האוניברסיטה הפתוחה

המחלקה למתמטיקה ולמדעי המחשב

סקירה השוואתית של שיטות המשלבות רשתות בייסיאניות ולוגיקה עמומה

עבודה מסכמת זו הוגשה כחלק מהדרישות לקבלת תואר

"מוסמך למדעים" M.Sc. במדעי המחשב

באוניברסיטה הפתוחה

החטיבה למדעי המחשב

על-ידי

שי קריגמן

העבודה הוכנה בהדרכתה של ד"ר מיריי אביגל

ספטמבר 2011

תוכן העניינים

רשימת טבלאות:

רשימת איורים:

1 מבוא

תחום הבינה המלאכותית כולל בתוכו ניסיונות שונים לפתח את יכולות המחשבים כך שיפעלו בדרך המחקה את החשיבה האנושית ולפתור בעיות שדרך החשיבה האנושית פותרת בקלות. במהלך השנים נעשו ניסיונות רבים לפתח שיטה אחת המצליחה לבצע משימה זו, אולם עד היום לא נמצאה שיטה כזו המסוגלת לבצע זאת בצורה מושלמת. כיום, כולל תחום הבינה המלאכותית מספר כלים ושיטות שהתפתחו כל אחד באופן עצמאי. בין השאר ניתן למנות את האלגוריתמים האבולוציוניים, לוגיקה עמומה, רשתות נוירונים ורשתות בייסיאניות. כלים אלו מצליחים לתת מענה למגוון רחב של בעיות שתחום הבינה המלאכותית מנסה לפתור, אולם אף אחד מהשיטות הללו אינו פותר את כל הבעיות. כל אחת מהשיטות מכוונת לסוג מסוים של בעיות, אם כי יש חפיפה בין קבוצות הפתרונות שכל שיטה מסוגלת לספק. לאחרונה, לאחר ביסוס כל שיטה בפני עצמה והבנה מעמיקה יותר של החסרונות והיתרונות של כל שיטה, נערכו מחקרים רבים המנסים לייצור שיטות היברידיות המשלבות שתי שיטות או יותר כדי לפתור בעיה מסוימת ובכך להתגבר על החסרונות של כל שיטה בעזרת היתרונות של השיטה האחרת. עבודה זו תתרכז בשילוב של שתי שיטות מתחום הבינה המלאכותית: לוגיקה עמומה (fuzzy logic) ורשתות בייסיאניות (Bayesian Networks).

הלוגיקה העמומה פותחה ידי זאדה (Zadeh), במחצית השנייה של שנות השבעים של המאה העשרים [1], כדי לנסות לתת מענה לסוג בעיות שבהן הבעיה או פתרונה או שניהם כאחד אינם ניתנים לניסוח בצורה לוגית. בעיות אלו בדרך כלל עוסקות בבעיות שבני אדם מתארים אותן או את פתרונן בצורה אנושית מעורפלת ולא ניתן להמיר ביטויים עמומים אלו לשפת המחשב – הלוגיקה הרגילה. המטרה שלשמה פותחה הלוגיקה העמומה היא לשמש כחוליה מקשרת בין תיאור הבעיה בצורה אנושית על ידי שפה עמומה לבין "יכולתו" של המחשב להבין רק שפה קשיחה – בינארית וחד משמעית. הלוגיקה העמומה מציעה שיטה בעזרתה ניתן לבטא מצבים עמומים בשיטות מוגדרות היטב, וכך לטפל בצורה ממוחשבת בהחלטות והסקה באופנים הקרובים יותר לאופן בו אנשים מחליטים או מסיקים.

רשתות בייסיאניות מבוססות על חוק בייס. חוק זה מאפשר לחשב הסתברות מותנית של מאורע כאשר יודעים דווקא את ההסתברויות המותנות ההפוכות. הוא נוסח על ידי תומאס בייס [2]. חוק זה מהווה את הבסיס לסטטיסטיקה בייסיאנית. רשת בייסיאנית היא דרך לבנות מודל המציג נתונים וקשרים סיבתיים לא ודאיים בין מאורעות שונים. המודל מבוסס על הסתברות ותיאוריית הגרפים. הוא תואר לראשונה על ידי פרל (J. Pearl) [3]. מודל זה הפך לכלי חשוב המשמש לתיאור מידע לא ודאי בעולם הבינה המלאכותית. רשתות בייסיאניות מאפשרות לתאר את האירועים השונים והתלויות בינם לבין עצמם כגרף מכוון כאשר כל אירוע הוא צומת ותלות היא קשת מכוונת מהגורם לתוצאה וההסתברות המותנה היא הערך שהקשת מקבלת. דרך תיאור זו חוסכת זמן חישוב רב ומיותר ומאפשרת לבנות מודל של מערכות אמיתיות המורכבות ממאורעות הקשורים זה

לזה בקשרים סיבתיים הסתברותיים. מודל כזה מאפשר לחזות את התנהגות המערכת ולבחור את ההתנהגות המתאימה עבור כל תרחיש. היתרונות הגדולים של שיטה זו, לעומת שיטות אחרות לבניית מודלים, הם יכולת הניבוי שלה בתנאי אי-ודאות והיכולת לייצור מתוך ים המידע את הקשרים הסיבתיים בין הנתונים השונים. וכיום היא משמשת רבות במחקר גנטי, כרית נתונים, ניתוח פינאנסי, אבחון פגמים ובקרה תעשייתית.

כאמור, לשתי הגישות יש יתרונות רבים והן משמשות ככלים רבי עוצמה לפתרון בעיות באינטליגנציה מלאכותית. אבל יש לכל אחת מהן גם לא מעט חסרונות. העבודה תתמקד בעבודות המשלבות את שתי השיטות ומנסות לנצל את היתרונות של שיטה אחת כדי להתגבר על החסרונות של השיטה האחרת.

שתי השיטות פותחו כל אחת בפני עצמה והן זכו למימושים והישגים לא מעטים, שתיהן מהוות כלים חשובים ורבי ערך בתחומים שונים של בינה מלאכותית. לכל שיטה יש מרחב הבעיות שהיא מסוגלת לפתור ובהן היא נותנת פתרון טוב (כלומר פתרון נכון תוך צריכת מינימום משאבי מחשוב). לכאורה אין תחומי חפיפה בין מרחבי הבעיות שכל שיטה יכולה לפתור אבל כפי שנראה בהמשך יש לא מעט בעיות שניתנות לפתרון בעזרת שתי השיטות. לדוגמה בעיות סיווג (classification). סיווג הוא משימה חשובה בתחומים רבים בהם זיהוי תבניות (pattern recognition), קבלת החלטות וכריית מידע (data mining). מטרת הסיווג הוא להחליט לאיזו קבוצה שייך כל אחד מנתוני הכניסה. שתי השיטות, הן בייסיאנית והן העמומה, יכולות לבצע משימות סיווג שונות במידה כזו או אחרת של הצלחה ולכל גישה יתרונותיה וחסרונותיה. כאשר מתבצעת ההחלטה באיזו שיטה לנקוט כדי לפתור בעיה מסוימת יש אפשרות לפעול בשתי דרכים. אפשר לנתח את היתרונות והחסרונות של כל שיטה, וכמוכן לנתח את הבעיה העומדת בפנינו ולהחליט באיזו שיטה עדיף לפתור את הבעיה. והאפשרות השנייה היא לנסות לשלב בין שתי השיטות כך שכל שיטה תספק את יתרונותיה ותחפה בכך על חסרונותיה של השיטה השנייה.

בעבודה זו נתאר את שתי השיטות ואת יתרונותיה וחסרונותיה של כל שיטה. ננסה לאבחן את מרחב הבעיות שכל שיטה מתאימה לפתרונן ונבדוק מה המשותף בין שני המרחבים של השיטות. כמו כן נבחן שיטות שונות המשלבות לוגיקה עמומה עם רשתות בייסיאניות. שיטות היברידיות אלו מנסות לנצל את היתרונות של שיטה אחת כדי לפתור את הבעיות של השיטה השנייה.

העבודה מכילה חמישה פרקים: בשני הפרקים הראשונים נדון בכל שיטה בפני עצמה, בלוגיקה העמומה בפרק 2 וברשתות בייסיאניות בפרק 3. מטרתם של שני פרקים אלו היא לתת מבוא לשתי השיטות לחסרונותיהן ויתרונותיהן כהכנה לפרקים העיקריים: בפרק 4 נשווה בין מרחב הבעיות שרשת בייסיאנית מסוגלת לפתור לבין מרחב הבעיות שלפתרונן נועדה הלוגיקה העמומה. בפרק 5 נתאר שלוש שיטות המשלבות רשת בייסיאנית ולוגיקה עמומה בדרכים שונות: פרק 5.1 יתאר דרך לבנות את מערך הכללים של מערכת לוגיקה עמומה בעזרת רשת בייסיאנית. בפרק 5.2 נראה כיצד ניתן לנצל את הלוגיקה עמומה כדי לאפשר משתנים

רציפים ברשת בייסיאנית ובפרק 5.3 נתאר גישה המנסה לבנות רשת בייסיאנית עבור משתנים עמומים. לבסוף, בפרק 6 נסכם את הדברים וננסה להעלות מסקנות.

2 מבוא ללוגיקה עמומה

מערכות מחשבים מבוססות על גישה ספרתית ובינארית. המבנה הבסיסי של תכנית מחשב הוא הספרה הבינארית (bit) והיא מייצגת את הבחירה באחד משני מצבים, כן או לא, 0 או 1. אולם, תופעות רבות, ובעיקר תופעות שבני אדם חשים בהם, אינן דווקא בינאריות. אנו נוהגים לתאר את מזג האוויר כ"די חם", את מצב תזרים המזומנים כ "מצוין" או "בעייתי", את מראהו של אדם אחר כ"די יפה" וכדומה. אבל, כאמור, מערכות ספרתיות דטרמיניסטיות ובינאריות אינן יכולות להתמודד עם נתונים עמומים שכאלה.

לוגיקה עמומה פותחה כדי לנסות לתת מענה לסוג בעיות שבהן לא ניתן להגדיר את הבעיה בצורה דטרמיניסטית. המטרה שלשמה פותחה הלוגיקה העמומה היא לשמש מן חוליה מקשרת בין תיאור הבעיה בצורה אנושית על ידי שפה עמומה לבין "יכולתו" של המחשב להבין רק שפה קשיחה – בינארית וחד משמעית. הלוגיקה העמומה מציעה שיטה בעזרתה ניתן לבטא מצבים עמומים בשיטות מוגדרות היטב, וכך לטפל בצורה ממוחשבת בהחלטות והסקה באופנים הקרובים יותר לאופן בו אנשים מחליטים או מסיקים.

מערכות לוגיקה עמומה משמשות בהקשרים יישומיים רבים. ישנם יישומים פיננסיים, כמו למשל בקבלת החלטות לגבי רכישת נכסים, זיהוי וקבלת החלטות לגבי אפיקי השקעה או אישור הלוואה, וסיוע בתהליך האיחזור (השליפה) של נתונים מתוך מאגרי מידע. מערכות לוגיקה עמומה משמשות תפקיד מרכזי במערכות ייצור ובבקרת פעולתן של מכונות מסובכות. למשל, הלוגיקה העמומה נמצאה מתאימה לבקרת הפעולה של מנועים. מערכות תחבורה ומיזוג אוויר.

כאמור, מערכות לוגיקה עמומה הינן שפה מקשרת בין התיאור העמום של שפת האדם לשפה הקשיחה שמבין המחשב. לכן בעצם מערכות לוגיקה עמומה הן הרחבה של מערכות מומחה רגילות, ומשום כך הן סובלות מחסרונות דומים לאלה של מערכות מומחה מסורתיות. למשל, קשה מאד לבנות מערכות לוגיקה עמומה לומדות.

2.1 עקרונות פעולה

כאמור לעיל, הלוגיקה העמומה פותחה במטרה לאפשר תיאור של מצבים פיסיים בצורה התואמת יותר את החשיבה האנושית שאינה רואה את המציאות במונחים מוחלטים של כן או לא אלא בצורה יותר מעורפלת. גישה זו פותחה לראשונה על ידי זאדה [1]. זאדה פיתח את תורת הקבוצות העמומות ועל בסיס גישה זו מבוססת הלוגיקה העמומה. ממש כמו שיש תאימות בין תורת הקבוצות והפעולות שניתן לבצע בין קבוצות לבין הלוגיקה המסורתית, כך יש תאימות בין תורת הקבוצות העמומות לבין הלוגיקה העמומה. תורת הקבוצות

העמומות, בדומה לתורת הקבוצות הרגילה עוסקת בשייכותם של פרטים לקבוצה. אולם בעוד שבתורת הקבוצות הרגילה ישנן שתי אפשרויות – או שהפרט שייך לקבוצה או שלא – הרי שבתורת הקבוצות העמומות ישנם מצבי ביניים בהם השייכות של איבר לקבוצה לא מלאה. עבור כל קבוצה ניתן להגדיר פונקצית שייכות (**function membership**) שתחום ההגדרה שלה הוא כל הפרטים בעולם, והיא מגדירה לכל פרט מספר בתחום [0,1] המגדיר את "דרגת השייכות" של הפרט לקבוצה. הערך "1" פירושו שייכות מלאה. והערך "0" פירושו אי-שייכות. כל מספר ביניהם מציין שייכות חלקית (ברור אם כן שהקבוצות הרגילות הן מקרה פרטי של קבוצה עמומה כשלכל פרט יש ערך של אחד או אפס בלבד).

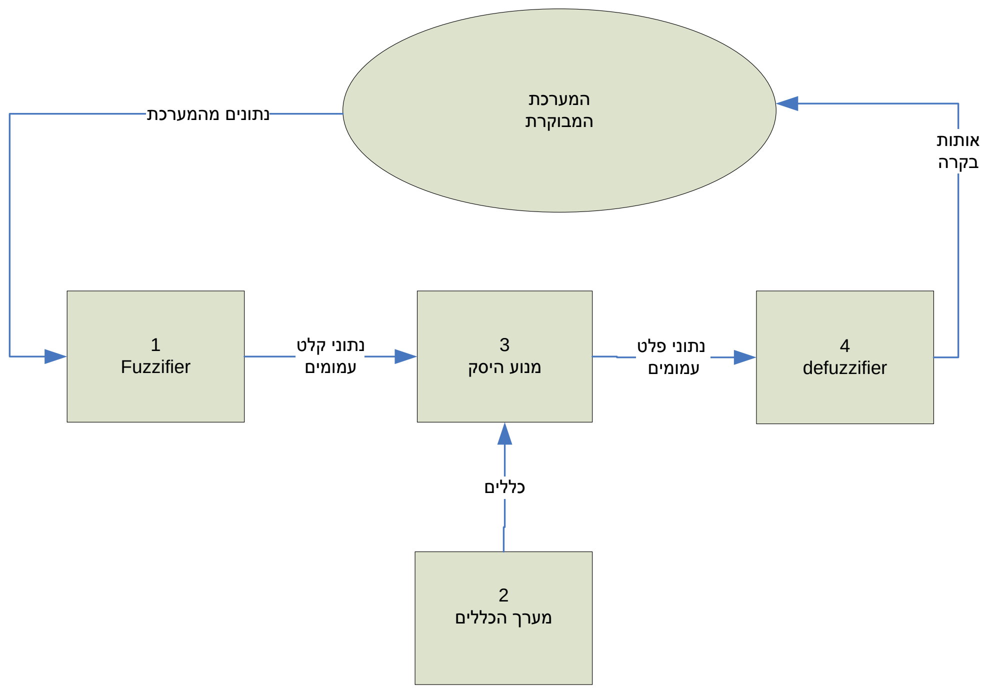
תורת הקבוצות העמומות מגדירה פעולות שונות הניתנות לביצוע על קבוצות עמומות המקבילות לפעולות בקבוצות רגילות לדוגמה: איחוד וחיתוך קבוצות. תורת הקבוצות העמומות מאפשרת להגדיר לוגיקה עמומה הכוללת את כל הפעולות על קבוצות כמו איחוד, חיתוך, מציאת הקבוצה ההופכית ואת האופרטורים המקבילים להן (AND, OR ו- NOT). הלוגיקה העמומה לא מגדירה בצורה חד משמעית את התכונות הללו ויש כמה וכמה דרכים להגדיר אותן. המקובל ביותר הוא להשתמש בפונקצית מינימום ל-AND, פונקצית מקסימום ל- OR והמשלים ל-1 ל- NOT (בצורה זו כללי דה-מורגן ותכונות נוספות מוסיפים להתקיים).

במסגרת זו לא נוכל לעמוד על כל עקרונותיה של הלוגיקה העמומה. נדגיש כאן את היסודות והמונחים הדרושים להבנת עבודה זו. לפרטים נוספים הקורא מופנה לספרי המחקר והלימוד הרבים (לדוגמא [4]).

2.2 ארכיטקטורה של מערכת לוגיקה עמומה

מערכת לוגיקה עמומה [5] כוללת שלושה מרכיבים: המרה של נתוני הכניסה למונחים עמומים, הפעלת מערך כללים על מונחים אלו כשהתוצאה היא מונחים עמומים חדשים. בשלב האחרון מומרים מונחים אלו לנתוני היציאה הדרושים.

מערכות לוגיקה עמומה מיושמות בדרך כלל על בסיס המודל של ממדני (Mamdani Model) [6] המכונה גם RBFS (rule-based fussy system). הרעיון הבסיסי במודל זה הוא המרה של נתוני הכניסה למונחים עמומים, הפעלת מערך כללים על מונחים אלו כשהתוצאה היא מונחים עמומים חדשים. בשלב האחרון מומרים מונחים אלו לנתוני היציאה הדרושים. איור 1 מתאר דיאגראמה בסיסית של המודל.



איור 1: תרשים מלבנים של מערכת בקרה מבוססת לוגיקה עמומה

כפי שאפשר לראות, מודל זה מורכב מארבעה מרכיבים:

1. **Fuzzifier**- מקבל את נתוני הקלט במונחים כמותיים וממיר אותם למונחי הלוגיקה העמומה. מרכיב זה משמש כמתאם בין הנתונים הנמדדים בדרך כלל בצורה כמותית (לדוגמה מדידות של חיישנים מסוגים שונים) לבין מערכת הלוגיקה העמומה המנתחת נתונים בעלי אופי עמום. המרה זו מתבצעת בעזרת פונקציות שייכות. פונקציות אלו יוצרות זיקה בין המונחים הכמותיים (לדוגמה טמפרטורה של חיישן חום) לבין המונחים העמומים (לדוגמה: חם מאד, קריר, קר מאד). כל משתנה קלט יכול לקבל מספר ביטויים עמומים כאשר לכל ביטוי כזה תהיה פונקצית שייכות מתאימה.
2. **מערך הכללים (RB-rule base)** של המערכת מורכב מאוסף תנאים (if-then rules) עמומים. כללים אלו מתארים קשרים לא-לינאריים איכותיים בין צמד משתנים עמומים (דוגמה לכלל כזה יכולה להיות: "אם התנור חם מאוד אז צריך להנמיך במעט את האש").
3. **מנוע ההיסק** – מפעיל את הכללים הנמצאים בבסיס הכללים על נתוני הקלט ומייצר את נתוני הפלט במונחים עמומים.
4. **defuzzifier**- מקבל את נתוני הפלט ממנוע ההיסק ומתרגם אותם למונחים כמותיים הנדרשים ביציאה מן המערכת (לדוגמה: הזרם הדרוש להפעלת מנוע). גם כאן כמו ב-fuzzifier לכל משתנה

פלט ישנם ביטויים עמומים המתארים אותו. לכל ביטוי ישנה פונקצית שייכות ונדרשת פעולת המרה הפכית המתרגמת את הביטוי העמום למספר כמותי המאפשר לבקר ולשנות את אותה מערכת הנשלטת על ידי מערכת הלוגיקה העמומה.

