

מודלים במסחר אלקטרוני - תרגיל בית 1

תאריך הגשה: 24/5/2006

המרצה: פרופ' משה טננהולץ. המתרגל: מר אלון אלטמן.

שאלה 1

- א. לאיזו מחלקה (class) שייכת הכתובת הבאה: 135.123.45.67? נמקו.
- ב. ציינו לכל אחד מהשרותים הבאים האם כדאי להפעיל אותו מעל TCP או UDP. נמקו!

- (1) העברת דואר אלקטרוני
- (2) העברת הודעות קצרות (SMS)
- (3) טלפון באינטרנט
- (4) מערכת ניהול רשתות
- (5) אתר מסחר אלקטרוני
- (6) מערכת אימות כרטיס אשראי

- ג. אם ברשת נתונה פועל יותר משרת DHCP אחד המקצים כתובות IP מאותו מרחב כתובות, איזו בעיה עלולה לקרות?

שאלה 2

נתונים 3 דפי web: דף הקורס מסחר אלקטרוני (A), homepage של משה (B) ו-homepage של אלון (C) עם הקשרים הבאים: דף הקורס מקשר אל הדף של משה ואל הדף של אלון, הדף של משה מקשר אל הדף של אלון והדף של אלון מקשר אל דף הקורס מסחר אלקטרוני ואל עצמו.

א. מה ה-PageRank של כל דף? נתון $d = 85\%$ (כלומר 15% סיכוי לעבור לאתר אקראי).

ב. מה ה-hubbusiness ומה ה-authority של כל דף?

שאלה 3

נתון אוסף טרנזקציות market-basket הבא:

מספר לקוח	פריטים שנקנו
1	A, D, E
1	A, B, C, E
2	A, B, D, E
2	A, C, D, E
3	B, D, E
3	C, D
4	A, B, C
4	A, D, E
5	A, B, E
5	A, B, E

- א. יש לחשב את ה- support עבור המוצרים: $\{E\}$, $\{B, D\}$, ו- $\{B, D, E\}$.
- ב. יש לחשב את רמת הביטחון של ה- association rules הבאים: $E \rightarrow BD$, $BD \rightarrow E$. יש להסביר את התוצאות.
- ג. יש לבנות מטריצת market-basket עפ"י הנתונים בטבלה הנ"ל לפי מספרי הל-קוחות כאשר סל עבור לקוח הוא איחוד כל הסלים שהוא קנה. יש לפתור את סעיפים א' ו-ב' על מטריצה זו.
- ד. נתון כי s_1 היא רמת הביטחון של r association rule עפ"י הטבלה הנתונה, וכי s_2 היא רמת הביטחון של r association rule עפ"י המטריצה שבנית בסעיף ג'. האם ישנו קשר ($=, >, <$) בין s_1 ו- s_2 ? יש להוכיח את תשובתך!

שאלה 4

נתונים 5 וקטורים המתארים רכישות 10 ספרים ע"י 10 לקוחות:

A=0101010101
 B=0011001100
 C=1110010001
 D=1001001001
 E=0001111100

- א. יש לחשב jaccard measure בין כל זוג וקטורים לעיל.
- ב. יש לבצע clustering לקבוצת הווקטורים הנ"ל לפי אלגוריתם hierarchical clustering. יש לעצור כאשר נותרים שני clusters. יש לפרט את שלבי ההרצה ואת העץ (ההיררכיה) המתקבל.
- את המרחק בין שני clusters יש לחשב לפי מינימום ה- jaccard distance בין שני וקטורים, אחד מכל cluster. להזכירכם, jaccard distance בין וקטורים מוגדר כאחד פחות ה- jaccard measure ביניהם.

שאלה 5

יש לבצע clustering לקבוצת הנקודות הבאה לפי אלגוריתם k-means תוך שימוש במרחק אוקלידי (L_2) עבור $k = 3$. יש לפרט את שלבי ההרצה ואת ה-clusters המתקבלים.

הנקודות: $\{(2, 20), (2, 5), (8, 4), (5, 8), (7, 5), (6, 4), (1, 2), (4, 9)\}$.

