

תרגיל ריצה – הסקה ברשת בייסיאנית

הבעיה

עליכם לממש רשת בייסיאנית ואלגוריתמי הסקה הסתברותית שונים, ולהשוות ביניהם. האלגוריתמים שעליכם לממש:

1. הסקה פשוטה, ללא שום שיפורים, כמו שמתואר בשקף 67 במצגת.
2. Variable elimination, עם הורדת משתנים מיותרים בהתחלה (כמו שמתואר בשקף 91 במצגת), כאשר סדר האלימינציה של המשתנים הוא לפי סדר ה-ABC.
3. כמו 2, כאשר אתם קובעים בצורה היוריסטית את סדר האלימינציה של המשתנים.

מכיוון שזמני הריצה שונים ממחשב למחשב, הפרמטר שאותו התוכנית תמדוד הוא מספר פעולות הכפל והחיבור שנעשו.

שימו לב: באלגוריתמים 2 ו-3 יש צורך גם לקבוע את סדר הכפלת ה-factors כאשר מבטלים משתנה. סדר זה יבחר כך שיווצרו כמה שפחות שורות בפקטור החדש. במידה ויש כמה אפשרויות שקולות, יש לבחור באפשרות שיוצרת פקטור שסכום ה-ASCII של המשתנים שלו הוא הקטן ביותר.

מימוש

קלט

התוכנית תקרא את כל הקלט שלה מקובץ יחיד - input.txt. הקובץ יכיל את מבנה הרשת, ואת השאילתות השונות. לתרגיל מצורף קובץ input.txt לדוגמה, שמכיל את הרשת הבייסיאנית של האזעקה שלמדנו בכיתה (מופיעה בשקף 59 במצגת), וקובץ input2.txt שמכיל רשת בייסיאנית בה חלק מהמשתנים אינם בינאריים.

בכל קובץ השורה הראשונה תתחיל תמיד במילה Network, כדי לסמן שזהו החלק המתאר את הרשת. בשורה השנייה יהיו רשומים שמות המשתנים. שם של משתנה הוא רצף של אותיות וספרות ללא רווחים. לאחר מכן, עבור כל משתנה יהיו רשומים הערכים האפשריים שלו, מי הם ההורים שלו ברשת, וה- CPT. כדי לתאר את ה- CPT תהיה שורה עבור כל צירוף ערכים של הורי המשתנה, ואז ההסתברות עבור כל ערך של המשתנה. לדוגמה, למשתנה B בקובץ input.txt אין הורים, ולכן רשום none בחלק של ה- parents, וב- CPT שלו רשום $P(B=true) = 0.001$. למשתנה A בקובץ input.txt יש שני הורים, B ו-E, ולכן כשרשומה השורה הבאה ב- CPT שלו $P(A=true|B=false,E=true) = 0.29$. אומר ש- $P(A=true|B=false,E=true) = 0.29$. למשתנה C בקובץ input2.txt יש 2 הורים, A ו-B, כאשר A הוא בינארי, אבל B ו-C מקבלים 3 ערכים אפשריים, ולכן כשרשומה השורה הבאה ב- CPT שלו $P(C=stay|A=true,B=maybe) = 0.2$. אומר ש- $P(C=go|A=true,B=maybe) = 0.3$ וגם $P(C=stay|A=true,B=maybe) = 0.2$.

לאחר מכן יתחיל החלק של השאילתות, שיסומן במילה Queries. עבור כל שאילתה יהיה רשום בקובץ באיזה אלגוריתם יש להשתמש: 1, 2 או 3, לפי מה שתואר קודם. השאילתה תמיד תהיה במבנה הבא: $P(Q=q|E1=e1, E2=e2, \dots, Ek=ek)$. ז"א השאילתה תמיד תשאל על ההסתברות לקבל ערך עבור משתנה query אחד, כאשר נתונים מספר משתני evidence. במידה ותוצאת השאילתה כבר נתונה באחד מה- CPT, אין לחשב אותה אלא לתת אותה מיד. במידה ויש ערך עם הסתברות 0 יש להוריד אותו מהטבלה ולא להכניס לחישובים.