לדוגמא, נתאר מנוע המבוקר על ידי בקר המבוסס על לוגיקה עמומה. המנוע מסיע מנוף ממקום למקום כשמטרת מערכת הבקרה היא להניע את המנוע כך שיגיע למקום היעד במהירות אך בצורה חלקה וללא זעזועים הנגרמים משינויים חדים במהירות. נתוני הקלט למערכת הבקרה הם מהירות המנוע ברגע נתון והמרחק ליעד ברגע זה. נתון הפלט של מערכת זו יהיה הזרם החשמלי למנוע הקובע את מהירותו. ה-fuzzifier ימיר את הנתונים המספריים אודות המהירות והמיקום לביטויים עמומים שיכולים להיות: "מהירה", "ממוצעת", "איטית" ו"איטית מאוד" עבור מהירות המנוע, ו"רחוק", "קרוב" ו"קרוב מאוד" עבור המיקום. הפונקציה המקשרת בין הנתונים המספריים לבין הביטויים העמומים נקראת פונקצית שייכות. בשלב הבא, במנוע ההיסק מופעלים על הנתונים הכללים הנמצאים במערך הכללים. כללים אלו יכולים להיות: "אם היעד קרוב הקטן את המהירות" או "אם עדיין היעד רחוק והמהירות איטית, הגדל את המהירות בצורה משמעותית". כפי שאפשר לראות גם נתון הפלט מבוטא במושגים מעורפלים: "הקטן מהירות" ו"הגדל מהירות בצורה משמעותית". תפקידו של ה-defuzzifier יהיה להמיר ביטויים אלו למספרים לדוגמא הקטנת הזרם החשמלי למנוע ב-1mA או הגדלת זרם זה ב-5mA (בהתאמה לביטויים לעיל).

מערכות לוגיקה עמומה יכולות לפתור בעיות סיווג שבהן מאפייני האובייקטים מאופיינים בעמימות. במערכות אלו, משתני הכניסה למערכת מתארים את המאפיינים של האובייקטים שאנו מעוניינים לסווג ומשתנה היציאה הוא המחלקה אליה משתייך האובייקט. יש לשים לב לעובדה שבמערכת זו הסיווג אינו בהכרח קשיח, זאת אומרת שאובייקט יכול להיות שייך לשתי מחלקות בדרגות שייכות שונות. כלל אופייני במערכת זו יכול להיות מוצג בדרך הבאה:

$$R_k : \text{IF } X_1 \text{ is } A_{1l_1} \text{ AND ... AND } X_n \text{ is } A_{nl_n} \text{ THEN } Class = C_j$$

כאשר: R_k הוא מזהה של הכלל, $X_1..X_n$ הם המאפיינים של האובייקט העומד לסיווג (ומאפיינים אלו הם לשוניים-עמומים), $A_{1l1}..A_{nln}$ הם ביטויים לשוניים-עמומים אפשריים עבור כל משתנה עמום, ולבסוף C_j הוא מחלקה אליה שייך האובייקט (וכאמור לעיל המחלקה יכולה להיות עמומה או לא).

כאשר קיימים כמה כללים במערך הכללים המייחסים אובייקט כלשהו לכמה מחלקות שונות, יש צורך לקבוע איזה כלל מכריע ולאן לשייך בסופו של דבר את האובייקט. קיימות דרכים שונות כיצד להגיע להחלטה הנכונה אודות סיווגו של האובייקט. הגישה הקלאסית ((Classic Fuzzy Reasoning Method (CFRM) בודקת את מידת ההתאמה בין כל המאפיינים לבין כל כלל תוך שימוש ב-t-norm. התוצאה תהיה שהכלל שיקבע את המחלקה אליה שייך האובייקט הוא בעל מידת ההתאמה המכסימלית.

2.3 הקשיים שבפיתוח מערכת לוגיקה עמומה

פיתוח מערכת לוגיקה עמומה כולל בתוכו הן את הגדרת פונקציות השייכות של נתוני הקלט השונים והן את בנית מערך הכללים הנדרש. בין השיקולים הנדרשים לפיתוח מערכת כזו, ניתן למנות את בחירת פונקצית השייכות לכל אחד מנתוני הקלט ואת הפרמטרים השונים של פונקציות אלו וכן את הגדרת הכללים הנכונים שיובילו בסופו של דבר לפתרון הבעיה עבור כל מרחב נתוני הקלט האפשרי. שתי פעולות אלו אינן טריוויאליות ודורשות מומחיות רבה הן בהבנת עקרונות מערכות לוגיקה עמומה והן בבעיה המסוימת המיועדת לפתרון בעזרת מערכת זו. כדי לבנות את מערך הכללים יש צורך במומחים המכירים את הפתרון לבעיה ומסוגלים לנסח אותו בכללים. הצורך במומחיות זו מקשה מאוד על פיתוח מערכות מבוססות לוגיקה עמומה שכן לעיתים קרובות אין מומחים בעלי ידע, לעיתים אין הם מסוגלים לנסח את הפתרון המוכר להם בכללים (כאשר לדוגמה יש להם יכולת אינטואיטיבית לפתרון הבעיה) ולעיתים כלל לא ידוע על פתרון קיים לבעיה.

אחת הדרכים העיקריות להתמודדות עם קשיים אלו היא שימוש באלגוריתמי למידה, לדוגמה רשת נוירונים. בשיטה זו, מערכות לוגיקה עמומה משולבות ברשתות נוירונים כך שרשת הנוירונים משמשת כתחליף למערך הכללים. רשת הנוירונים "לומדת" את מערך הכללים של המערכת, ואילו המערכת הלוגית מיישמת את הפעלת הכללים. לאחר שלב האימון של רשת הנוירונים, היא מסוגלת לקבל את נתוני הקלט (לאחר שהופעלו עליהם פונקציות השייכות המפותחות עדיין בשיטה הרגילה) ולהוציא במוצא את הפלט הרצוי. אחד החסרונות המרכזיים של שיטה זו הם שלא תמיד ניתן לבצע את שלב האימון במערכת (לדוגמא כשהיא אינה בשליטתנו). חסרון שני הוא שאין בידינו בסופו של דבר את מערך הכללים הניתן להבנה אנושית באשר לטיבה של המערכת ודרך התנהגותה.

דרך אחרת להתמודדות עם קשיים אלו היא להשתמש באלגוריתמי חיפוש שונים על מרחב הבעיה תוך ניסיון למצוא את הפתרון האופטימאלי ואז לקבע את הפרמטרים והכללים שהובילו לפתרון זה. הקושי שבדרך זו הוא שמרחב החיפוש גדול מאוד, מספר הכללים האפשריים לבניית מערכת הכללים ודרגות החופש בהגדרת פונקציות השייכות והפרמטרים שלהן יוצרים מרחב חיפוש גדול מאוד שאינו מאפשר שימוש ראלי באלגוריתמי חיפוש סטנדרטיים. השימוש בדרך זו מוגבל בעיקר למצב בו יש מידע סביר אודות הכללים ופונקציות השייכות הדרושים לבניית המערכת. אלגוריתמי החיפוש משמשים לצורך שיפור ביצועי המערכת וכוונון עדין יותר של המערכת כך שתפתור יותר טוב את הבעיה. דוגמה לאלגוריתם חיפוש היא השימוש באלגוריתמים אבולוציוניים לפיתוח מערכת לוגיקה עמומה. אלגוריתמי חיפוש משמשים הן להגדרת פונקציות השייכות והן לפיתוח מערך הכללים אם כי בדרך כלל קשה לאפיין בן זמנית הן את מערך הכללים והן את פונקציות השייכות. לכן רוב הגישות מטפלות בנפרד בשני האספקטים הללו, אם כי יש גישות המנסות לטפל בו-זמנית הן בפונקציות השייכות והן במערך הכללים.

בעבודה זו נבחן דרך נוספת לבניה אוטומטית של מערך הכללים הפעם תוך שימוש בשיטות בייסיאנית. שיטות אלו מאפשרות למידה של תכונות המערכת מתוך מערך מידע הקיים אודות המאפיינים של המערכת. שיטות אלו מנוצלות כדי לאפיין את המערכת על סמך הגישה הבייסיאנית ולאחר מכן מוצע אלגוריתם לגזירת מערך כללים עמומים מתוך המודל בייסיאני. בפרק 5.1.2 נתאר מחקר המשווה מערכת לוגיקה עמומה שנבנתה בעזרת הגישה האבולוציונית לבין מערכת שנבנתה בעזרת הגישה הבייסיאנית.

2.4 ניתוח הבעיות שלוגיקה עמומה משמשות להן פתרון

כאמור, מטרתם המרכזית של מפתחי הלוגיקה העמומה הייתה לאפשר תיאור "אנושי" של בעיות ובכך לנצל ידע אנושי לפתרון בעיות, ידע שבעליו אינו יודע לנסח אותו במושגים בינאריים המאפשרים המרתו לשפה שהמחשב מסוגל להבין. הלוגיקה העמומה מאפשרת שימוש בביטויים עמומים מחד גיסא, ויכולת לתרגמם לשפה שהמחשב מבין מאידך.

חוסר הבהירות האנושית נובע מכמה וכמה סיבות שונות זו מזו. שתי הסיבות העיקריות המצריכות שימוש בלוגיקה עמומה הן חוסר בהירות הנובע מחוסר ידע (לדוגמה, חקלאי אינו יודע את ההשפעה המדוייקת של כמות המים על גידול מסוים), וחוסר בהירות הנובע מאקראיות במציאות (לדוגמה שני צמחים המקבלים אותה כמות של מים, האחד ייבול והשני יפרח). שני מצבים אלו מטופלים באופן מסורתי על ידי שיטות הסתברותיות שהרשת הבייסיאנית הינה אחת מהן ונתאר זאת בהמשך. סוג אחר של חוסר בהירות, נובע מעצם ההמשגה האנושית, ולטיפול בסוג זה, בעיקר פותחה הלוגיקה העמומה (אם כי נעשים בה שימושים מעניינים מאוד לבעיות מהסוגים הקודמים, לדוגמה בפרק 5.2 נתאר שימוש בלוגיקה עמומה דיסקרטיזציה של משתנה רציף). בני אדם יוצרים מושגים מופשטים כדי לתאר את העולם. כל מושג מתאר קבוצה של עצמים אמיתיים בעלי תכונה משותפת. הבעיה היא שרוב, אם לא כל, התכונות בעולם הטבעי הן רציפות, וקשה, ואפילו בלתי אפשרי, לקבוע את גבולות התכונה. דוגמה לכך היא שאלת הערימה: אבן אחת אינה ערימת אבנים. הרבה אבנים הן ערימת אבנים. מה עם שתי אבנים? הן אינן ערימה. ושלוש? וארבע? מתי האבנים הללו הופכות לערימה? כך כמעט כל מושג דורש את הגדרת הגבול שלו וגבול זה אינו ברור וחד משמעי. הלוגיקה העמומה פותרת בעיה זו על ידי הגדרת מידת השייכות לקבוצה מסוימת לא בצורה החלטית של כן או לא, אלא בצורה הדרגתית. כך לדוגמה אוסף של 9 אבנים יכול להיות ערימה ב-70%.

2.4.1 לוגיקה עמומה ובעיות סיווג

הבעיה שתוארה לעיל נקראת "בעיית הערימה" והיא מתארת בעיה עקרונית בדרך האנושית להגדרת מושגים: כיצד יש להתייחס למצבי הקצה, מצבים בהם התכונות המגדירות את המושג לא קיימות באופן מובהק בפרט כלשהו ועולה השאלה האם המושג המופשט חל עליו או לא. בעולם המחשבים בעיה כזו היא בעיית סיווג.

הסיווג הוא שיוך של אובייקט מסוים לקבוצה אחת מבין כמה אפשריות על פי תכונותיו. משימת סיווג ניתנת לתיאור פורמאלי בדרך הבאה: בהינתן קבוצה של אובייקטים (או תבניות) $E = \{e_1, e_2, ..., e_M\}$, כאשר כל אחד מהם מוגדר על ידי n מאפיינים. כמו כן נתונה קבוצה של מחלקות: $C = \{C_1, C_2, ...,C_J\}$. הסיווג הוא שיוך אובייקט בעל מאפיינים מסוימים המוגדרות בעזרת הווקטור $ep = (ap_1, ap_2, ..., ap_n)$ לאחת המחלקות C_j. הקושי הטמון בפתרון בעיות סיווג הוא שלא תמיד ניתן להתאים באופן מלא את האובייקט לקבוצה אחת ולעיתים הוא מתאים אף לשתי קבוצות או אף יותר. יתכן שעל פי מספר מאפיינים הוא מתאים לקבוצה אחת ועל פי מאפיינים אחרים הוא מתאים לקבוצה אחרת. הלוגיקה העמומה מטפלת במצבים אלו על ידי זה שהיא יכולה לשייך את האובייקט ליותר מקבוצה אחת, כאשר מידת השייכות לקבוצה כלשהי יכולה לנוע בין 1, שפירושו התאמה מלאה לקבוצה, לבין 0, שפירושו חוסר התאמה מלא, וכל מספר בין אחד לבין אפס.

ניתן להבין את הלוגיקה העמומה כטיפול יותר עדין, מדויק והדרגתי במקרי הקצה של הקבוצה. במקום האילוץ לשייך נתון מסוים לקבוצה אחת בלבד, ניתן לומר שהוא "קרוב" לקבוצה מסוימת יותר מאשר לקבוצה הסמוכה. בכך המעבר בין קבוצות אינו חד אלא הדרגתי.

כפי שנראה בהמשך אפשר לבצע דיסקרטיזציה של משתנה רציף בעזרת קבוצות עמומות כך שמאבדים פחות מידע אודות המעברים בין הקבוצות. דיסקרטיזציה רגילה קובעים גבול ברור וחד בין הקבוצות השונות. לדוגמה, נניח שהנתון הרציף הוא טמפרטורה ואנו מבצעים דיסקרטיזציה לשתי קבוצות: "חם" – מעל מאה מעלות ו"קר" - מתחת למאה מעלות. דיסקרטיזציה רגילה, 101 מעלות ייחשבו כ"חם" ו-99 מעלות כ"קר". לעומת זאת דיסקרטיזציה עמומה, 101 מעלות ייחשבו כ"חם" עם מידת שייכות של 0.6 ו"קר" עם מידת שייכות של 0.4. לעומת זאת 99 מעלות ייחשבו כ"חם" עם מידת שייכות של 0.4 ו"קר" עם מידת שייכות של 0.6. בדרך זאת נשמר מידע אודות מקרי הקצה של הקבוצות השונות.

2.4.2 יישומי הלוגיקה העמומה

מערכות לוגיקה עמומה משמשות בהקשרים יישומיים רבים. ישנם יישומים פיננסיים, כמו למשל בקבלת החלטות לגבי רכישת נכסים, זיהוי וקבלת החלטות לגבי אפיקי השקעה או אישור הלוואה, וסיוע בתהליך האחזור (השליפה) של נתונים מתוך מאגרי מידע. מערכות לוגיקה עמומה משמשות תפקיד מרכזי במערכות ייצור ובבקרת פעולתן של מכונות מסובכות. למשל, הלוגיקה העמומה נמצאה מתאימה לבקרת הפעולה של מנועים. מערכות תחבורה ומיזוג אוויר. למרבה ההפתעה יש ללוגיקה עמומה השפעה גם על התנהגות של מערכות אמיתיות. לדוגמה בקרת מנועים: כאשר אנו רוצים לקרב משהו ליעד כלשהו על ידי מנוע, נוצרת תופעה של תנודות, כיוון שהמנוע אינו יכול לעצור בבת אחת הוא עלול לעבור את היעד ואז נצטרך להחזירו ליעדו בתנועה יותר איטית וחוזר חלילה. מסתבר שכאשר בונים מערכת בקרה מבוססת לוגיקה עמומה אפשר

להגיע ליעד בתנועה אחת חלקה ללא עצירות וקפיצות. בצורה זו מופעלות רכבות ביפן כך שלמרות המהירות הגבוהה אליה הן מגיעות עדיין הנוסעים אינם מרגישים את ההאטה וההאצה בכניסה ויציאה מהתחנות.

2.4.3 החסרונות של לוגיקה עמומה

הקושי המרכזי בבניית מערכת לוגיקה עמומה הוא הצורך במידע מדויק ומלא על המערכת. אמנם מידע זה יכול להיות מנוסח בצורה עמומה כלומר לא באופן מספרי מדויק אלא תוך שימוש בביטויים עמומים המובנים יותר על ידי בני אדם, ולכן מידע זה יכול להתנסח על ידי מומחים המכירים את המערכת שאינם בהכרח מומחי מחשבים. אולם אין דרך אחרת לבנות את המערכת ללא מידע זה ולכן אם הוא חסר או לא מדויק הדבר יפגע באופן משמעותי באיכות המודל.

חסרון נוסף הוא שהלוגיקה העמומה מאפשרת מספר דרגות חופש רבות בבחירת פונקציות השייכות ובבחירת פונקציות החיתוך והאיחוד. ולעיתים גם מומחים המכירים היטב את המערכת אינם מסוגלים לבנות פונקציות אלו. גם בנייה נכונה של פונקציות השייכות אינה קלה וגם כאן יש מגוון אין סופי של אפשרויות הן בבחירת הפונקציות המתאימות, הן בתכונות שלהן והן במספר הקבוצות העמומות שכל משתנה יכול לקבל. קושי זה הביא למחקרים רבים שניסו לשלב שיטות לומדות כמו רשתות עצבים מלאכותיות או אלגוריתמים גנטיים במערכת לוגיקה עמומה. גם שיטות בייסיאנית יכולות לשמש למטרה זאת שכן פותחו שיטות רבות ללמידה של רשת בייסיאנית על סמך בסיסי נתונים.

3 מבוא לרשתות בייסיאניות

רשת בייסיאנית היא מודל שנועד לתאר מערכת כלשהי המורכבת מאירועים הקשורים זה בזה בקשר סיבתי [7]. הרשת מורכבת משני מרכיבים. הראשון, גרף מכוון ללא מעגלים בו כל צומת מייצג אירוע כלשהו במערכת והקשתות המחברות את הצמתים, מייצגות את התלות בין אירועים שונים. לדוגמה, ברשת בייסיאנית המנתחת מחלות שונות, הצמתים יכולים להיות גורמי מחלה וסימפטומים, כאשר הקשתות יתארו קשר בין סימפטומים לגורמי מחלה. המרכיב השני ברשת בייסיאנית הוא טבלה המכילה את כל קשתות הגרף ונותנת לכל קשת ערך מספרי המציין את ההסתברות המותנה להתרחשות התוצאה בהינתן שהגורם התרחש. בהינתן רשת בייסיאנית ניתן להשתמש בה לניתוח הסתברויות במערכת הן ביחס לגורמים והן ביחס לתוצאות. כלומר, אפשר לקבל מידע אודות המאורעות השונים במערכת ולבדוק מה ההסתברות אודות המאורעות לגבם אין מידע. אם הידע הקיים הוא אודות הגורמים, הרשת הבייסיאנית תשמש לניתוח התוצאות האפשריות, ואם המידע הקיים הוא אודות התוצאות של המערכת, אזי תשמש המערכת לניתוח הסיבות האפשריות לתוצאות אלו. למרות האפשרות לנתח את שני הכיוונים, יתרונה הגדול של הרשת הוא בניתוח הגורמים, שכן בדרך כלל פשוט יותר לחשב את ההסתברות המותנית להתרחשות התוצאה כאשר הגורמים ידועים. אולם כאשר התוצאה ידועה ויש צורך לנתח את ההסתברות של הגורמים לתוצאה זו, החישוב נעשה קשה יותר.