יש להתחיל עם המרכזים ההתחלתיים הבאים $x_1 = (2, 20); x_2 = (2, 5); x_3 = (8, 4)$.

שאלה 6

הטבלה הבאה מכילה נתונים לגבי קונפיגורציות של מחשבים בהם פגע ה-worm לאחרונה וכאלה שלא נפגעו:

מערכת הפעלה	דפדפן	תוכנת דואר	נפגע?
Linux	Firefox	Thunderbird	לא
Windows XP	Internet Explorer	Outlook	כן
Windows XP	Firefox	Outlook	כן
Windows XP	Internet Explorer	Thunderbird	לא
Windows 2000	Internet Explorer	Outlook	כן
Windows 2000	Firefox	Outlook	לא
Windows 2000	Internet Explorer	Thunderbird	כן

א. יש לבנות עץ החלטה ע"י שימוש באלגוריתם ID3 על מנת לקבוע האם מחשב בעל קונפיגורציה מסוימת יפגע מה-worm או לא.

יש לבחור את הרמה הבאה לבניית העץ לפי האנטרופיה המשוקללת של הקבוצות המתקבלות (ולא האנטרופיה המקסימלית)

ב. בהסתמך על עץ ההחלטה שבנית, יש לקבוע האם מחשב בעל מערכת הפעלה Linux שמריץ Firefox ו-Outlook ייפגע מה-worm. ומה לגבי מחשב בעל Windows XP שמריץ Firefox ו-Thunderbird?

שאלה 7

בגלל תכיפות הבחירות בארץ הוחלט למחשב את מערכת הבחירות.

אזרחי המדינה מעוניינים לבצע הצבעה חשאית לבחירת ראש הממשלה.

לשם כך הם שולחים את ההצבעות "ש" או "פ" לשופט חשין, שכולם סומכים עליו. בהמשך נדון במספר שיטות למימוש ההצבעה ברשת מחשבים. נאמר ששיטה כלשהי עומדת בתנאי הסודיות, אם איש מלבד השופט חשין אינו מסוגל לגלות מהי הצבעתו של משתתף אחר כלשהו. לשופט חשין מותר לדעת מהן ההצבעות של המשתתפים.

א. על-מנת שהקולות לא יעברו בצורה גלויה, הוחלט להשתמש באלגוריתם RSA להצפנה. כל משתתף יצפין את ההצבעה שלו ("ש" או "פ") תחת המפתח הפומבי של חשין וישלח את התוצאה לחשין. האם שיטה זו עומדת בתנאי הסודיות? נמקו!

ב. כיצד ניתן לשפר את הפרוטוקול המוצע בא', כאשר עדיין חייבים להשתמש בה-צפנה ב-RSA, כך שההצבעה לא תיחשף?

ג. יו"ר ועדת הבחירות בחר בשיטה שעומדת בקריטריונים של סעיפים א' וב'. לאחר ההצבעה, התגלה כי אחד המשתתפים זייף הצבעות של משתתפים אחרים. הציעו שיטת הצבעה שחסינה לזיופים ועומדת בתנאי סעיפים א' וב'.

שאלה 8

שאלה זו עוסקת באלגוריתם multistage PCY למציאת frequent pairs.

א. הוצע, לשם יעילות, להשתמש באותה פונקציית hash בכל שלבי האלגוריתם. כיצד תשפיע פעולה זו על ביצועי האלגוריתם? נמקו את תשובתכם!

ב. לשם חסכון בזיכרון, הוצע שבכל שלב יישמר רק ה-bitmap מהשלב הקודם ולא יישמרו bitmaps אחרים ו/או רשימת ה-frequent items. כיצד תשפיע פעולה זו על ביצועי האלגוריתם? נמקו את תשובתכם!

ג. הוצע לשפר את האלגוריתם ולשמור בכל שלב את תוצאות הספירה עצמה של ה-hash table ולא רק bitmap. כיצד ישפיע שינוי זה על ביצועי האלגוריתם, אם אנו מניחים שכל הזיכרון הפנוי מנוצל לאחסון ה-hash table?