פלט

הפלט ייכתב לקובץ output.txt. בשורה הראשונה בקובץ יש לכתוב את תוצאת השאילתה הראשונה (מעוגלת ל-5 מקומות אחרי הנקודה), פסיק, מספר פעולות החיבור שנדרשו ע"י האלגוריתם כדי לענות על השאילתה, פסיק, ואז מספר פעולות הכפל שנדרשו. בשורה הבאה יש לכתוב את אותם הנתונים עבור השאילתה השנייה וכן הלאה. אין להוסיף שום סימן או תו אחר בקובץ. שימו לב: בפעולת הנירמול יש לספור את פעולות החיבור, אך אין לספור שם פעולות כפל (כי יש שם חילוק).

בנוסף יש להגיש קובץ וורד details.docx. בתחילת הקובץ יש לכתוב את פרטי המגיש (שם ות.ז.). לאחר מכן יש לתאר במילים את היוריסטיקה בה בחרתם להשתמש כדי לקבוע את סדר האלימינציה של המשתנים באלגוריתם 3.

אופן הניקוד

- קוד נכון, שמממש את האלגוריתמים, ומחזיר את התוצאה המבוקשת על כל הקלטים החוקיים.
- איכות הפונקציה היוריסטית בה בחרתם להשתמש.
- קוד מתועד וקריא (שמות משתנים ופונקציות משמעותיים).
- הגשה בזמן.

פרטי ההגשה

- ההגשה ביחידים בלבד. אין לפרסם את המטלה ב-GitHub. אין להעתיק אחד מהשני או ממקור חיצוני. תתבצע בדיקת העתקות. תרגיל מועתק (גם בחלקו) יגרור פסילה מיידית של 2 הצדדים המעורבים, בלי קשר למי העתיק ממי.
- ניתן לכתוב את התוכנית ב-Java בלבד, והיא צריכה להתקמפל ולרוץ בגרסה 1.8.0_144. שם המחלקה בה נמצאת פונקציית ה-main יהיה Ex1. יש להשתמש ב-default-package בלבד (ללא תתי תיקיות). חובה להגיש את קבצי המקור.
- אין לממש GUI. קובץ ה-input.txt וקובץ ה-output.txt יהיו באותה ספרייה בה נמצאת התוכנית, ולכן אין לציין ספרייה ספציפית בקוד שאתם מגישים (במידה וכן, ירדו על כך נקודות).
- יינתן קלט ופלט לדוגמה. ודאו שתוכניתכם עובדת אתו כמו שצריך, אך זהו לא הקלט היחיד אותו תיבדק התוכנית.
- התוכנית תיבדק דרך ה-command line ולא ב-eclipse. לכן, כדי לוודא שהתוכנית שלכם עובדת עליכם להעתיק את קבצי המקור ואת הקובץ input.txt שניתן כדוגמה לאחת הספריות במחשב, לפתוח command line ולהריץ `javac *.java` ואז `Ex1.java`. התוכנית תיצור את הקובץ output.txt באותה ספרייה והוא צריך להיות זהה לקובץ output.txt שניתן כדוגמה.
- ההגשה נעשית דרך מערכת הבדיקות האוטומטית. בכתובת: <http://submit.org.il/ariel> שם המשתמש שלכם הוא שם המשתמש שהוקצה לכם באי-מייל שלכם באריאל, כמו שכתוב במכלול. לדוגמה, אם הדואר שלכם במכלול הוא israel.israeli@msmail.ariel.ac.il, אז שם המשתמש שלכם הוא: israel.israeli. הסיסמה היא 4 הספרות האחרונות של תעודת הזהות.
- ברגע שתגישו את המטלה תקבלו למייל הפרטי שלכם שרשום במכלול (לא המייל באריאל) את תוצאות ההגשה-האם התרגיל התקמפל, ואם כן מה היה הפלט שלכם בהשוואה למה הפלט הרצוי. תוכלו כמובן להגיש שוב ולדרוס את ההגשה הקודמת. שימו לב, המייל מהמערכת עלול להגיע לתיקיית הספאם. תבדקו שם וסמנו אותו כ-"not spam". בנוסף, המערכת לא עובדת כמו שצריך עם מיילים של walla.
- חשוב להבהיר: הציון של הבדיקה האוטומטית הוא לא הציון הסופי שלכם במטלה. בבדיקה עצמה ייבדקו קלטים נוספים וייבדק הקוד שלכם. הציון כן מיועד לתת לכם אינדיקציה לגבי מצב התרגיל שלכם.
- תאריך הגשת התרגיל- 07.12.20

בהצלחה!