3.1 עקרונות יסוד

3.1.1 הסתברות מותנה:

הסתברות מותנית היא מושג שעוזר לנו כאשר יש בידנו מידע חלקי בלבד על המצב, ואנחנו רוצים לדעת את המידע החסר. "הסתברות המותנית של מאורע A בהינתן מאורע B" היא הסיכוי להתרחשותו של A, בהנחה ש- B אכן התרחש. כאשר אנו יודעים שהמאורע B קרה, אזי אנו יודעים שיש לבדוק רק את המקרים הנכללים ב-B שבהם גם A התרחש. ההנחה מכווצת, כביכול, את מרחב המדגם, וכך אפשר לחשב הסתברות מותנית בדרך הבאה:

משוואה 1:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

בסימון P(A|B) יש משמעות לכיוון של A ו-B. במקרה זה אנו יודעים שהמאורע B קרה ושואלים מה ההסתברות שגם המאורע A קרה. הכיוון ההפוך: P(A|B) אומר שאנו יודעים שהמאורע A קרה ושואלים מה ההסתברות שהמאורע B קרה.

3.1.2 משפט בייס

האם יש קשר בין שתי ההסתברויות המותנות הללו? התשובה היא כן ומשפט בייס מגדיר אותה [2]:

מתוך הגדרת ההסתברות המותנית מתקיימים השיוויונים הבאים:

משוואה 2:

$$P(A\cap B) = P(A|B)P(B)$$

משוואה 3:

$$P(A\cap B) = P(B|A)P(A)$$

ומכאן נקבל את **משפט בייס**:

משוואה 4:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

לביטויים המרכיבים נוסחה זו שהיא הבסיס לכל פורמליזם בייסיאני ניתנו שמות שונים:

P(A) - **ההסתברות האפריורית** למאורע A היא ההסתברות להתרחשות המאורע A כפי שידוע לנו לפני שיש לנו כל מידע נוסף על הגורמים להתרחשותו של מאורע זה.

P(A|B) - **ההסתברות האפוסטריורית** של המאורע A תלויה כבר בתוצאות "ניסויים" או מידע נוסף שיש לנו אודות ההתרחשות של הגורמים להתרחשות המאורע A. כלומר ידוע לנו ש-B התרחש ורוצים לדעת מה ההסתברות שגם A התרחש.

P(B|A) - היא ההסתברות שאם המאורע A קרה יקרה בעקבותיו המאורע B נקראת **הנראות** (likelihood).

P(B) - היא ההסתברות התיאורטית שיקרה המאורע B שהוא המאורע שגורם למאורע A.

במשפט בייס אנו יוצאים מתוך ההנחה ש-B אכן קרה כבר ואנו שואלים מה ההסתברות ש-A קרה גם הוא. הטענה הטמונה במשפט בייס היא שההסתברות להתרחשותו של המאורע A תלויה גם בשאלה מה ההסתברות שמאורע B יקרה. זו טענה מפתיעה במבט ראשון כיוון שהתפיסה המיידית היא שאם ידוע לנו ש-B התרחש, הרי להסתברות ש-B התרחש אין השפעה על ההסתברות ש-A יתרחש. אולם משפט בייס טוען שאכן יש

קשר בין ההסתברויות, ולצורך חישוב ההסתברות להתרחשותו של A יש צורך לדעת את ההסתברות להתרחשות של B.

עד כה הראנו את הקשר בין שני מאורעות A ו-B כאשר המאורע A תלוי במאורע B בלבד. מה קורה כאשר המאורע A תלוי במספר גורמים? ניתן לתאר את המאורע B כמאורע שהוא תוצאה של גורמים אלו (שנקרא להם Bi) כלומר B יתרחש אם כל המאורעות Bi יתרחשו וניתן לחשב את ההסתברות להתרחשות B בעזרת חישוב המאורעות Bi. אם גורמים אלו אינם תלויים זה בזה (mutually exclusive), הרי ש-P(B) נתונה על ידי הנוסחה הבאה:

משוואה 5 :

$$P(B) = \sum_{j=1}^{m} P(B|Aj)P(Aj)$$

כאשר Aj הם כל המאורעות היכולים לגרום למאורע B.

נוסחה זו מאפשרת את הכללת משפט בייס, שנכתב לעיל עבור שני מאורעות, עבור תלות במאורעות רבים: אם $\{Bi\}i \in K$ חלוקה של מרחב המדגם שבה לכל החלקים הסתברות חיובית, אז לכל מאורע A בעל הסתברות חיובית מתקיים (לכל B_k):

משוואה 6:

$$P(Bk|A) = \frac{P(A|Bk)P(Bk)}{\sum_i P(A|Bi)P(Bi)}$$

משפט בייס הוא כלי שימושי מאוד בעולם האינטליגנציה המלאכותית כיוון שבעזרתו ניתן לבצע כמה פעולות חשובות כגון סיווג (classification) והיסק (inference):

3.1.3 סיווג בייסיאני (Bayesian Classifier):

סיווג בייסיאני הוא אחד הדרכים לפתור בעיות סיווג. סיווג הוא החלטה לשייך עצם מסוים למחלקה אחת מבין כמה אפשרויות על פי המאפיינים שלו. בסיווג בייסיאני משתמשים במידע סטטיסטי הקיים אודות המחלקות ואודות האיבר האופייני לכל מחלקה.

נניח שאנו רוצים לבצע סיווג בייסיאני. נגדיר את Ω כמרחב כל המחלקות האפשריות. עבור כל אובייקט שאנו רוצים לסווג, אנו צריכים להחליט לאיזו מחלקה ω הוא שייך. לשם כך יש צורך לדעת את פילוגי ההסתברות הבאים:

1. P(ω) - פילוג ההסתברות של המחלקות השונות על המרחב Ω. פילוג זה מתאר את שכיחות ההופעה של המחלקות השונות, כלומר אם נבצע סיווג מלא של כל הנתונים האפשריים – מה תהיה שכיחות ההופעה של כל מחלקה ω. P(ω) נקרא פילוג אפריורי של ω.
2. לכל מחלקה ω ∈Ω, נתון פילוג הסתברות מותנה p(x |ω) על מרחב הקלט X. פילוג זה מתאר כיצד נראה איבר אופייני מכל מחלקה. הפילוג המותנה נקרא לעיתים פונקצית הסבירות.

בעיית הסיווג דורשת לחשב בעצם את p(ω |x) שהיא ההסתברות עבור כל קלט חדש x להיותו שייך לקבוצה ω. משפט בייס מאפשר בעצם להשתמש בשני נתונים אלו כדי לחשב הסתברות זו עבור כל אחת מן המחלקות. התוצאה של תהליך הסיווג תהיה בחירת המחלקה בעלת ההסתברות p(ω |x) הגבוהה ביותר מבין כל המחלקות.

3.1.4 הסקה בייסיאנית (Bayesian Inference):

בהסקה בכלל ובהסקה בייסיאנית בפרט, אנו מתעניינים בניסיון לאמת השערה מסוימת על סמך נתונים או מידע שיש לנו אודות המערכת. לדוגמה מערכת לאיתור תקלות ברכב תנסה לשער השערות אודות מהות התקלה על סמך בדיקות במוסך. הבדיקות מספקות נתונים אודות מגוון תסמינים, ועל סמך נתונים אלו ננסה להסיק מהי התקלה שגרמה לתסמינים אלו (או ליתר דיוק מהי התקלה שיש לה את ההסתברות המקסימאלית להופעת תסמינים אלו).

לפתרון שתי בעיות אלו, הן סיווג בייסיאני והן הסקה בייסיאנית, עושים שימוש מסיבי במשפט בייס. אולם פתרון משפט זה במלואו הוא בעיית NP-hard [14], ולכן לא ניתנת לפתרון ממוחשב עבור מערכות הכוללות מספר רב של אירועים. כדי לאפשר זמן חישוב יותר סביר פותחה הרשת הבייסיאנית שהיא בעצם מבנה נתונים המקל על חישובי משפט בייס במערכות מורכבות.

3.1.5 רשת בייסיאנית (Bayesian Network):

רשת בייסיאנית היא דרך לבנות מודל המציג נתונים וקשרים סיבתיים לא ודאיים [7]. רשתות בייסיאניות מאפשרות לתאר את האירועים השונים והתלויות בינם לבין עצמם כגרף מכוון, כאשר כל אירוע הוא צומת ותלות היא קשת מכוונת מהגורם לתוצאה, וההסתברות המותנה היא הערך שהקשת מקבלת.

כדי להבהיר את המושגים השונים נביא דוגמה לרשת בייסיאנית פשוטה (הלקוחה מתוך [25]). בהמשך נשתמש בדוגמה זו כדי להבהיר נושאים שונים. הרשת בדוגמה מתארת מערכת פשוטה לניטור מצב הרטיבות

של משטח דשא (ראה איור 2). המערכת מכילה ארבעה משתנים בוליאניים המתארים ארבעה אירועים המרכיבים את המערכת (אות גדולה מתארת את המשתנה ואות קטנה מתארת ערך שהוא יכול לקבל):

C – מתאר את מצב העננות. (ערך c פירושו שמיים מעוננים ו- c~ (קרי: not c) פירושו שמיים בהירים).

R – משתנה הקובע האם ירד גשם (r- ירד גשם, r~ – לא ירד).

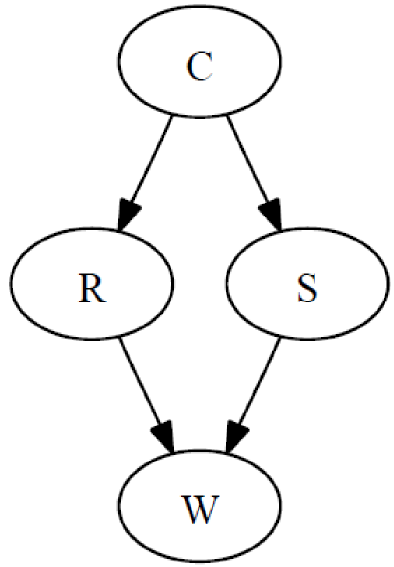
S – מתאר את מצב הממטרות (s – ממטרות עובדות, s~ – הממטרות אינן עובדות).

W- קובע את מצב הדשא (w – הדשא רטוב, w~ – הדשא יבש).

בהתאם לכך הרשת הבייסיאנית המתארת מערכת זו מכילה ארבעה צמתים שכל אחד משויך לאירוע אחר. קשת מכוונת מקשרת בין שני אירועים שיש בין אחד לשני כאשר הכיוון של הקשת הוא מהסיבה לתוצאה (אם ירד גשם הדשא יהיה רטוב). אם אין קשר סיבתי בין שני אירועים, גם לא תהיה קשת המחברת אותם זה לזה. בנוסף לכך אם אין קשר סיבתי ישיר בין שני אירועים, לא תהיה קשת מחברת (מצב העננות אינו משפיע ישירות על הרטיבות של הדשא). לעיתים לימוד עובדות נוספות והוספת אירועים חדשים למערכת יכולה לבטל או להוסיף תלויות בין אירועים (קשתות בין צמתים). לדוגמא במערכת הקושרת בין גובה התלמיד לבין יכולת הקריאה שלו, הוספת גיל התלמיד מבטלת את התלות בין גובהו ליכולת הקריאה שלו. במצב כזה הקשת המחברת בין הצומת המתאר את גובה התלמיד לצומת המתאר את יכולת הקריאה שלו תמחק מהגרף. במקומה ניצור קשת בין הצומת המתאר את גיל התלמיד לבין הצומת המתאר את יכולת הקריאה שלו (ותתכן גם קשת בין הצומת המתאר את גובה התלמיד לבין הצומת המתאר את גילו).

בנוסף לכך כוללת הרשת הבייסיאנית שבדוגמה שלנו ארבע טבלאות, אחת לכל צומת (או אירוע). הטבלה מתארת את ההסתברויות המותנות להתרחשות האירוע שמיוצג על ידי צומת זה בתלות בהורים שלו. כך לדוגמה, הטבלה השלישית מתארת את ההסתברות שירד גשם כתלות במצב העננות: במצב בו לא יהיו עננים (~ בטבלה מציין את אופרטור השלילה) יש הסתברות של 0.2 לגשם (הסיבה שההסתברות היא לא אפס כיוון שהגדרת ה"אין עננים" אינה שמים בהירים לחלוטין אלא מצב ביניים – שמים מעוננים חלקית נחשבים כ"אין עננים"). ואילו בנוכחות עננים ההסתברות לגשם תהיה 0.8. כמובן שעל פי הגדרת ההסתברות, כל שורה בטבלה צריכה להסתכם לאחד.

C		p(c)	p(~c)
		0.5	0.5
C → S		p(s)	p(~s)
~c		0.5	0.5
c		0.9	0.1
C → R		P(r)	P(~r)
~c		0.8	0.2
c		0.2	0.8
R, S → W		p(w)	p(~w)
p(~r)	p(~s)	0	1
p(~r)	p(s)	0.9	0.1
p(r)	p(~s)	0.9	0.1
p(r)	p(s)	0.99	0.01



איור 2: דוגמה לרשת בייסיאנית פשוטה

הגישה הבסיסית של תורת ההסתברות מחייבת חישוב מלא של כל ההסתברויות, הן את ההסתברות לקבל כל ערך אפשרי של כל משתנה. והן את כל ההסתברויות המותנות הקושרות בין כל שני משתנים. גישה זו אמנם נכונה מתמטית אבל בלתי ניתנת למימוש במערכות מורכבות מציאותיות משתי סיבות מרכזיות. הראשונה היא שיש קושי בלתי ניתן לפתרון לחשב או אפילו להעריך את ההסתברויות המותנות הללו. לדוגמה, אם נשתמש במערכת הפשוטה שתוארה לעיל, קשה או אפילו בלתי אפשרי למצוא את ההסתברות המותנה בין המאורעות "השמים מעוננים" ו-"הדשא רטוב". הבעיה השנייה היא שגם אם ניתן ערכים לכל ההסתברויות הללו עדיין מספר החישובים הנדרש גדל אקספוננציאלית עם מספר המשתנים. במקרה שתואר לעיל, עבור ארבעה משתנים יש צורך למצוא 24 הסתברויות מותנות כאלו. מלבד מספר החישובים הנדרש בדרך כלל קשה מאד לחשב הסתברויות אלו במצבים מציאותיים. אמנם שתי בעיות אלו ניתנות לפתרון חלקי על ידי ההנחה שכל המשתנים שאינם תלויים זה בזה שכן במקרה זה אין צורך לחשב את ההסתברות המותנית בין האירועים שהרי הם זרים זה לזה. למרות זאת עדיין קיים קושי רב בחישוב ההסתברויות המותנות השונות. לעומת זאת בני אדם יכולים להעריך בקלות יחסית אי תלות בין משתנים (לדוגמה: "מחר ירד גשם" ו"תוך חמש שנים תפרוץ מלחמה גרעינית"). משמעות הדברים היא שניתן לבנות את מערכת התלויות בין המשתנים השונים לפני חישוב ההסתברויות שכן מערכת זו אינה תלויה בערכי ההסתברות. עובדה זו חשובה מאוד, שכן קל הרבה יותר לבנות את הרשת הבייסיאנית כאשר בשלב הראשון נדרש רק המידע הקושר בין האירועים

השונים ומגדיר את הקשרים הסיבתיים הראשוניים בין האירועים. רק בשלב הבא, לאחר שנבנה הגרף והוגדרו הצמתים והקשתות המחברות אותם זה לזה, אפשר לגשת למילוי הטבלאות של ההסתברויות. לדרך זו יש שני יתרונות: ראשית, מספר ההסתברויות הוא המינימאלי הנדרש לניתוח המערכת ושנית, רק ההסתברויות הקושרות מאורעות בעלי תלות ישירה מחושבות והסתברויות אלו קל יחסית לחשב או להעריך.

המשמעות המעשית של הטענות הללו היא שבבואנו לנתח מערכת המורכבת ממספר רב של משתנים בעלי הסתברויות שונות, יש לבנות קודם את מערכת התלויות בין המשתנים הללו תוך הסתמכות על מקורות ידע שונים ולא מתמטיים. רק לאחר שנבנתה מערכת כזו ניתן להוסיף לה את הידע המתמטי הנדרש אודות ההסתברויות השונות כדי להפיק ממנה את המידע הדרוש.

אחד ממבני הנתונים המתאימים ביותר לבניית מערכת כזו הוא הגרף. גרף בנוי מצמתים המקושרים זה לזה בקשתות. גרף כזה בו הצמתים מייצגים את המשתנים והקשתות את התלויות בין המשתנים. קיומה של קשת בין שני צמתים מעיד על תלות בין שני המשתנים המיוצגים על ידי שני הצמתים.

ההצגה של הסתברות משותפת כרשת בייסיאנית איננה יחידה. למעשה, לכל סדר שנבחר על המשתנים המקריים אנו עלולים לקבל רשת אחרת. בדרך כלל ננסה לסדר את המשתנים לפי סדר סיבתי – מהסיבות הראשוניות לכוון התוצאות. סדר זה ייתן לנו לרוב רשת דלילה המנצלת אי תלויות מותנות רבות.

את הטבלאות המופיעות ברשת הבייסיאנית, קל יותר למלא ע"י מומחה או ע"י מחקר סטטיסטי. נשים לב שכל מילוי של הטבלאות (בתנאי שההסתברויות שבכל שורה מסתכמות ל – 1) נותן לנו תאור של הסתברות משותפת תקינה. במצב זה, אין חשש שטעות במילוי הטבלאות יגרום לסתירה, כלומר למצב לא אפשרי מבחינת הסתברותית.

רשת בייסיאנית הינה, אם כן, מודל יעיל לייצוג פונקציות הסתברות משותפת של סט משתנים אקראיים לשם הסקה הסתברותית וקבלת החלטות. המודל מורכב משני מרכיבים עיקריים: מבנה הרשת וסט הסתברויות מותנות. מבנה הרשת מיוצג באמצעות גרף מכוון, כאשר קיום של קשר בין שני צמתים (משתנים) מייצג תלות של שני המשתנים בגרף. סט ההסתברויות המותנות מכמת את חוזק הקשרים.

ניתן להשתמש ברשת בייסיאנית ללימוד ההתנהגות של המערכת הממודלת בשני כיוונים שונים [8]: ראשית אפשר להתקדם מהגורמים לתוצאות. אם ידועה התרחשות מסוימת מה ההשפעה של ידיעה זו על התרחשויות המושפעות ממנה? השפעה זו מכונה התפשטות קדימה (forward propagation) כלומר משורש הגרף לעבר העלים שלו. גם ניתוח הכיוון ההפוך אפשרי אודות לחוק בייס: אם ידוע מצבו של אירוע מסוים מה זה אומר על ההסתברות להתרחשות האירועים הגורמים לאותו אירוע? ניתוח כזה נקרא התפשטות אחורה (back propagation) שכן בניתוח כזה אנו נעים במעלה הגרף מהעלים לכיוון השורש.

את האלגוריתמים שפותחו לשימוש ברשת בייסיאנית לצורכי הסקה ולסיווג ניתן לחלק לשני סוגים [9]. הסוג הראשון הוא אלגוריתמים מדויקים המבוססים על חישוב מלא של משפט בייס תוך ניצול הרשת הבייסיאנית לחיסכון בזמן החישוב. היתרון הגדול של אלגוריתמים אלו הוא מידת הדיוק שלהם ופשטותם. אולם הם יעילים רק עבור רשתות בייסיאניות קטנות יחסית בגלל מספר חישובי ההסתברות הרב הנדרש. עבור מערכות מורכבות, רמת הסיבוכיות גבוהה מידי. לצורך זה פותח הסוג השני של אלגוריתמים המבצעים קירובים שונים החוסכים זמן חישוב ומאפשרים תוצאה מקורבת אבל עדיין בתחום הדיוק המבוקש.

