20I6D100K</i> ($\sigma_r = 6\%$)	176
6.34 Runtime of H-TKRIMP on <i>T40I10D100K</i> ($\sigma_r = 2\%$)	177
6.35 Runtime of H-TKRIMP on <i>T40I10D100K</i> ($\sigma_r = 4\%$)	177
6.36 Runtime of H-TKRIMP on <i>T40I10D100K</i> ($\sigma_r = 6\%$)	178
6.37 Memory usage of H-TKRIMP on <i>accidents</i>	178
6.38 Memory usage of H-TKRIMP on <i>chess</i>	179
6.39 Memory usage of H-TKRIMP on <i>connect</i>	179
6.40 Memory usage of H-TKRIMP on <i>mushroom</i>	180
6.41 Memory usage of H-TKRIMP on <i>pumsb</i>	180
6.42 Memory usage of H-TKRIMP on <i>pumsb*</i>	181
6.43 Memory usage of H-TKRIMP on <i>BMS-POS</i>	181
6.44 Memory usage of H-TKRIMP on <i>retail</i>	182
6.45 Memory usage of H-TKRIMP on <i>T10I4D100K</i>	182
6.46 Memory usage of H-TKRIMP on <i>T20I6D100K</i>	183
6.47 Memory usage of H-TKRIMP on <i>T40I10D100K</i>	183
6.48 Scalability of H-TKRIMP ($k : 500, \sigma_r = 6$)	184
6.49 Scalability of H-TKRIMP ($k : 10,000, \sigma_r = 6$)	184