3.1.6 למידה של רשת בייסיאנית

בניית רשת בייסיאנית המתארת בעיה אמיתית דורשת לרוב מאמץ רב. יש צורך במומחה המכיר היטב את תחום הבעיה. מומחה זה צריך למלא את כל טבלאות ההסתברות המותנה בצורה מדויקת. כמובן שלא תמיד ניתן למצוא מומחה כזה, אולם לעיתים קרובות ניתן להשיג דוגמאות רבות לבעיה. דוגמה כזו מורכבת מווקטור שלם או חלקי של ערכים של המשתנים המקריים. שתי עובדות אלו, הקושי או אפילו אי יכולת למצוא מומחה והיכולת למצוא דוגמאות, יצרו מוטיבציה גדולה לפתח אלגוריתמים המסוגלים ללמוד רשת בייסיאנית מתוך אוסף של דוגמאות. למידה זו יכולה להיות מציאת ערכי הפרמטרים בטבלאות ההסתברות של רשת נתונה או אפילו לכלול את למידת המבנה של הרשת עצמו.

מטרת הלמידה היא למצוא רשת בייסיאנית המתארת בצורה הטובה ביותר אוסף נתון של נתונים אודות המשתנים של המערכת. כלומר לחוקר יש אוסף של נתונים המתאר מדידות שונות של המערכת ועל סמך אוסף נתונים זה הוא רוצה לבנות רשת בייסיאנית המתארת את הקשר בין הנתונים. אילו משתנים של המערכת משפיעים ישירות על אחרים ועל כן יש ליצור קשת בין הקודקודים המייצגים משתנים אלו. בין אלו משתנים אין כלל קשר. סוג אחר של למידה הוא הניסיון לבנות את טבלאות ההסתברויות של המשתנים השונים. ניתן אם כן לחלק את הלמידה לשני שלבים. בשלב הראשון יש צורך ללמוד את מבנה הרשת – אלו קודקודים מחוברים לאלו קודקודים אחרים. בשלב השני יש צורך ללמוד את טבלאות ההסתברות.

ניתן לחלק את השיטות השונות ללימוד מבנה הגרף לשתי קטגוריות מרכזיות. השיטה הראשונה מבוססת על ניתוח התלויות בין הנתונים שבבסיס הנתונים. כיון שרשת בייסיאנית בעצם מהווה מודל של תלויות בין משתנים שונים, האלגוריתמים השונים השייכים לשיטה זו, משתמשים במבחנים שונים למציאת תלויות בין המשתנים בבסיס הנתונים ובהתאם לתלויות אלו בונים את הרשת. האלגוריתם TPDA המתואר בפירוט ב-[10] הוא דוגמה לאלגוריתמים מסוג הזה. אלגוריתם זה מתבצע בשלושה שלבים. בשלב הראשון מייצרים רשת בייסיאנית מקורבת בה כל שני משתנים המופיעים בבסיס הנתונים כקשורים זה בזה מחוברים בקשת. בשלב השני מוסיפים קשתות בין משתנים על פי מבחני אי-תלות מותנה (Conditional independence test). בשלב האחרון מצמצמים את מספר הקשתות כך שיישארו רק הקשתות המחברות משתנים התלויים זה בזה. לבסוף, נקבעים הכיוונים של הקשתות.

תהליכי הלמידה המבוססים על הקטגוריה השנייה מתבצעים על ידי בדיקת אוסף הרשתות האפשריות עבור אוסף הנתונים הקיימים ובדיקת מידת ההתאמה של כל רשת כזו לנתונים. על ידי שיטות חיפוש היוריסטיות ניתן לסקור את מרחב הרשתות האפשריות, להתאים לכל רשת מספר המתאר עד כמה השימוש ברשת שנבנתה תואם לנתונים הקיימים. ולבחור את הרשת המתאימה ביותר כלומר בעלת ערך ההתאמה המכסימלי. אפשר להחיל תהליך זה על השלב הראשון בלבד על השני או על שניהם ביחד.

במסגרת עבודה זו לא אוכל לתאר את כל שיטות הלמידה הקיימות. אסתפק בדוגמה אחת חשובה שכן בהמשך נשתמש בה לבניית רשת בייסיאנית (ראה פרק 5.1.11). שיטה זו נקראת אלגוריתם K2 [11]. האלגוריתם מממש חיפוש היוריסטי המוצא את מבנה הרשת הבייסיאנית המייצג בצורה הטובה ביותר מאגר נתונים נתון. האלגוריתם מתואר באיור 3.

```
procedure K2;
{Input:
        • A set of n nodes
        • an ordering on the nodes
        • an upper bound u on the number of parents a node may have
        • a database D containing m cases.}
{Output:
        • For each node, a printout of the parents of the node.}
  for i:= 2 to n do
     πi := Ø;
     Pold := f(i, πi); {This function is computed using Equation 20.}
     OKToProceed := true;
     While OKToProceed and | πi | < u do
           z := next node in Pred(xi);
           Pnew := f(i, πi U {z});
           if Pnew > Pold then
                 Pold := Pnew;
                 πi := πi U {z};
           else OKToProceed := false;
     end {while};
     write('Node: ', xi, ' Parent of xi: ',_i);
  end {for};
end {K2};
```

איור 3: האלגוריתם K2

נתוני הכניסה של האלגוריתם הם רשימה מסודרת של משתנים (n_i) המשמשים צמתים ברשת הרצויה, מספר מקסימאלי אפשרי של הורים לכל צומת, ומאגר נתונים (D) המכיל אוסף (m) של מקרים מסודר בטבלה בה כל שורה מציינת מקרה נפרד וכל עמודה משתנה אחר מ-n המשתנים האפשריים. כאמור, רשימת המשתנים

מסודרת כך שכל משתנה יכול להשפיע על המשתנים הבאים אחריו אבל לא על משתנים המופיעים לפניו ברשימה. מטרת הרשימה המסודרת הוא להקטין את מרחב החיפוש ובכך לייעל את זמן החיפוש של האלגוריתם: כאשר מחפשים את המשתנים המשפיעים על משתנה כלשהו יש צורך לבדוק רק את המשתנים הנמצאים לפניו ברשימה. המשתנה הראשון ברשימה הוא שורש העץ ועליו לא משפיע שום משתנה אחר. האלגוריתם מתחיל במשתנה השני. לולאת ה-for שבשורות 4-17 עוברת על כל אחד מהצמתים ברשת ומייצרת את רשימת ההורים שלו (π_i) בדרך הבאה: בשלב הראשון מניחים שאין למשתנה שום הורה (שורה5) ונבדקת מידת ההתאמה של העץ לנתונים (שורה 6). מידת ההתאמה נקבעת על ידי הפונקציה $f(i,\pi_i)$ המוגדרת על ידי המשוואה הבאה:

משוואה 7:

$$f(i,\pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!$$

כאשר:

q_i – מספר האיברים ברשימת כל הצירופים האפשריים של ערכי ההורים של הצומת ה-i. (אם לדוגמא ל-v_i יש שני הורים בעלי ערכים בוליאניים אזי $q_i=4$)

π_i – רשימת ההורים של הצומת ה-i

r_i- מספר הערכים שהצומת ה-i יכול לקבל (לדוגמה אם הוא משתנה בוליאני, $r_i = 2$)

Nijk – הוא מספר המקרים ב-D בהם המשתנה v_i מקבל את הערך במקום ה-k (בווקטור הערכים האפשריים של v_i) וערכי הערכים של הורי v_i הם האיבר ה-j ברשימת הצירופים האפשריים שלהם (ראה q_i לעיל)

Nij – מוגדר על ידי $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$

משוואה זו קובעת את מידת ההתאמה שכן ככל שהפער בין הנתונים לבין המסקנה של ניתוח הרשת יותר קטן, כך תוצאת המשוואה שהיא מידת ההתאמה תהיה יותר גדולה [12].

בהמשך, בלולאה שבשורות 8-15, מוסיפים צומת אחרי צומת מהצמתים שקדמו לצומת n_i. וכל פעם נבדקת מידת ההתאמה (שורה 10) אם היא יותר גדולה מהקודמת מוסיפים את הצומת לרשימת ההורים של המשתנה הנבדק.

לאחר שנמצאה הרשת האופטימאלית לתיאור הנתונים, מתבצע השלב הבא: בניית טבלאות ההסתברות. שלב זה יותר פשוט ויכול להתבצע על ידי הערכה בייסיאנית (Bayesian estimation) [12] בדרך של חישוב תדירות הופעת הנתונים בבסיס הנתונים אותו לומדים.

3.2 ניתוח הבעיות שרשתות בייסיאניות משמשות להן פתרון

רשתות בייסיאניות נועדו בעיקר להתמודד עם חוסר וודאות אודות המערכת אותה רוצים לנתח. חוסר הוודאות של מאורע, ולכן הצורך בחישוב או הערכה של ההסתברות שלו, עשויים לנבוע משני גורמים:

1. חוסר ודאות הנובע מהאקראיות שבטבע. דוגמה לכך היא התפרקות חומר רדיואקטיבי. אין אנו יכולים לדעת באיזה אטום ספציפי ובאיזה מועד מדויק תתרחש ההתפרקות הבאה, אם כי ניתן לדעת זאת במונחים סטטיסטיים-הסתברותיים.
2. חוסר ודאות הנובע ממידע חלקי הנמצא בידינו. דוגמה לכך היא הטלת מטבע, שעשויה להסתיים באחת משתי תוצאות אפשריות, בלא שתהיה לנו דרך לחזות איזה תוצאה תצא בהטלה מסוימת, אם כי, בהינתן כל הנתונים הפיסיקליים הרלוונטיים - ניתן תיאורטית לחשב את התוצאה. דוגמה נוספת היא השאלה "האם בשבוע הבא תפרוץ מלחמה?", שעליה נדרש המודיעין לענות. אף שייתכן שהאויב יודע תשובה ברורה לשאלה זו, המידע שבידי המודיעין הוא מידע חלקי, שלפיו יש לתת הערכה להסתברות של מאורע זה.

שאלה פתוחה היא האם כל הסתברות מהסוג הראשון היא למעשה הסתברות מהסוג השני, כלומר האם האקראיות שבהתפרקות אטומים רדיואקטיביים נובעת רק מחוסר ידיעה שלנו על כל הכוחות הפועלים, או שאקראיות היא חלק בלתי נפרד מהטבע.

באופן מסורתי, נפתרה בעיית חוסר הוודאות בעזרת שיטה שקבעה מקדמי ביטחון שנועדו להבטיח מסקנות נכונות בכל מקרה. כמובן ששיטה זו מאלצת את המשתמש בה לוותר על דיוק והחלטות מסוימות בגלל הצורך במקדמי ביטחון. הסטטיסטיקה הבייסיאנית אומצה, כתחליף למקדמי הביטחון, באופן נרחב בכלי בינה מלאכותית להסקה ולסיווג הסתברותיים, שכן היא מאפשרת שימוש במידע רב יותר כשההסתברות משמשת ככלי בקרה למידת אמינותו של מידע זה.

3.2.1 יישומי רשתות בייסיאניות

לרשתות בייסיאניות יש מגוון די נרחב של יישומים. לא נוכל לספק במסגרת עבודה זו סקירה מלאה ומפורטת של כל היישומים. נדגים רק כמה מן היישומים הנפוצים כדי להדגים את כוחה של השיטה ואת מרחב הבעיות להן היא משמשת לפתרון.

השימוש הנפוץ ביותר ברשת בייסיאנית הוא ככל הנראה באשף התשובות של תוכנת האופיס של מיקרוסופט שהחל את דרכו כבר באופיס 95 ונמצא גם בכלי תמיכה נוספים של מיקרוסופט [42].

שימוש נוסף הוא במערכות מומחה כאשר רשתות בייסיאניות מאפשרות להוסיף את המימד ההסתברותי הקיים במערכות רבות. הדוגמה המפורסמת ביותר היא QMR-DT (Quick Medical Reference (decision-theoretic מערכת זו [43] יצרה גרף בן שתי שכבות האחת מחלות (600) והשנייה סימפטומים (4000) של המחלות. הקשתות ציינו את הקשר בין הסימפטומים למחלות. מטרת מערכת זו הייתה זיהוי המחלה על פי הסימפטומים שלה.

סוג אחר של מערכות מומחה הוא מערכות לניתוח כשלים ובעיות במערכות מכאניות גדולות בבתי חרושת וכדומה. נראה שהמערכת הראשונה מבוססת רשת בייסיאנית היא Vista שפותחה על ידי אריך הורביץ עבור NASA [44]. מערכת זו ניתחה בזמן אמת נתונים של מערכות הנעה של מעבורות חלל והצביעה על כשלים אפשריים ואף הציעה פתרונות אפשריים לכשלים אלו.

דוגמה נוספת לשימוש ברשתות בייסיאניות הוא בניתוח גנים למציאת קשרים ומבנים של גנים [45]. הנפח הגדול של המידע המתקבל מאיסוף כזה מקשה מאוד על גילוי תבניות פעילות משותפת. שיטות אלו הניבו זיהוי מהיר ויעיל של מנגנונים שכבר פוענחו בעבר במאמץ גדול בהרבה, והצביעו על קשרים שלא היו ידועים לפני כן.

3.3 החסרונות של רשתות בייסיאניות

לרשת בייסיאנית יש חסרון משמעותי והוא הצורך במספר חישובים הגדל אקספוננציאלי עם מספר המשתנים ולמעשה היא בעיית NP שלמה [13], ולכן בלתי ניתנת למימוש וחישוב באופן פשוט עבור מערכות גדולות ומורכבות. כמובן שבעיות אלו הן הבעיות המעניינות ומטרידות את מדעני המחשב.

חיסרון נוסף הוא הרגישות הגבוהה של רשתות בייסיאניות לאי דיוק בערכי ההסתברויות המותנות. הרגישות הזו מצטרפת לחוסר היכולת לקבוע ערכים אלו בצורה מדויקת. חוסר יכולת זו גורמת להנחות פשטניות שבדרך כלל אינן תואמות מקרים אמיתיים רבים. התוצאה היא שהרשת הבייסיאנית יוצרת מסקנות לא אמינות.

3.3.1 משתנים רציפים

רשתות בייסיאניות פותחו במקורן עבור משתנים בדידים. כאשר מדובר במשתנים בדידים, הקשרים בין משתנים ברשת הבייסיאנית מתוארים בעזרת התפלגות הסתברות משותפת בין משתנים אקראיים. עבור רשתות כאלו קיימות שיטות מדויקות ומבוססות לניתוח הקשר בין המשתנים והסקת מסקנות (לדוגמה [15]). כאשר מדובר במשתנים רציפים, המצב אינו פשוט. דרך חישוב ההסתברות המותנית לא ישימה לגבי משתנים רציפים ויש צורך במציאת פתרונות אחרים.

דרך אחת לתיאור משתנה רציף יכולה להתבסס על שימוש במשפחה מסוימת של התפלגות רציפה. אחת ההתפלגויות שנעשה בהן שימוש רב לתיאור משתנים רציפים ברשת בייסיאנית היא התפלגות גאוסיאנית [16]. בדרך זו, מניחים שלכל האירועים בעלי הערכים הרציפים יש התפלגות גאוסיאנית, ומשתמשים

בנתונים אודות האירוע כדי להגדיר את הפרמטרים של ההתפלגות הבייסיאנית. שינוי נוסף הוא הטבלה המתארת את ההסתברות המותנית שבין האירוע לבין הגורמים לאותו אירוע, שהרי לא ניתן לתאר בטבלה את כל הערכים האפשריים כפי שנעשה עבור משתנה בדיד [17]. החיסרון המרכזי של גישה זו שהיא מגבילה את ניתוח הנתונים להתפלגות אופיינית מסוימת (גאוסיאנית) שאינה בהכרח מתארת נכונה את ההתפלגות המותנית כפי שהיא במציאות, כאשר כל היתרון של הגישה הבייסיאנית הוא שהיא אינה מניחה הנחות מוקדמות אודות אופי ההסתברות. חסרון נוסף הוא שאין כיום שיטות טובות לטיפול במודל כזה ובעיקר, אין אלגוריתמים ללמידה אוטומטית של הרשת במצב זה. קושי גדול יותר הוא להתמודד עם מערכות היברידיות הכוללות גם משתנים בדידים וגם משתנים רציפים. הגישה העיקרית לטיפול במערכות אלו נקראת cg-potentials בשיטה זו מעריכים את המאפיינים של הפונקציה הגאוסיאנית ראה תיאור מפורט יותר ב-[18]. החסרונות של שיטה זו הם שניים: ראשית, חוסר הדיוק של המאפיינים של ההתפלגות הגאוסיאנית של המשתנים הרציפים. ושנית בלתי אפשרי לטפל בשיטה זו במשתנה בדיד התלוי במשתנה רציף. קיימות גם שיטות נוספות המבוססות על קירובים אך הן אינן מדויקות ולכן לא שמישות.

דרך אחרת להתמודד עם משתנים בעלי ערכים רציפים היא לבצע דיסקרטיזציה, כלומר לחלק את התחום הרצוף של ערכי המשתנה לקבוצות של ערכים ולהתייחס בצורה זהה לכל הערכים הנמצאים באותו תחום בצורה זהה. בצורה כזו הפכנו משתנה רציף למשתנה בדיד. החיסרון בשיטה זו הוא שאנו מאבדים מידע על המשתנה שיתכן שהוא חשוב: יתכן ששני ערכים הנמצאים באותו תחום ולכן מטופלים באותה דרך, הם, למעשה, בעלי השפעה מרחיקת לכת על המשתנים האחרים. המרה זו היא בדרך כלל שרירותית מידי ולכן פוגעת באמינות התשובות שהרשת הבייסיאנית מספקת [19]. כדי להתמודד עם חיסרון זה פותחו שיטות רבות המבצעות דיגיטציה כזו בצורה מתוחכמת המאפשרת מינימום של איבוד מידע [20] ומקטינות את חוסר הדיוק. ברור ששיטות אלו מגדילות את הסיבוכיות של תהליך הלמידה וממילא מקטינות את גודל הרשתות בהן ניתן לטפל. דרך אחרת לבצע דיסקרטיזציה היא להשתמש במומחים המכירים היטב את המערכת ומסוגלים לאפיין את התחומים בהם אין משמעות רבה להבדל בין הערכים הרציפים [20]. כמובן שדרך זו אפשרית רק כאשר קיימים מומחים המסוגלים לספק מידע זה.

כפי שנראה בפרק 5.2 הלוגיקה העמומה מספקת דרך להתמודד עם קושי זה בכך שההמרה לערכים בדידים מתבצעת בעזרת פונקציות עמומות שמשמרות את המידע הקיים בנתונים הרציפים באופן מסוים.

4 השוואת מרחב הבעיות של רשת בייסיאנית ולוגיקה עמומה

כפי שראינו שתי השיטות פותחו במקורן למטרות שונות לחלוטין. בעוד שלוגיקה עמומה נועדה לחקות את החשיבה האנושית הרי שהרשת הבייסיאנית נועדה להתמודד עם חוסר הוודאות שלנו אודות העולם האמיתי. לכאורה אין שום קשר בין מרחב הבעיות שרשת בייסיאנית נועדה לפתור לבין מרחב הבעיות שלוגיקה עמומה מסוגלת להתמודד איתן. ולמרות זאת, שתי השיטות נוצלו להתמודדות עם בעיות שונות לחלוטין מיעדן המקורי. התוצאה היא שיש לא מעט בעיות ששתי השיטות מצליחות לתת להן פתרון. מה הם התחומים בהם שתי השיטות יכולות לעזור? – יש לציין ששתי השיטות מתמודדות עם חוסר בהירות ואי וודאות ולכן יש בהחלט בעיות ששתי השיטות יכולות לפתור. מטבע הדברים מפתחי הלוגיקה העמומה באים מעולם שונה לחלוטין מעולם הסטטיסטיקה הבייסיאנית וההיפך ולכן אין כמעט מחקרים המשווים בין שתי השיטות.

דוגמה להשוואה בין שתי השיטות אפשר למצוא ב-[21]. במחקר זה נבנה מודל האמור לבדוק את מידת ההתאמה של נהרות לשמש כבית גידול לצב מים מסוים. הנהרות הוגדרו על פי קצב הזרימה בהם ועל פי העומק של המים. המאמר מתאר שלוש שיטות למידול, הראשונה אינה מעניינו (היא מבוססת על ממוצע משוקלל) ושתי השיטות האחרות הינן לוגיקה עמומה ומודל בייסיאני. בשתי השיטות כל משתנה (עוצמת זרימה ועומק) חולק לשלוש קבוצות (טוב, בנוני וגרוע ביחס למידת ההתאמה לדרישות הצבים) כך נוצרות תשע מחלקות אפשריות (זרימה טובה-עומק טוב, זרימה טובע-עומק בינוני וכן הלאה). בשני המודלים השתמשו בהגדרות זהות להגדרת העומק ועוצמת הזרימה. ניתוח התוצאות של המודלים השונים מראה שבאופן כללי שתי השיטות נותנות חיזוי דומה והן דומות זו לזו אך בפרטים היו הבדלים. המאמר הנזכר מתאר את ההבדלים אולם אינו מנתח אותם ומסביר את מקורם.

שתי השיטות משמשות גם למידול של מערכות לצורך סימולציה. מודל ממוחשב של מערכת הוא ניסיון לייצג מאפיינים מסוימים בהתנהגותה של מערכת, למטרות שונות. המודל מנסה לפשט את המערכת ולקחת בחשבון רק את המאפיינים המשפיעים על התנהגות המערכת החשובה לצרכינו. הצורך בשיטות מידול כמו רשת בייסיאנית או לוגיקה עמומה נובע מכך שבדרך כלל אין מידע מלא אודות הנתונים של המודל. מידע חלקי או עמום דורש שיטות מתאימות להשלמת המידע כדי שהמודל יהיה שלם ויהיה ניתן להשתמש בו לתיאור המערכת.

שתי השיטות מספקות את היכולת לבנות מודלים במצב בו המידע לא ודאי, אבל לרשת בייסיאנית יש יתרון משמעותי ביותר. יתרון זה הוא קיומן של שיטות שונות המאפשרות לימוד מבני של הנתונים. כלומר זיהוי הקשרים הסיבתיים בין מאגר נתונים גדול ללא ידע מוקדם [22]. לעומת זאת פיתוח רשת לוגיקה עמומה דורש ידע מלא קודם אודות מערך הכללים המקשר בין נתוני הקלט ונתוני הפלט. יש צורך במומחים

המסוגלים לנסח מערך כללים המבוסס על ביטויים עמומים המתאר את התנהגות המערכת. אולם, כאמור בפרק 2.3 לא תמיד יש את הידע הנדרש לפיתוח מערך הכללים [23].

תחום נוסף המשותף לשתי השיטות הוא תחום הסיווג. כפי שהוסבר בפרק 2.4.1, סיווג הוא שיוך של אובייקטים לקבוצות על פי המאפיינים שלהם. בתחום זה רב המשותף לשתי השיטות. הן הלוגיקה העמומה והן הרשת הבייסיאנית משתמשות במחלקות רכות (soft classes) כאשר כל איבר לא שייך למחלקה אחת בלבד באופן חד-משמעי כמו בתורת הקבוצות הקלאסית. משתנה יכול להיות שייך לקבוצה באופן חלקי וכתוצאה מכך, יכול להשתייך ליותר ממחלקה אחת עם מספר המציין את מידת שייכותו למחלקה זו. רשת בייסיאנית ולוגיקה עמומה נחלקות במשמעות של "מידת השייכות" לקבוצה. בלוגיקה עמומה, משמעות מידת השייכות החלקית היא עמימות בהגדרת הקבוצה לעומת זאת ברשת בייסיאנית מידת השייכות החלקית מתארת את ההסתברות לשייכות מלאה בקבוצה [24].

ההבדל הזה יוצר הבדל בדרך בה מחושבת מידת השייכות הזאת ובפרשנות שלה. כך לדוגמה, להבדל זה יש משמעות מעשית בחישוב השייכות של קבוצה שהיא חיתוך של שתי קבוצות אחרות. בשיטה הבייסיאנית החיתוך הוא מכפלת ההסתברויות ואילו בלוגיקה עמומה אין דרך חד משמעית לחישוב החיתוך ויש מספר פונקציות לחישוב זה. הדרך המקובלת היא פונקצית MIN הלוקחת את הערך המינימאלי שבין שתי הפונקציות עבור כל ערך בתחום ההגדרה של הפונקציה.

אחת הדרכים לחיקוי תהליך החשיבה האנושי מכונה "היסק מבוסס מקרים" (case-base reasoning). העיקרון הוא פשוט: כאשר אנו נתקלים במצב לא צפוי, אנו נזכרים במקרים דומים למצב החדש. דרך הפעולה שלנו תהיה לחקות עד כמה שניתן את המצבים המוכרים לנו. תוכנה מבוססת CBR מנסה לחקות גישה זו על ידי בניית בסיס ידע רחב עד כמה שניתן המבוסס על ידע וניסיון של מומחים, ניתוח הבעיה העומדת בפני התוכנה וזיהוי דרך הפעולה הרצוי תוך שימוש בבסיס הידע הקיים. איסוף המידע וארגונו הוא אם כך, שלב מאד חשוב בתהליך. כדי שאפשר יהיה להשוות את המקרה המטופל למקרים הקיימים בבסיס הידע, יש צורך לסווג מקרים אלו לקבוצות שונות בעלות קשר ענייני בין המקרים בכל קבוצה. סיווג בייסיאני (Bayesian classifier) ולוגיקה עמומה הן שתי דרכים לביצוע סיווג זה. סיווג בייסיאני בעזרת קשרים סטטיסטיים בין אירועים שונים, ולוגיקה עמומה בעזרת קבוצות עמומות ופונקציות שייכות. היתרון הגדול של סיווג בייסיאני הוא שקיימות שיטות המאפשרות אוטומציה של בניית הרשת הבייסיאנית (ראה פרק 3.1.6). אולם החיסרון הגדול של סיווג בייסיאני הוא שהוא מתקשה מאד בטיפול בנתונים בעלי ערכים רציפים.

ראינו אם כן שיש לא מעט בעיות שניתנות לפתרון הן על ידי לוגיקה עמומה והן על ידי רשת בייסיאנית, לכן באותם מקרים בהם שתי השיטות יכולות לספק פתרון יש צורך לשאול באיזו גישה כדאי להשתמש לפתרון הבעיה. כיצד אם כן ניתן לקבוע איזו גישה עדיפה? – קשה לתת תשובה חד משמעית לשאלה זו. כאמור לעיל, מעט מאד מחקרים נעשו סביב שאלה זו. למעשה, מצאתי מחקר אחד בלבד שעשה השוואה כזו בין שתי

השיטות (כנזכר לעיל) וגם הוא לא ענה לשאלה זו אלא הראה שניתן לפתור אותה בעיה בעזרת שתי הגישות. יש אמנם מצבים בהם קל יחסית לענות על שאלה זו. לדוגמה, מערכת לוגיקה עמומה מורכבת ממערך כללים המנוסח בצורה מובנת אף לאדם שאינו בקי בשפות מחשב. לכן כאשר יש צורך בפתרון שבני אדם יוכלו בעזרתו להבין את ההתנהגות הפנימית של המערכת, את מערכת היחסים בין המרכיבים שלה, ולא רק הסתפקות ביחס שבין משתני הכניסה למשתני היציאה. במצבים כאלו, לוגיקה עמומה תספק פתרון יותר מוצלח, שכן יתרונה הגדול הוא שמערך הכללים העמומים מספק יכולת הבנה אנושית של המבנה הפנימי של המערכת. רשת בייסיאנית לעומת זאת לא פשוטה להבנה, במיוחד לא לאנשים שהסטטיסטיקה אינה חלק מהכשרתם המקצועית.

לעומת זאת, לרשת בייסיאנית יש יכולת למידה, כלומר קיימים אלגוריתמים שאפשר לבנות באמצעותם רשת בייסיאנית מתוך אוסף נתונים גם כאשר אין ידע קודם אודות המערכת ואין מומחים המסוגלים לתאר אותה. ללוגיקה העמומה אין פתרון במצב זה שכן ללא ידע מוקדם, מלא, אודות המערכת לא ניתן לבנות מערכת לוגיקה עמומה. בהנחה שיש בסיס נתונים המתאר את התנהגות המערכת, ניתן להשתמש באלגוריתמי למידה של רשת בייסיאנית כדי לפתור את הבעיה. בפרק הבא נראה כיצד ניתן לשלב את שתי השיטות כדי להרוויח משתיהן גם ללמוד את המערכת בעזרת אלגוריתם של רשת בייסיאנית וגם לפתח מודל המובן לבני אדם.

5 דרכים לשילוב רשתות בייסיאניות ולוגיקה עמומה

עד לשלב זה ניתחנו את התכונות והמאפיינים של כל שיטה כשהיא לעצמה. בפרק זה נתאר כיצד ניתן לשלב רשת בייסיאנית עם לוגיקה עמומה כדי לייצור שיטות היברידיות חדשות. בשיטות אלו מנסים להשתמש ביתרונות של שיטה אחת כדי לבנות מערכת מהשיטה השנייה תוך ניצול יתרונות אלו כדי לחפות על החסרונות של השיטה הבסיסית.

באופן כללי ניתן לדבר על שלוש גישות אפשריות:

גישה אחת [11] שתתואר בפרק 5.1, בונה מערכת לוגיקה עמומה. כאמור לעיל החיסרון המרכזי של הלוגיקה העמומה הוא הצורך במומחים. בשיטה זו משתמשים בשיטות בייסיאניות כדי ללמוד את הקשרים בין הנתונים. לאחר שיש בידנו מידע על הקשרים בין האירועים במערכת, נעשה שימוש במידע הזה כדי לבנות את מערך הכללים של מערכת הלוגיקה העמומה. תהליך זה הוא אוטומטי ואינו מצריך ידע של מומחים אודות המערכת ובכך יתרונו הגדול לעומת בנייה רגילה של מערכת לוגיקה עמומה. מדוע לא לבנות את המודל כולו בעזרת הרשת הבייסיאנית? בגלל חוסר היכולת להבין את משמעות הרשת הבייסיאנית. כאשר קיים צורך לבנות מודל המובן לבני אדם, נבנה מערכת לוגיקה עמומה. ואם לא קיים מידע המאפשר לבנות את מערך הכללים בדרך רגילה, נשתמש ברשת בייסיאנית כדי לבנות את מערך הכללים הנדרש למערכת הלוגיקה העמומה.

גישה אחרת [25] (שתתואר בפרק 5.2) לשילוב שתי השיטות היא להשתמש בלוגיקה עמומה כדי לפתור בעיות של הרשת הבייסיאנית. הבעיה המרכזית של רשת בייסיאנית היא חוסר היכולת לטפל במשתנים רציפים. הרעיון הוא לבצע המרה של משתנים רציפים לערכים בדידים. בדרך זו השימוש בלוגיקה עמומה מסתיים לאחר ההמרה לערכים בדידים ובכך מתאפשר השימוש ברשת בייסיאנית. הרשת הבייסיאנית מפותחת אחר כך ללא תלות בלוגיקה עמומה.

הגישה הזו בעצם נותנת משמעות אחרת למושג העמימות ובעצם נותנת לו פרשנות סטטיסטית. אפשר לתאר את פונקצית השייכות במונחים הסתברותיים. מידת השייכות לקבוצה מסוימת היא ההסתברות שאדם מן השורה ישייך את האיבר לקבוצה. לדוגמה האם אדם זה גבוה או לא? אמנם גבהו הוא חד משמעי 180 ס"מ אך האם הוא "גבוה"? הלוגיקה העמומה תבנה פונקצית שייכות בה לדוגמה הערך 180 ס"מ יהיה שייך למושג "גבוה" בערך 0.7. על פי ההסבר ההסתברותי משמעות הדבר היא ש- 70% מבני האדם יתארו אדם זה כגבוה.

הגישה השלישית [23] שתתואר בפרק 5.3, חולקת על תפיסה זו. על פי גישה זו יש הבדל מהותי בין מידע עמום למידע לא וודאי ולכן יש צורך בטיפול נפרד בכל סוג של מידע ולא ניתן לשלב אותם לכדי מספר אחד. על פי גישה זו יש לשלב בצורה אינהרנטית את הפונקציות העמומות בתוך הרשת הבייסיאנית. על פי גישה זו

יש לבנות רשת בייסיאנית עמומה שתכיל שני סוגים של מידע. סוג אחד של מידע הוא מידת העמימות של האירוע, כלומר מידת השייכות של כל ערך אפשרי למחלקות השונות האפשריות לאותו אירוע (לדוגמה: "גבוה", "בינוני" ו"נמוך"). הסוג השני של המידע יהיה ההסתברות להתרחשות כל אחת מן האפשרויות הללו. שני סוגים של מידע אלו צריכים להיות קיימים בו זמנית ברשת הבייסיאנית ולא ניתן לבצע רדוקציה של אחד מהם לשני. בנוסף לכך צריכה להיות אפשרות להעביר את שני סוגי המידע הללו במעלה או במורד הרשת במהלך הניתוח שלו.

בפרקים הבאים נתאר כל גישה בצורה מפורטת ונדון ביתרונותיה וחסרונותיה של כל אחת מן הגישות הללו.

5.1 רשת בייסיאנית ללימוד מערך הכללים של מערכת לוגיקה עמומה

בפרק זה נתאר שיטה המשלבת בדרך מעניינת לוגיקה עמומה ושיטות למידה של רשת בייסיאנית [26,11]. המטרה היא לבנות מערכת לוגיקה עמומה כפי שתוארה בפרק 2.2. על פי גישה זו אין צורך במידע של מומחים אודות המערכת כדי לבנות את מערך הכללים. מערך הכללים נבנה בצורה אוטומטית תוך שימוש ברשת בייסיאנית. הבסיס לבנייה הוא אוסף של נתונים אודות המערכת. בסיס נתונים זה מקשר בין המשתנים השונים של המערכת. בעזרת אלגוריתם למידה נבנית רשת בייסיאנית. בשלב הבא מופעל אלגוריתם המתרגם את הקשרים של הרשת הבייסיאנית לכללים עמומים המרכיבים את מערך הכללים של מערכת הלוגיקה העמומה. הצורך לבנות דווקא מערכת לוגיקה עמומה נובע מהיותה מובנת לחשיבה האנושית. כאשר משתמשים במודל לתיאור מערכת, בדרך כלל לא מתעניינים במבנה הפנימי של המודל. מטרת המודל היא לקבל את משתני הכניסה כפונקציה של משתני הכניסה (מודל כזה נקרא קופסה שחורה - black box). אולם לעיתים יש צורך להבין את המבנה הפנימי של המודל כדי להבין טוב יותר את המערכת הממודלת (קופסה לבנה – white box). במצב כזה אנו רוצים מודל מובן. וכאן מתגלה חסרונה של הרשת הבייסיאנית: אין היא מובנת בצורה פשוטה. יתכן שמומחי סטטיסטיקה יוכלו לנתח רשת בייסיאנית אבל בדרך כלל מי שמנתח את המודל הוא מומחה בתחום המערכת ולא בסטטיסטיקה. עבורו מודל של מערכת לוגיקה עמומה יהיה מובן ואותו הוא יוכל לנתח.

5.1.1 תיאור השיטה

האלגוריתם המאפשר שימוש ברשת בייסיאנית לייצור מערך הכללים של המערכת מורכב מארבעה שלבים:

1. דיסקרטיזציה של מבנה הנתונים תוך שימוש בתהליך היוצר קבוצות עמומות עבור כל משתנה (שלב זה זהה למתואר בפרק 5.2 המתאר שיטות דומות). בצורה זו נוצר בסיס נתונים חדש. שלב זה הכרחי רק כאשר המערכת כוללת משתנים רציפים והוא נועד לאפשר שימוש באלגוריתמים של רשת בייסיאנית, שכאמור אינם יכולים לטפל היטב במשתנים רציפים (ראה ביתר פירוט בפרק 5.2).
2. פיתוח רשת בייסיאנית מתוך בסיס הנתונים הקיים אודות המערכת לאחר השינוי שהתבצע בו במהלך השלב הקודם. הפיתוח של הרשת מתבצע תוך שימוש באלגוריתם K2 (שתואר לעיל בפרק 3.1) זהו תהליך אוטומטי שאינו מצריך הבנה או ידע מוקדם אודות המערכת ובכך יתרונו הגדול לעומת השיטות הרגילות של פיתוח מערכת לוגיקה עמומה.
3. חילוץ כללים מהמסווג הבייסיאני שנוצר בשלב הקודם. שלב זה הוא החלק המורכב של התהליך ובו טמון החידוש שבאלגוריתם זה. בהמשך נפרט את דרך הביצוע של שלב זה.

4. בשלב האחרון, נבנית מערכת לוגיקה עמומה כאשר הפונקציות השייכות שלה נלקחות ממה שנוצר בשלב הראשון לעיל, מערך הכללים שלה מבוסס על הכללים שנוצרו בשלב הקודם ובשלב זה יוצרים גם את פונקציות ההמרה של משתני היציאה ממשתנים עמומים למשתנים רציפים המתאימים למערכת האמיתית (ראה פרק 2.2).

כאמור לעיל, השלב השלישי הוא החידוש הגדול באלגוריתם זה. השלב הראשון מתואר כאמור בפירוט בפרק 5.2 ומשותף לשיטות רבות המטפלות ברשתות בייסיאנית הכוללת משתנים בעלי ערכים רציפים. השלב השני מבוסס על האלגוריתם K2 כמקובל בלמידה של רשתות בייסיאנית ואילו השלב הרביעי אופייני לבניית מערכות לוגיקה עמומה. החידוש אם כן הוא בשלב השלישי וכמובן בשילוב שיטות מעולם הלוגיקה העמומה עם שיטות בייסיאנית. לכן אתאר כאן בפירוט את האלגוריתם שמממש את התהליך בשלב השלישי.

תהליך ההמרה של הרשת בייסיאנית לכללים עמומים מתבסס על גישת MAP (maximum A posteriori). על פי גישה זו, נבחר הערך עבורו פונקצית ההסתברות המותנית נותנת ערך מקסימאלי. בשיטה זו נוצר כלל אחד עבור כל צירוף אפשרי של משתני הכניסה, וזיהוי המחלקה נקבע על פי מאפיין המחלקה בעל ההסתברות הגבוהה ביותר (MAP). תהליך זה יכול לייצור כמות עצומה של כללים שכן בדרך כלל יש מאות או אלפי משתנים במודלים הסתברותיים וכל הצירופים האפשריים הוא בעצם מכפלת מספרי כל הערכים האפשריים. כדי להקטין את מספר הכללים, ניתן להשתמש בעובדה שבדרך כלל הרבה משתנים אינם רלוונטיים לתהליך. לכן אפשר להקטין את זמן החישוב על ידי התרכזות רק בחלקים הרלוונטיים של הרשת בייסיאנית. כדי לבצע זאת האלגוריתם משתמש בכיסוי מרקוב (Markov Blanket). כיסוי מרקוב הוא בעצם אוסף כל הצמתים שכולל את האבות של צומת מסוים, הבנים שלו והאבות האחרים של בנים אלו. הוכח שאוסף זה מספיק כדי להבין ולאפיין את התנהגות המשתנה ואין צורך לבדוק את התנהגותם של צמתים אחרים ברשת. בדרך זו מספר הכללים קטן תוך כדי תהליך חילוץ הכללים מתוך הרשת בייסיאנית. למרות תהליך זה עדיין מספר הכללים אינו אופטימאלי ובדרך כלל הוא מכיל יותר כללים מהמינימום הדרוש לביצוע סיווג נכון, לכן יש צורך להוסיף שלב נוסף לאחר סיום האלגוריתם שיוציא ממערך הכללים את הכללים המיותרים. בסיום שלב זה קיבלנו אוסף של כללים. ראוי לשים לב שהרשת הבייסיאנית ששימשה אותנו לחילוץ הכללים הכילה בעצם ביטויים עמומים. הסיבה לכך היא ששלב 1 בתהליך הפך את ערכי הנתונים של בסיס הנתונים לערכים עמומים (ערכי Vi באיור 4 הם ביטויים עמומים).

איור **4** מתאר בצורה פורמאלית את הפרוצדורה המאפשרת לייצור את מערך הכללים על ידי שימוש במסווג בייסיאני. הפרוצדורה מקבלת בתחילתה את המסווג הבייסיאני ואת המאפיין שעל פיו רוצים לסווג. כמובן שאם נדרש סיווג עבור יותר ממאפיין אחד, יש צורך להריץ אלגוריתם זה כמה פעמים כמספר המאפיינים

```
procedure BC_Rule_Extraction;
input:
        BC: Bayesian Classifier
        X1: Class attribute
output: RSR {Reduced Set of Rules}
begin
    1.  RSR <- Ø {reduced set of rules is empty}
    2.  CMB <- MB(X1) {Markov Blanket of X1}
    3.  for i:= 2 to N do
    4.    Vi <- all possible values variable Xi can assume
    5.  for i:= 2 to N do
    6.    ji <- |Vi|
    7.  RI <- 1 {rule index}
    8.  for k2:= 1 to j2 do
    9.    for k3:=1 to j3 do
    10.     …..
    11.      for kN:=1 to jN do
    12.      begin
    13.        X2 = v2k2 and X3 = v3k3 and... and XN= VNkN {rule antecedent}
    14.        use MAP to determine the class value Val_Class
    15.       define rule RRI as:
    16.          if X2 = V2k2 and X3 = V3k3 and ... and XN = vNkN then X1 = Val-Class
    17.       RSR <- RSR U {RRI}
    18.       RI <- RI + 1
    19.     end
    20. RSR <- remove_superfluous_rules(RSR)
end
```

איור 4: אלגוריתם לחילוץ כללים ממסווג בייסיאני. לקוח מתוך [11]

המעניינים אותנו. הפונקציה מחזירה RSR שהוא אוסף הכללים העמומים לפיהם אפשר לבצע את הסיווג בעזרת מערכת לוגיקה עמומה. בשורה 1 מותחל אוסף זה ובשלב זה הוא ריק. כדי למצוא את כל המאפיינים המאפשרים לסווג מקרה כלשהו, אין צורך לבדוק את כל המשתנים האפשריים שיכולים להגיע לכמה מאות. כפי שנאמר לעיל אפשר להסתפק רק במשתנים המהווים 'שמיכת מרקוב' של המאפיין. לכן שורה 2 יוצרת את CMB - מערך של המשתנים הרלוונטיים על ידי מציאת המשתנים המהווים 'שמיכת מרקוב' של המאפיין ובכך אנו מקטינים את הסיבוכיות של האלגוריתם בצורה משמעותית. בשורות 3-4 נבנה מערך Vi עבור כל אחד מהמאפיינים שמוכלים ב- CMB מערך זה מכיל את כל הערכים שמאפיין זה יכול לקבל. ערכים אלו הם ערכים עמומים כיוון שכזכור, בשלב הקודם כל המשתנים של המסווג הבייסיאני קיבלו ערכים עמומים לכן בשורות 3-4 יש מערך סופי של ביטויים עמומים שהמשתנה יכול לקבל. בשורות 5-6 מאותחלים המשתנים המכילים את מספר הביטויים העמומים עבור כל אחד מהמאפיינים שנמצאים ב-CMB. בשורות 7-19 נוצר כלל אחד עבור כל צירוף אפשרי של הביטויים העמומים עבור כל אחד מהמאפיינים שנמצאים ב-CMB.

הכלל נוצר על ידי חישוב ה-MAP עבור הצירוף המסוים של הביטויים העמומים בעזרתו נקבע לאיזו מחלקה (Class_Val) שייך המאפיין X1 בשלבים אלו אנו מגדירים את הצמתים (המשתנים) המשפיעים על המאפיין לפיו אנו רוצים לסווג את הנתונים (כיסוי מרקוב). שלב זה מבוסס על האינטואיציה לפיה ההסבר הטוב ביותר לפיסת מידע הוא מצב העולם המסתבר ביותר מבחינה סטטיסטית בהינתן המידע הזה. כלומר, כל כלל נוצר על ידי ערך המשתנה עבור נתון כלשהו והמחלקה המסתברת ביותר סטטיסטית בהינתן ערך זה. בשורה 20 מסולקים כללים בעלי נתונים עודפים ממערך הכללים. אפשר לסלק בקלות כללים שבהם X1 מקבל ערך זהה ללא תלות בערך שמשתנה מסוים מקבל. לדוגמה אם משתנה מסוים מקבל שני ביטויים עמומים אפשריים, ועבור כל אחד מהם הכלל שנוצר בשורה 16 נותן ערך זהה ל-X1 הרי שאפשר לצמצם שני כללים אלו לכלל אחד המתעלם מערכו של אותו משתנה. בצורה זו ניתן לצמצם משמעותית את מספר הכללים ב-RSR שנוצר.

5.1.2 ניתוח הבעיות ששיטה זו מועילה לפתרונן

השיטה מתאימה למצבי בהם יש צורך ללמוד על המערכת מתוך אוסף נתונים מספריים מצד אחד ויש צורך בהבנה אנושית של המודל מצד שני. יתרונה הגדול הוא הן בהיותה לומדת מנתונים מספריים והן ביכולתה ליצור מערך כללים עמומים הניתנים להבנה במושגים וחשיבה אנושיים.

לא הוגדרו מאפיינים מיוחדים של המערכות המטופלות היטב על ידי אלגוריתם זה. במבחן השוואתי לשיטה זו [11] נלקחו ארבעה מאגרי נתונים מתוך UCI Machine Learning Repository. עבור כל מאגר מידע נבנו מערכות לוגיקה עמומה בעזרת שתי שיטות, זו שלנו המשתמשת ברשת בייסיאנית לבניית המערכת, והשיטה של וונג ומנדל (Wang & Mendel (WM) method). זו שיטה מסורתית לבניית מערכות לוגיקה עמומה מתוך בסיס נתונים קיים אודות המערכת [27]. בשיטה זו המידע מסודר בצורה של צמד: נתוני כניסה ונתון היציאה התואם את נתוני הכניסה הללו. כל צמד כזה מומר לכלל עמום. בשלב הבא השיטה מציאה דרכים לביצוע אקסטרפולציה כדי לכסות מקרים בהם המידע אינו מכסה את כל האפשרויות. בשיטה זו אין צורך במידע של מומחים אנושיים אודות המערכת ואיסוף מידע אודות התנהגות המערכת כ"קופסא סגורה" (נתוני כניסה מול נתוני יציאה) מספיק כדי לבנות מערכת לוגיקה עמומה המתארת מערכת זו. השיטה של WM משמשת כאמת מידה (benchmark) לבדיקת תכונותיהם של שיטות חדשות לבניית מערכות לוגיקה עמומה.

כל אחת מהמערכות נבנתה עם שלושה, חמישה או שבעה ערכים עמומים. טבלה 1 מציגה את התוצאות המאפיינות את המערכות שנבנו בעזרת שתי שיטות אלו. נבדקו שתי איכויות של המערכות: מספר הכללים במערך הכללים וממוצע קצב הסיווגים הנכונים (Average Correct Classification Rates - ACCR).

domain	BayesFuzzy		WM	
	ACCR(%)	#Rules	ACCR(%)	#Rules
Diabetes 3	71.1	8.0	100.0	25.0
Diabetes 5	74.3	16.3	99.2	88.3
Diabetes 7	76.5	37.1	98.8	169.8
Average	**73.9**	**20.4**	**99.3**	**94.3**
Cars 3	83.9	4.6	84.8	21.6
Cars 5	63.7	13.6	79.2	53.4
Cars7	50.9	5.2	68.9	83
Average	**66.1**	**7.8**	**77.6**	**52.6**
Iris 3	95.7	7.7	100	14.7
Iris 5	94.7	21.4	100	45.6
Iris 7	94.7	42	96.7	67
Average	**95.0**	**23.7**	**98.9**	**42.4**
Machine 3	93.3	9.5	94.0	12.5
Machine 5	93.2	50.5	95.0	28.7
Machine 7	86.4	109.04	87.9	39.1
Average	**90.9**	**56.3**	**92.3**	**26.7**
Total Average	**81.5**	**27.0**	**92.0**	**54.0**

טבלה 1: תוצאות ההשוואה בין מערכות לוגיקה עמומה שנוצרו על ידי שיטת WM ועל ידי BayesFuzzy. לקוח מתוך [11].

כפי שאפשר לראות מן הטבלה, המערכות שנוצרו על ידי האלגוריתם המתואר פה, היו בממוצע בעלי מספר כללים קטן יותר מאלו שנוצרו השיטה המסורתית. מצד שני, מערכות אלו היו פחות איכותיות כלומר טעו יותר פעמים

בסיווג. לא נעשה ניסיון להסביר את ההבדלים בין בסיסי המידע השונים. קיימים מאפיינים רבים נוספים שלא נבדקו במחקר זה כמו מידת הסיבוכיות של כל אחת מן השיטות והיחס בין גודל מאגר הנתונים ואיכות הפתרון.

בעבודה נוספת [26] נערכה השוואה בין השיטה בייסיאנית לייצור מערך כללים, לבין שיטה אבולוציונית בה אלגוריתם גנטי מייצר את הכללים. ההשוואה נעשתה תוך שימוש באותו מערך נתונים בו השתמשו בעבודה שתוארה לעיל. כמו בעבודה הקודמת, גם כאן נלקחו ארבעה בסיסי נתונים ועבור כל אחד מהם נבנו 3 מערכות לוגיקה עמומה, עם 3, 5 ו-7 משתנים עמומים. בכל אחד מהמקרים היו 5 מאפיינים לכל מחלקה. המערכות נבנו בעזרת ארבע שיטות שונות: שתי שיטות מבוססות אלגוריתמים אבולוציוניים (GA I, GA II) בהם מערכת הלוגיקה העמומה נבנית בעזרת אלגוריתם אבולוציוני הבונה מערכות לוגיקה עמומה רבות ובוחר את הטובה ביותר בשיטות אבולוציוניות, השיטה המבוססת על רשת בייסיאנית (BF) והשיטה הקלאסית (WM). הנתונים מוצגים בשתי טבלאות. טבלה 2 מתארת את מספר הכללים שנוצרו בכל אחת מן השיטות. העמודה הימנית מתארת את מספר הכללים האפשריים עבור כל מערכת ושאר העמודות מתארות את מספר הכללים שנוצר בכל שיטה. טבלה 3 מציגה את איכות הסיווג של כל שיטה כאשר המספר 100 מייצג הצלחה של מאה אחוז בסיווג.

Domain	Total	GA I	GA II	BF	WM
Diabetes 3	162.0	14.2	28.4	8.0	24.0
Diabetes 5	1,250.0	51	67.4	16.3	84.4
Diabetes 7	4,802.0	38.6	75.4	37.1	159.2
MPeG 3	243.0	13.0	16.4	4.6	20.2
MPeG 5	3,125.0	36.8	46.0	13.6	46.2
MPeG 7	16,807.0	34.4	51.2	5.2	76.6
Iris 3	243.0	7.6	13.8	7.7	15.0
Iris 5	1,875.0	15.0	41.2	21.4	44.8
Iris 7	7,203.0	54.4	61.4	42.0	67.2
Machine 3	243.0	6.8	12.8	9.5	13.8
Machine 5	3,125.0	21.6	25.4	50.5	29.0
Macghine 7	16,807.0	26.0	27.8	109.04	33.0

טבלה 2: השוואת מספר הכללים במערכות לוגיקה עמומה מיוצרות אוטומטית. לקוח מתוך [26]

Domain	GA I	GA II	BF	WM
Diabetes 3	100.0	99.77	71.1	91.1
Diabetes 5	99.97	99.39	74.3	89.0
Diabetes 7	92.9	94.9	76.5	86.9
Average	**97.62**	**97.53**	**73.9**	**89.0**
MPeG 3	87.51	86.32	83.9	79.3
MPeG 5	78.69	74.12	63.7	76.8
MPeG 7	64.8	51.81	50.9	62.3
Average	**77.0**	**70.15**	**66.1**	**72.8**
Iris 3	99.71	98.7	95.7	100.0
Iris 5	100.0	100.0	94.7	100.0
Iris 7	98.32	97.3	94.7	94.7
Average	**99.34**	**98.67**	**95**	**98.23**
Machine 3	94.2	95.1	93.2	93.7
Machine 5	94.3	92.0	93.2	95.6
Machine 7	91.4	86.9	86.4	92.8
Average	**93.3**	**91.33**	**90.9**	**94.03**

טבלה 3: השוואת איכות הסיווג במערכות לוגיקה עמומה מיוצרות אוטומטית. לקוח מתוך [26]

מתוך שתי הטבלאות ניתן לראות שמצד אחד, מערכי הכללים שנוצרו בשיטה בייסיאנית היו קטנים יותר מאשר בשיטות האחרות, אבל מצד שני איכות הפתרון פחות טובה. קשה לדעת מתוך תיאור המחקר מה המקור להבדלים אלו: האם הבעיה של איכות הסיווג קיימת כבר ברשת בייסיאנית או שתהליך חילוץ הכללים הוא הבעייתי והוא יוצר פחות כללים מן הנדרש לסיווג בעל איכות יותר גבוהה. שאלה נוספת היא, האם ניתן לקבוע את נקודת האופטימום הרצויה בין איכות הסיווג לבין מספר הכללים כלומר להחליט מראש שמוכנים לוותר על אחוזי איכות לטובת מערכת קטנה יותר או שהדרישה היא לאיכות מושלמת של פתרון ולהגדיל לשם כך את מערך הכללים.

יתרון ברור בכל זאת קיים לשיטה בייסיאנית. בשיטה הבייסיאנית הקטנת מספר הכללים נעשית על ידי בחירת המאפיינים הרלוונטיים ביותר לתהליך הסיווג ולכן שיטה זו מתאימה ביותר כאשר באפליקציות בהן יש מספר גדול ומשמעותי של נתונים לא רלוונטיים. השיטה הבייסיאנית מאפשרת לסנן בקלות יחסית נתונים אלו אודות לשימוש בשמיכת מרקוב. בשיטה זו, לא רק מספר הכללים קטן אלא גם מידת המורכבות שלהם. בכך היא מאפשרת להתרכז רק בנתונים הרלוונטיים לסיווג ולהקטין בצורה משמעותית את זמן החישוב של המערכת הלוגיקה העמומה.

יתרון נוסף של השיטה בייסיאנית הוא כאשר למחלקות השונות יש יותר ממאפיין אחד, שכן במצב זה יש צורך לבנות רק רשת בייסיאנית אחת וממנה לגזור X מערכי כללים כאשר X הוא מספר המאפיינים שיש למחלקות. יתרון זה עדיין לא נבחן במחקר [26].

5.2 שימוש בלוגיקה עמומה לתיאור משתנים רציפים ברשת בייסיאנית

המערכות הקיימות בעולם כוללות, בדרך כלל, משתנים בדידים, רציפים או שילוב של שני הסוגים. כאשר רוצים להשתמש ברשת בייסיאנית לצורך מידול של מערכות כאלו, יש לקחת בחשבון עובדה זו. אולם כפי שתואר בפרק 3.3 השיטות הסטנדרטיות הקיימות כיום לבניית רשת בייסיאנית מתקשות לטפל במשתנים בעלי ערכים רציפים, ובמיוחד כאשר המערכת מכילה גם משתנים רציפים וגם משתנים בדידים. כאמור לעיל הפיתרון האופטימאלי כיום הוא לבצע דיסקרטיזציה של המשתנים הרציפים לתחומים בדידים. אולם חסרונה המרכזי של גישה זו הוא איבוד המידע הכרוך בכך, שכן כל הערכים בתחום אחד מנותחים באופן זהה ללא קשר למיקומם בתוך התחום. אם, לדוגמה, יש לנו משתנה שמתאר טמפרטורה ונבצע עליו דיסקרטיזציה ונקבע את נקודת המעבר בנקודת הרתיחה של המים, הרי ש-101 מעלות ינותחו אחרת מ-99 מעלות למרות שברור שההבדל כמעט חסר משמעות. מצד שני 90 מעלות ו-99 מעלות ינותחו בצורה זהה. לכן יש צורך לבצע דיסקרטיזציה חכמה שתגרום למינימום איבוד מידע וזאת, כמובן תוך מאמץ חישובי מינימאלי.

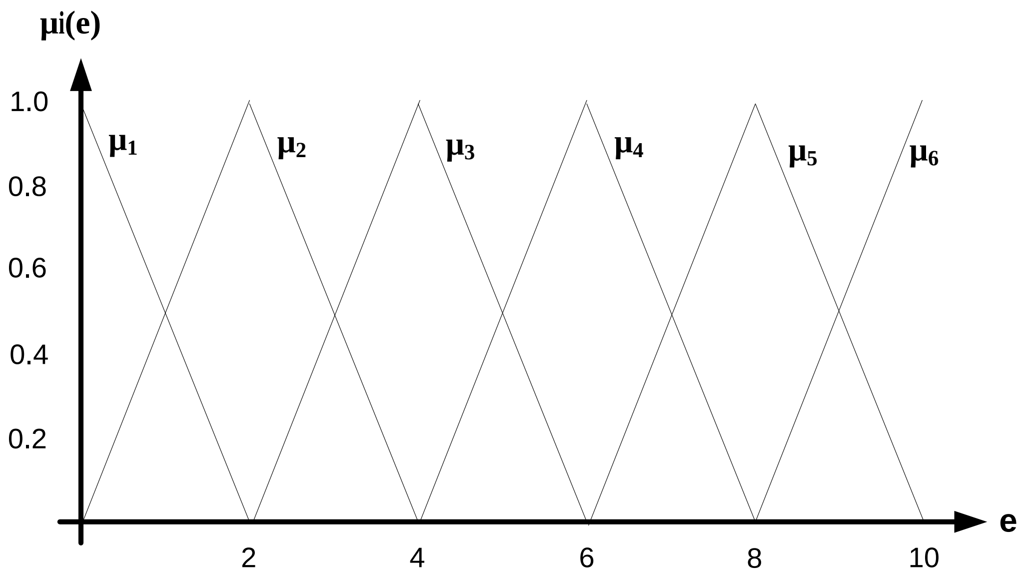
במצב זה נכנסת הלוגיקה העמומה לתמונה ומספקת פיתרון נוח ופשוט לבעיית דיסקרטיזציה: חלוקת תחום ההגדרה של המשתנה הרציף לקבוצות עמומות. בצורה זו מתבצעת דיסקרטיזציה שאינה מאבדת את כל המידע אודות נקודות הקצה של כל תחום. יש לציין ששימוש זה בלוגיקה עמומה לא נובע כלל מניסיון להגיע להבנה אנושית או מאופיים העמום של המשתנים אלא רק מהרצון לבצע דיסקרטיזציה שאינה מאבדת מידע אודות נקודות המעבר בין הקבוצות השונות. לעיתים, אופי הנתונים הקיים מחייב שימוש בשיטה זו כיוון שהנתונים הקיימים אודות המערכת הם עמומים. הסיבות לעמימות זו יכולות להיות עקב אי דיוק, אי יציבות של המדידות או כל סיבה אחרת שאינה מאפשרת מדידה מדויקת של תכונות המערכת, ולעיתים פשוט אלו הנתונים הקיימים. במצבים אלו יש צורך מראש להשתמש בשיטות של לוגיקה עמומה כדי לשלב מידע זה במודל בייסיאני המתאר את המערכת (וראה על כך ביתר פירוט בפרק הבא).

בפרק זה אתאר כיצד ניתן לבצע המרה של משתנים רציפים לערכים בדידים [28]. בדרך זו השימוש בלוגיקה עמומה מסתיים לאחר ההמרה לערכים בדידים ובכך מתאפשר השימוש ברשת בייסיאנית. הרשת הבייסיאנית מפותחת אחר כך ללא תלות בלוגיקה עמומה [29] אולם יש לשימוש במשתנים עמומים השפעה על המשוואות המרכיבות את משפט בייס הן בגלל הצורך להתאימן לאופי של משתנה עמום והן בגלל הרצון לנצל את האופי העמום כדי ליעל את החישובים ובכך לחסוך זמן ריצה ולהגדיל את רמת הסיבוך של המערכת בה ניתן לטפל.

5.2.1 תיאור השיטה

5.2.1.1 דיסקרטיזציה בעזרת לוגיקה עמומה

משתנה עמום הוא משתנה שיכול לקבל ערכים עמומים, בדרך כלל לשוניים. לדוגמה מידת החום של גוף מסוים יכולה להימדד על ידי טמפרטורה ואז המשתנה יהיה רציף לדוגמה בין (-100) מעלות צלסיוס לבין +100 מעלות צלסיוס. מדידת החום יכולה להיות משתנה עמום אם היא יכולה לקבל את אחד הערכים הבאים: {לוהט, חם מאד, פושר, קר}. במקרה זה מדובר על משתנה עמום בעל ארבעה ערכים אפשריים. הקישור בין ערכי מדידה רציפים כמו טמפרטורה לבין הביטויים העמומים נעשה בלוגיקה העמומה בעזרת פונקצית שייכות המצמידה לכל ערך מספרי את מידת השייכות שלו לערך עמום. פונקצית השייכות הפשוטה ביותר היא הפונקציה המשולשת הניתנת לתיאור בדרך גראפית כפי שמובא באיור הבא:



איור 5: דוגמה לחלוקת תחום רצוף לביטויים עמומים בעזרת פונקציות שייכות

באיור זה הערכים הרציפים האפשריים הם בין 0 לבין 10 ונעשתה דיסקרטיזציה לשישה תחומים. חלוקה זו דורשת הגדרה של שש פונקציות שייכות עבור כל אחד מהערכים העמומים האפשריים. בדוגמה זו ניתן לראות שכל ערך מספרי מאפס עד עשר יכול להיות משויך לביטוי עמום אחד או שניים בלבד. לדוגמא הערך e=3 משויך לקבוצה העמומה e2 עם 0.5 מידת שייכות וכן ל-e3 וגם כאן עם מידת שייכות של 0.5. לעומת זאת 6 משויך לקבוצה העמומה e4 עם מידת שייכות של 1.

לצורך דיסקרטיזציה של משתנה רציף, אין משמעות לביטויים העמומים אלא רק למספר הקבוצות והתכונות של פונקציות השייכות. פרמטרים אלו נקבעים על פי רמת הדיוק הנדרשת מצד אחד ורמת הסיבוכיות שנגרמת

כתוצאה מהגדלת מספר החישובים ככל שמספר הקבוצות גדל. לא מצאתי בספרות מחקרים שעסקו בפונקציות שייכות שונות בכל המחקרים שבדקתי הייתה התייחסות רק לפונקציות משולשות. אבל נעשתה השוואה לגבי התוצאות של דיסקרטיזציה למספר קבוצות משתנה.

5.2.1.2 חישוב ההסתברויות המותנות עבור משתנים עמומים

הבעיה המרכזית בשימוש במשתנים עמומים ברשת בייסיאנית היא איך להתחשב בפונקציות השייכות של המשתנים במהלך חישוב ההסתברויות של המערכת. הקושי נובע מכך שלכל משתנה יכולים להיות כמה ערכים רלוונטיים בו-זמנית. עבור משתנים בדידים, כל חישוב יכול לכלול ערך אחד בלבד. לדוגמה, אם יש משתנה המציין את כמות הגשם שיורד. אם נבחר משתנה בדיד בעל שני ערכים: יורד או לא יורד גשם, הרי שבכל חישוב של מצב המערכת יש ערך אחד בו צריך להתחשב. אפשר גם לבחור משתנה עם יותר משני ערכים: גשם עד 30mm, בין 30mm לבין 70mm, ויותר מ-70mm. גם פה ברגע שנמדוד את כמות הגשם המשתנה יקבל ערך חד-משמעי, ונוכל לנתח את התנהגות המערכת עבור ערך זה בלבד. כאשר עוברים למשתנים עמומים כל הדיון משתנה: נניח שיש לנו משתנה עמום: כמות הגשם, שיכול לקבל שלושה ערכים: "טפטוף", "גשם ממוצע" ו"מבול". אם נמדוד כמות של 10mm, על פי פונקציות השייכות יתכן שמדובר ב"טפטוף" במידת שייכות של 0.2 ו"גשם ממוצע" במידת שייכות של 0.8. בחישוב ההסתברויות ברשת הבייסיאנית יש צורך להתחשב הן בכל הערכים האפשריים (בדוגמה שלנו שניים) והן במידת השייכות של הנתון לערכים אלו (בדוגמה שלנו 0.2 ו-0.8). כיצד ניתן לבצע זאת ברשת בייסיאנית?

מצאתי בספרות המחקר שלוש שיטות שונות לפתרון בעיה זו. נתאר שלוש שיטות אלו וננסה להשוותן זו לזו.

1. הגישה האחת הופעלה כדי לבנות מערכת המלצה לבחירת מוסיקה לשמיעה בהתאם לאווירה [30]. בגישה זו, כל מדידה של המשתנים מועתקת לווקטור, בו כל איבר מציין את מידת השייכות לאחד הביטויים האפשריים. לדוגמה, מדידת 27 מעלות תתורגם לווקטור (0.4 ,0.61 ,0) כאשר יש שלושה ביטויים: קר, מתון וחם. משמעות הווקטור היא ש-27 מעלות הם 'קר' במידת שייכות 0, 'מתון' במידת שייכות 0.61 ו-'חם' במידת שייכות 0.4. לאחר שכל ערך מדידה של משתנה מומר לווקטור שייכות, ניתן לחשב את וקטור השייכות של נתוני הכניסה כולם כמכפלה של הווקטורים של כל משתנה זה בזה על פי הנוסחה הבאה:

משוואה 8

$$\begin{aligned}\text{Fuzzy Evidence} &= FMV_{node_1} \times FMV_{node_2} \times \cdots \times FMV_{node_n} \\ &= (\mu_{s_1}, \mu_{s_2}, \ldots) \times (\mu'_{s_1}, \mu'_{s_2}, \ldots) \times \cdots \\ &= ((\mu_{s_1} \times \mu'_{s_1} \times \cdots), (\mu_{s_2} \times \mu'_{s_2} \times \cdots), \cdots)\end{aligned}$$

משוואה 8 מייצרת ווקטור אחד של מידות שייכות עבור כל נתוני הכניסה. כאשר:

FMV_{modex} – הוא ווקטור המייצג משתנה עמום אחד (s_i) ו-μ,μ'.. הם מידות השייכות של הנתון. כל איבר בווקטור זה מייצג את מידת השייכות עבור ביטוי עמום אחד. בעזרת ווקטור זה ניתן לחשב את ההסתברות לתוצאה מסוימת בהינתן נתוני הכניסה העמומים. החישוב יהיה סכום ההסתברויות של כל אפשרות כפול מידת השייכות המנורמלת של אפשרות זו. משוואה 9 מתארת חישוב זה:

משוואה 9

$$P(x_{target} = s_k^{target} \mid \text{Fuzzy Evidence}) = \sum_{\forall E_r} \frac{P(x_{target} = s_k^{target} \mid E_r)\mu_{E_r}(e)}{\sum_{\forall E_r} \mu_{E_r}(e)}$$

כאשר:

s_k^{target} - ההסתברות לתוצאה מסוימת

Fuzzy Evidence – נתוני הכניסה העמומים כפי שהוגדר במשוואה 8.

E_r – אפשרות כלשהי לקבל את התוצאה

$P(x_{target} = s_k^{target} \mid E_r)$ - ההסתברות של אפשרות אחת המחושבת על ידי שיטות בייסיאנית סטנדרטיות.

$\mu_{E_r}(e)$ - מידת השייכות של האפשרות E_r.

משמעות המשוואה היא שההסתברות המותנית הרגילה של כל אפשרות מוכפלת במידת השייכות של אפשרות זו ומחולק בסכום כל מידות השייכות לצורך נרמול.

כאמור, גישה זו נוסתה בפועל ונבנתה בעזרתה מערכת לבחירת מוסיקה. מערכת זו נבחנה בצורה סובייקטיבית על ידי אדם שבחן את תוצאות הבחירה. מבחן סובייקטיבי זה מקשה על השוואת גישה זו לגישות אחרות.

2. גישה אחרת [31], מגדירה מחדש את משפט בייס עבור משתנים עמומים בדרך הבאה: כאשר אנו רוצים לחשב את ההסתברות להתרחשות של אירוע של משתנה עמום ($\tilde{B}$) המותנה בהתרחשות אירוע אחר, A_j, נשתמש בהגדרה הבאה:

משוואה 10:

$$P(\tilde{B} \mid A_j) = \sum_{i \in I} \mu_{\tilde{B}}(B_i) P(A_j \mid B_i) P(B_i) \Big/ P(A_j)$$

כאשר:

$\tilde{B}$ - משתנה עמום

Bi - ביטוי עמום אחד

$\mu_B(B_i)$ - פונקצית השייכות שלו.

בצורה זו פונקצית השייכות משפיעה על ההסתברות כך שככל שמידת השייכות גבוהה יותר כך ההסתברות לערך זה תשפיע יותר.

כאשר יש צורך לבדוק את ההסתברות להתרחשות אירוע רגיל המותנה בהתרחשות אירוע עמום, נגדיר את משפט בייס בצורה הבאה:

משוואה 11:

$$P(B_i \mid \tilde{A}) = \sum_{j \in J} \mu_{\tilde{A}}(A_j) P(A_j \mid B_i) P(B_i) \Big/ P(\tilde{A})$$

גם כאן השפעתה של מידת השייכות היא דומה. במשוואה זו יש צורך להגדיר מחדש את ההסתברות השולית העמומה $P(\tilde{A})$ והיא תוגדר בדרך הבאה:

משוואה 12:

$$P(\tilde{A}) = \sum_{j \in J} \mu_{\tilde{A}}(A_j) P(A_j)$$

בשלב זה אנו יכולים להגדיר את משפט בייס עבור משתנים עמומים בדרך הבאה:

משוואה 13:

$$P(\tilde{B} \mid \tilde{A}) = \sum_{i \in I} \sum_{j \in J} \mu_{\tilde{B}}(B_i) \mu_{\tilde{A}}(A_j) P(A_j \mid B_i) P(B_i) \Big/ P(\tilde{A})$$

אם ננתח משוואות אלו נגלה שכדי להשתמש בהן יש צורך לבצע את השלבים הבאים:

א. יש להגדיר עבור המשתנה העמום את מספר הביטויים העמומים האפשרי עבורו ואת המשמעות של כל ביטוי.

ב. יש להגדיר את ההסתברות המותנה של כל ביטוי כזה.

ג. יש להגדיר את פונקצית השייכות של כל ביטוי.

אפשר לראות אם כך שטבלת ההסתברויות של הרשת הבייסיאנית גדלה בצורה הבאה: עבור כל משתנה עמום, כל ביטוי נוסף מוסיף שורה עבור ההסתברויות המותנות של בניו של אותו משתנה.

יתרונו הגדול של מודל זה הוא פשטותו. אמנם הוא מגדיל את הסיבוכיות הן בזמן והן של מבנה הנתונים אולם כל שיטה המערבת לוגיקה עמומה נאלצת לכך. חסרונו המרכזי של מודל זה הוא בכך

שניתן לבצע איתו הסקה בייסיאנית רק בעזרת האלגוריתם המדויק ואין דרך להשתמש באלגוריתמים מקורבים. חסרון זה מגביל את השימוש בדרך זו רק למערכות פשוטות יחסית שמספר המשתנים בהן מצומצם. חסרון נוסף הוא שהמידע העמום (מידת השייכות) הולך לאיבוד במהלך החישוב ולא בא לידי ביטוי בתוצאה. לא מצאתי מחקרים המתארים שימוש מעשי בשיטה זו (ב-[31] יש שימוש בדוגמה של מערכת לאיתור תקלות בטורבינת גז אבל אין נתונים על המערכת הזו היא משמשת רק כדוגמה לחישובים אפשריים בעזרת מודל זה) וכך לא ניתן להעריך את יעילותה של השיטה ואת מידת אמינותה ביחס לשיטות אחרות.

3. דרך אחרת, המתוארת ב-[32], להחיל את משפט בייס על משתנים עמומים היא להתחשב בפונקציות השייכות כדי לחשב את פונקצית הניראות של המשתנים. כזכור, פונקצית הנראות מתארת את ההסתברות להתרחשות התוצאה (v) כאשר ידוע שהתרחש אירוע מסוים (c). אם התוצאה היא משתנה עמום ניתן לחשב הסתברות זו כאינטגרל על כל הערכים העמומים האפשריים של אירוע זה. הגדרה זו מתוארת במשוואה הבאה:

משוואה 14

$$P(v|c) = \int_{x \in \vartheta i} \mu v(x) f(x|c) dx$$

כאשר:

C – אירוע מסוים

V- התוצאה עבורה מחושבת ההסתברות

f(x|c) - היא פונקצית צפיפות ההסתברות המותנה של הערך x כאשר ידוע ש-c קרה.

$\mu_v(x)$ - היא מידת השייכות

על ידי הצבת משוואה 14 בתוך משפט בייס (משוואה 6) ניתן להגדיר את משפט בייס עבור משתנים עמומים:

משוואה 15

$$P(c_j \mid A = a_i) = \frac{\int_{x \in v_{li}} u_{v_{li}}(x) f(x \mid c_j) dx P(c_j)}{\sum_{k=1}^{m} \left(\int_{x \in v_{li}} u_{v_{li}}(x) f(x \mid c_k) dx P(c_k) \right)}$$

הבעיה המרכזית בשימוש במשוואה 15 היא שיש צורך לדעת את f(x|c) שהיא פונקצית הצפיפות המותנית וכמובן שמידע זה בדרך כלל לא זמין. אם יש מידע על פונקצית הצפיפות של המשתנה הרציף, ניתן להשתמש במידע זה כדי למצוא את פונקצית הצפיפות של המשתנה העמום (לתיאור מפורט של שיטות אלו ראה [33]). אולם יתרונה הגדול של דיסקרטיזציה עמומה הוא היכולת להתמודד עם מצבים שבהם לא ידועה פונקצית הצפיפות של המשתנה. פתרון אחד הוא לבצע מדידות במעבדה כך שיאפשרו הערכה מקורבת של פונקצית הצפיפות. אך פתרון זה מצריך מדידות מקצועיות ומסובכות וכמובן שלא תמיד אפשריות [32]. בעיה זו הובילה לפיתוח פתרונות המאפשרים למצוא דרכים אחרות לתיאור הנראות של ההסתברות המותנית ללא ידיעת פונקצית הצפיפות המדויקת אלא על סמך קירוב שלה. ב-[34] מוצעת המשוואה הבאה כקירוב לפונקצית הצפיפות המותנית תוך הסתמכות על ההסתברות המותנית:

משוואה 16 :

$$f(x \mid c_j) = \frac{\sum_{i=1}^{n} W(x_i)}{U} \left(\frac{u_{x_1}(x)}{W(x_1)} P(x_1 \mid C_j) + \frac{u_{x_2}(x)}{W(x_2)} P(x_2 \mid C_j) + \ldots + \frac{u_{x_n}(x)}{W(x_n)} P(x_n \mid C_j) \right)$$

כאשר:

x_i - ביטוי עמום לאחר ביצוע דיסקרטיזציה של המשתנה הרציף

$W(x_i)$ - פונקציה המתארת את גודל האינטרבל של x_i

U - מתאר את הגודל של כל המרחב של x.

בכך בעצם הנראות מחושבת כסכום משוקלל של הנראות של כל הקבוצות העמומות המאפיינות את המשתנה. השקלול הוא ביחס ישר למידת השייכות של הנתון המספרי הרציף לביטוי x_i וביחס הפוך לאינטרבל של הביטוי ביחס לתחום המלא שהמשתנה יכול לקבל. בצורה זו מקבלים שהנראות תלויה הן בפונקציות השייכות והתכונות שלהן והן בנראות של כל ביטוי עמום. מתוך המשוואה ניתן לראות שככל שמספר הביטויים העמומים רב יותר כך החישוב יהיה מדויק יותר כלומר יותר קרוב לערך האמיתי של פונקצית הצפיפות המותנית.

ראינו אם כך שלוש שיטות שונות לחישוב ההסתברות המותנית של משתנה עמום. הדבר מעיד על חוסר בהירות בקשר שבין מידת השייכות וההסתברות המותנה. הסיבה לכך היא שמדובר על תהליך מלאכותי שאינו נובע מהגדרת המערכת. בסופו של דבר מטרת ההתחשבות במידת השייכות היא לתת משמעות למיקומו של

הערך של המשתנה בתוך הרצף למרות דיסקרטיזציה ולכן צורת החישוב אינה משמעותית כל עוד היא נותנת משקל נכון למידת השייכות (כלומר, מידת שייכות גבוהה מגדילה את המשקל של הנתון).

כפי שניתן לראות, בכל חישוב יש משמעות מאוד גדולה לפונקציות השייכות. משמעות זו מעלה את השאלה כיצד ניתן לקבוע לכמה ביטויים עמומים יש לחלק את מרחב המשתנה הרצוף ומה הן פונקציות השייכות המתאימות לאופי הנתונים. בדרך כלל במחקרים העוסקים בנושא פונקציות השייכות נבחרות באופן שרירותי כשהשיקול העיקרי הוא קלות החישוב ואין ניסיון להשוות בין פונקציות שונות (בדרך כלל משתמשים בפונקציה המשולשת כפי שמתואר באיור 5). גם לקביעת מספר הביטויים העמומים עבור המשתנה לא מוצגת דרך מדעית והקביעה היחידה היא שככל שמספר הביטויים רב יותר כך דיוק החישוב יהיה גבוה יותר [32]. יש לציין שהחשיבות המרכזית בדיסקרטיזציה עמומה הוא הוספת המידע אודות מיקום הנתון על פני התחום הרציף ואפיון זה מושג גם בפונקצית שייכות משולשת פשוטה. אמנם יתכן ויש חשיבות גם למאפיינים נוספים של פונקצית השייכות. לאחרונה הוצעו דרכים מורכבות הכוללות רשת עצבית מלאכותית ואלגוריתם גנטי לקביעת המאפיינים של פונקציות השייכות [35]. הצורך בשימוש בשיטות מורכבות אלו מראה את הקושי הרב בקביעת מאפיינים אלו.

5.2.2 ניתוח הבעיות ששיטה זו מועילה לפתרונן

תיאור של מימוש רשת בייסיאנית עם דיסקרטיזציה עמומה נמצא ב-[34]. בעבודה זו נבנתה רשת בייסיאנית המשמשת לסיווג תקלות במכוניות על פי מאפיינים של התקלה. לצורך בניית הרשת, נעשה שימוש בבסיס ידע שהתבסס על מקרים שנאספו במוסך לתיקון מכוניות. המידע כלל 250 מקרים של תקלות וכלל מאפיינים של התקלה והגורם לתקלה. בעזרת 225 מקרים נבנה בסיס ידע ומערכת בייסיאנית-עמומה לאיתור גורמים לתקלות במכונית על סמך תיאור התקלה. 25 המקרים הנוספים שימשו לבדיקת האלגוריתם. הוגדרו 15 מאפיינים של התקלה, מתוכם שניים היו משתנים רציפים שנותחו בעזרת פונקציות שייכות. אלגוריתם זה הושווה לאלגוריתמים סטנדרטיים של CBR והתוצאות מופיעות בטבלה 4.

	Totally match	Partially match	None match
Traditional CBR	13	8	3
ASCA CBR	12	10	3
Fuzzy CBR	10	12	4
Fuzzy-Bayesian CBR	13	11	1

טבלה 4: הדיוק של אחזור מקרים בשיטות השונות. מתוך [34].

התוצאות מראות על הבדלים לא מובהקים שאינם מראים על יתרון מובהק לשיטה זו או אחרת. יתכן שהבעיה נעוצה בכך שאחוז המאפיינים הרציפים היה קטן (שניים מתוך 15). מאפיינים אלו גם לא היו קשורים מיידית לתקלה (גיל המכונית ומספר הקילומטרים שנסעה) ולכן השפעתם על איכות הפתרון של האלגוריתם קטנה יחסית. אולם כאשר נעשה שימוש במספר קטן יותר של מקרים לבניית בסיס הידע, מידת הדיוק של האלגוריתם בייסיאני-עמום הייתה גדולה יותר יחסית לאלגוריתמים האחרים.

אם ננסה להכליל מתוך המחקרים שנעשו, הרי שהתועלת שבדיסקרטיזציה לקבוצות עמומות מועילה כאשר מתקיימים כמה תנאים: ראשית המשתנים הרציפים צריכים להיות בעלי השפעה מרכזית על המערכת. לעיל ראינו שבמערכת לאפיון תקלות במכונית הייתה לדיסקרטיזציה העמומה תועלת מועטה ביותר. הסיבה המרכזית לכך היא, לדעתי, שהמשתנים הרציפים שנבדקו היו גיל המכונית ומספר הקילומטרים שהיא נסעה. שני משתנים אלו אמנם משפיעים על אמינות הרכב אבל אין הם מאפיינים מובהקים של תקלה זו או אחרת.

תנאי נוסף ליתרון של דיסקרטיזציה עמומה הוא חוסר מידע או יכולת להשתמש בפונקצית התפלגות ידועה המאפיינת את המשתנים הרצופים. כאשר אפשר לאפיין משתנה מסוים כבעל התפלגות ידועה (ההתפלגות שמופיעה רבות במחקרים היא התפלגות גאוסיאנית), אין יתרון מהותי לדיסקרטיזציה בכלל.

דוגמאות נוספות לעבודות שמתארות מימוש מעשי של גישה זו הן [36] העוסקת בניהול סיכונים ו- [37] העוסקת בזיהוי שגיאות דפוס.

5.3 רשת בייסיאנית-עמומה

הדיון בפרק הקודם הראה כיצד ניתן לנצל את הלוגיקה העמומה כדי לבצע דיסקרטיזציה של משתנים רציפים תוך איבוד מינימאלי של מידע. באינטרפרטציה זו של הלוגיקה העמומה בעצם נעשה שימוש במונחים עמומים כדי לתאר את חוסר הוודאות שלנו ביחס להתרחשותו של מאורע מסוים. צורת התייחסות זו בעצם לא הבדילה בין המשתנה העמום לבין ההסתברות להתרחשותו וכל המשוואות שראינו השתמשו הן במידת השייכות והן בהסתברות כדי לחשב את ההסתברות החדשה לאור המידע החדש. לאחר חישוב זה, כל התכונות ה"עמומות" של המשתנה נעלמו ואיבדו את חשיבותן. כפי שראינו, הוספת העמימות כתכונה של המשתנה הייתה בעצם פעולה מלאכותית שנועדה רק לאפשר דיגיטציה של המשתנה הרציף. החלוקה לביטויים עמומים כמו גם הגדרת פונקציות השייכות היו מלאכותיים לחלוטין ובלתי תלויים באופי המשתנה. מטרתם היחידה הייתה הדיגיטציה ולאחר שזו בוצעה, לא היה ערך עצמי לעמימות.

בלוגיקה הקלאסית אנו אומרים, "אם יורד גשם, סימן שמעונן". בלוגיקה העמומה אנו אומרים, "אם גשום במידה x, סימן שמעונן לפחות במידה y". כלומר, חשבו על מצב שבו אנו יושבים בחדר ללא חלון ומקשיבים לרעש הגשם היורד על הגג. אם יורד גשם חזק מאוד, אנחנו יכולים להסיק מכך שהשמיים מאוד מעוננים. אבל אם לא יורד גשם בכלל, איננו יכולים לדעת עד כמה השמיים מעוננים. ייתכן שהם מאוד מעוננים אבל לא יורד גשם, וייתכן שהם אינם מעוננים בכלל, וכמובן שייתכן כל מצב באמצע. בהתאם, אם יורד גשם חלש, אנחנו יודעים שהשמיים לא יכולים להיות בהירים לחלוטין.

אולם לוגיקה עמומה פותחה מסיבות אחרות לגמרי. מטרתה המרכזית הייתה טיפול במונחים או מצבים שאינם ניתנים להגדרת חד משמעית. כדי להבהיר זאת נביא דוגמה פשוטה: נניח שאנו מעוניינים לתאר מערכת הכוללת מידע אודות מזג האוויר. נניח שקיים במערכת משתנה המתאר את מצב העננות. אפשר להגדיר משתנה זה במונחי הלוגיקה הרגילה: שמים מעוננים או לא. במצב זה המשתנה יכול לקבל את הערך true או false בהסתברויות שונות. אולם כאן אפשר לשאול כמה שאלות: איך נתייחס לשמים המעוננים רק בחלקם? מה מידת הכיסוי של השמים בה ניתן למשתנה את הערך true? מה לגבי משך זמן העננות? האם שעתיים של עננות מספיקות כדי להגדיר את השמים כמעוננים ולתת למשתנה את הערך true? שאלות דומות ניתן לשאול על משתנים רבים אחרים. אם נשאר בתחום מזג האוויר: משתנה המגדיר אם גשום או לא אף הוא עמום באותה מידה: האם טפטוף קל מוגדר כגשום כמו סופה? הלוגיקה העמומה מנסה לתת תשובות לשאלות אלו על ידי הגדרה של ביטויים עמומים: ניתן לתת למשתנה הגשם אפשרות לקבל ערכים עמומים: טפטוף, גשם קל, גשם כבד וכדומה.

מה משמעות הדברים ביחס לרשת בייסיאנית? – כאשר הרשת הבייסיאנית נבנית על מנת לתאר מערכת הכוללת משתנים שהם עמומים באופיים, יש משמעות רבה להתייחסות גם לאופי העמום של המשתנים. אם נשתמש בדוגמאות הקודמות, אם נבנה רשת בייסיאנית בה המשתנה "מעונן" קשור הסתברותית למשתנה "גשום" הרי שיש ערך לשאלות מהסוג: "מה ההסתברות שגשם קל התחולל בגלל עננות כבדה?" או האם עננות חלקית יכולה לגרום לגשם קל? ומה הסתברות לטפטוף במצב זה?" שאלות אלו לא ניתנות למענה ברשת בייסיאנית רגילה ואף לא בשיטות שתוארו בפרק הקודם שכן המידע העמום של המשתנים נעלם במהלך החישוב ולא ניתן לשחזור במעלה ובמורד הרשת הבייסיאנית. תורת ההסתברות הסטנדרטית לא מסוגלת לטפל במשתנים עמומים ויש צורך לפתח תיאוריה מתמטית מתאימה. ניסיון כזה קיים ב-[23] ולאחרונה ב-[38] ושם גם הפניה למקורות נוספים בנושא.

בפרק זה נדון בגישה המייחסת ערך עצמי לעמימות של המשתנה. התוצאה היא שיש צורך לשמר את תכונת העמימות של כל המשתנים ברשת הבייסיאנית, שכן אם יש ערך עצמי לעמימות הרי שערך זה משפיע גם על בניו של המשתנה כמו גם על הסקת המסקנות ביחס להוריו. כדי לאפשר שימוש בעמימות ברשת הבייסיאנית היה צורך להוסיף מידע זה לאלגוריתמים המטפלים ברשת הבייסיאנית. מצד שני, פותחו אלגוריתמים רבים לשימוש ברשת בייסיאנית ולכן, כדי לנצל את כל התכונות של רשת בייסיאנית עמומה נעשה ניסיון לצמצם למינימום את השינויים. בסופו של דבר, כפי שנראה בהמשך, שימור תכונת העמימות מצריך מבנה נתונים מורכב יותר ומספר חישובים רב יותר באופן משמעותי ולכן יש צורך באנליזה מדוקדקת לבחינת הצורך במידע העמום כאשר מתבצעת ההחלטה האם להשתמש במודל של רשת בייסיאנית עמומה לתיאור המערכת.

רשת בייסיאנית עמומה המבחינה בין מידת הודאות של הנתונים לבין העמימות שלהם מצויה ב-[39,25] שם יש גם תיאור של הייצוג המתמטי של רשת בייסיאנית עמומה והצגה פורמאלית שלו. לא נעשתה עבודה מעשית המדגימה שימוש בשיטה זו למימוש מערכת אמיתית.

5.3.1 תיאור המערכת

ניסיון לשמר את תכונת העמימות נעשה ב-[39,25]. הרעיון הבסיסי במחקר זה הוא שעמימות והסתברות אינן זהות זו לזו וכל משתנה יכול להיות בו זמנית גם עמום וגם לא-וודאי. לכן יש צורך לטפל בשני הפרמטרים הללו במקביל ולא לאחד אותם כפי שראינו בעבודות אחרות. לדוגמה משתנה המתאר את מצב העננות יכול להיות הן עמום: 70% עננות והן לא-וודאי: הסתברות של 50% לעננות של 70%. למרות הצורך להוסיף את העמימות למשתנים של הרשת הבייסיאנית, חשוב לשמור על מינימום של שינויים מרשת בייסיאנית רגילה ומה שיותר חשוב לאפשר שימוש באלגוריתמים המעדכנים את הרשת על פי תצפיות ואלגוריתמים ללמידה של הרשת. הרעיון הבסיסי המונח ביסודו של הפתרון הוא להגדיר התפלגות עמומה (fuzzy probability distribution). יש לא מעט דרכים להגדיר התפלגות עמומה. כדי לאפשר שימוש נוח ויעיל ברשת בייסיאנית

עמומה הוגדרה ההתפלגות העמומה כהתפלגות רגילה בתוספת פונקצית שייכות המגדירה את מידת השייכות של ההתפלגות כאשר כל התפלגות עמומה מגדירה מידת שייכות אחת בלבד. ההתפלגות העמומה קובעת מה ההסתברות לביטויים העמומים עם מידת השייכות המוגדרת בהתפלגות מסוימת. כדי להבהיר זאת נשתמש בדוגמה: נניח שיש לנו משתנה עמום בעל שלושה ביטויים עמומים אפשריים: low, mid, high, דוגמה להתפלגות עמומה של ביטוי זה יכולה להיות $\{high_{0.3}, mid_{0.5}, low_{0.2}\}_{0.2}$ הצומדיים מסמנים שמדובר בהתפלגות הסתברותית, המספר שאחר הצומדיים מסמן את מידת השייכות, והמספרים המוצמדים לכל ערך הם ההסתברות לקבלת ערך זה. משמעות הביטוי הוא שהמשתנה יכול לקבל את הערך high עם מידת שייכות של 0.2 בהסתברות של 30%, את הערך mid עם מידת שייכות 0.2 בהסתברות של 50% ואת הערך low עם מידת שייכות של 0.2 בהסתברות של 20%. בצורה כזו ניתן להגדיר אוסף של התפלגויות עמומות שכל אחת מתארת מצב אפשרי של האירוע: $\alpha = [. . . , \{f_{0.4}, t_{0.6}\}_{0.5}, . . .]$ כאשר כל ביטוי מגדיר התפלגות למידת שייכות אחרת. הסוגריים המרובעות מציינות שמדובר במשתנה עמום (להבדיל מהצומדיים שמסמנים, כאמור, התפלגות הסתברותית). כמובן שיש לתת משמעות מתאימה למידת השייכות כך שתהיה תואמת למציאות.

כעת, לאחר שהגדרנו התפלגות עמומה, ניתן לציין את ה"מצב" של אירוע כלשהו במערכת. כל אחת מהאפשרויות של האירוע יכולה להיות עמומה, כלומר לקבל יותר מערך אחד בו-זמנית. הדבר נובע מאופיו של משתנה עמום. לדוגמה, נניח שמצבו של המשתנה C נתון לנו על ידי המצב הבא:

$C = [f_{0.3}, t_{0.6}, \{f_{0.2}, t_{0.8}\}_{0.1}]$ כלומר ידוע לנו שהוא מקבל את הערך f במידת שייכות של 0.3 ואת הערך t במידת שייכות של 0.6. יחד עם זאת יש אפשרות של 20% לקבל את הערך f עם מידת שייכות של 0.1 ואפשרות של 80% לקבל את הערך t (עם אותה מידת שייכות). פירוש הדבר שמצבו הסופי של המשתנה יכול להיות אחת משתיים: או $C = [f_{0.4}, t_{0.6}]$ בהסתברות של 20% או $C = [f_{0.3}, t_{0.7}]$ בהסתברות של 80%. כפי שעולה מדוגמה זו, ההנחה במודל זה היא שמהתפלגות עמומה ניתן לגזור ערך אחד בלבד. ראוי לשים לב כבר בשלב זה שמדובר בהנחה של המודל שאינה חייבת להיות תואמת את המציאות הממודלת ולכן הנחה זו (וכפי שנראה בהמשך יש עוד הנחות כאלו) עלולה לפגוע באיכות המידול.

כיצד ניתן להשתמש בהגדרות אלו ברשת בייסיאנית עמומה? לשם כך נתאר שני תהליכים הכרחיים בבניית רשת בייסיאנית עמומה ובניצולה להסקת מסקנות על המציאות הממודלת: התפשטות קדימה והתפשטות אחורה (להגדרת המושגים, ראה במבוא).

התפשטות קדימה:

התפשטות קדימה ברשת בייסיאנית עמומה דומה מאוד להתפשטות ברשת רגילה. ברשת רגילה, כאשר יש מידע על אב, ניתן לחשב מחדש את ההסתברות המותנה למצבים השונים של הבנים שלו. גם ברשת בייסיאנית עמומה המצב דומה אלא שיש צורך לחשב את ההסתברות המותנה של הבן עבור יותר ממצב אחד שכן

בלוגיקה עמומה אנו מניחים שיש יותר מאפשרות נכונה אחת (ולכל אפשרות כזו מידת שייכות שונה). כדי לפשט את ההסבר כיצד מידע אודות אב משפיע על בניו ברשת הבייסיאנית נפריד לשני מצבים: הראשון הוא כאשר לבן יש רק אב אחד והשני כאשר יש יותר מאב אחד, ונתחיל מהמקרה הראשון הפשוט יותר.

כאשר יש רק אב אחד התהליך הוא פשוט וזהה כמעט לחלוטין להתפשטות ברשת בייסיאנית רגילה. כאשר, ברשת רגילה ידוע לנו מצבו של האב, אזי ניתן לעדכן את הטבלה המתאימה של הבן ולהתייחס רק לשורה הבודקת את ההסתברות המותנית בערך הידוע של הבן. ברשת בייסיאנית נעשית פעולה דומה רק שיש צורך להתחשב בכל הערכים האפשריים (שהרי תכונת העמימות היא שהמשתנה יכול לקבל כמה ערכים בו זמנית עם מידות שייכות שונות). המידע אודות האב עובר לבן ובנוסף מידת השייכות עוברת גם היא לבן. בסופו של התהליך, טבלת הבן תכיל מספר שורות זהה למספר השורות של האב, ובנוסף, את מידת השייכות. הפעולה מתבצעת עבור כל ערך אפשרי ללא תלות בערכים האחרים. כאן אנו נתקלים בהקלה נוספת של המודל שכן במציאות תתכן השפעה של כל ערך על כל הערכים האחרים. תכונת אי התלות נדרשת כדי לאפשר ביצוע ההתפשטות הזאת, אחרת לא ניתן לבצעה.

כדי לטפל במצב השני בו לבן יש יותר מאב אחד יש צורך לשלב את המידע של כל האבות. עבור כל מרכיב בטבלת האב תתבצע הכפלה קרטזיאנית עם כל מרכיב של האבות האחרים (מרכיב של טבלת האב מכיל את הסתברויות עבור כל אחד מהביטויים האפשריים עבורו במידת שייכות מסוימת). מידת השייכות של התוצאה תחושב על ידי product t-norm. (לדוגמה מספרית של התהליך, ראה [39]). כדי למנוע צורך בשלב נוסף של נורמליזציה של מידות השייכות, המודל מניח שסכום מידות השייכות של כל משתנה חייב להיות אחד. גם הנחה זו מגבילה את המודל ומטרתה פישוטו והקטנת הסיבוכיות. התוצאה של דרך חישוב זו תהיה שלבן יהיו שורות (ומספר חישובים) כמכפלת של האפשרויות של האבות. אם לבן יש R אבות ולכל אב i יש Xi ביטויים אפשריים אזי מספר השורות האפשריות של הבן יהיו:

משוואה 17:

$$N = \prod_{i=1}^{R} Xi$$

כמובן שמדובר במספר הולך וגדל של חישובים וגודל טבלאות שהופך את המודל ללא מעשי עבור רשתות מגודל מסוים (יש לזכור שגם רשתות בייסיאניות רגילות הן בעיות hard-NP). לכן יש צורך לנקוט בדרכים שיקטינו את מספר החישובים (בתמורה לוויתור על הדיוק כמובן). המחקר מציע ארבע דרכים להקלת החישובים נציג כאן שלוש מהן (וכמובן יתכנו שיטות נוספות) השיטה הרביעית מורכבת יותר ואינה מוסיפה תובנות משמעותיות (ראה [39]):

a. איחוד של מצבים: מיד אחרי חישוב כל המצבים של המשתנה, הם מאוחדים למצב אחד כאשר כל אחד מהמצבים משוקלל בהתאם למידת השייכות שלו. משמעות הגישה הזו היא שמשתנה עמום הופך להיות משתנה בדיד רגיל בתהליך המערבב מידת שייכות והסתברות ביחד. אנו מאבדים את העמימות של המידע (ובכך אנו חוזרים לגישה המאפיינת פתרונות מהסוג המוצע בפרק 5.2).

b. גישה אחרת היא לצמצם את מספר המצבים על ידי מחיקתם של המצבים בעלי מידת שייכות קטנה וזאת מתוך ההנחה שהשפעתם של מצבים אלו על התוצאה היא זניחה ולכן ניתן להתעלם מהם. יש שתי דרכים לקבוע אלו מצבים יוזנחו. דרך אחת היא לקחת בחשבון רק את K המצבים בעלי מידת השייכות הגבוהה ביותר. היתרון של דרך זו הוא בכך שמספר החישובים הופך להיות ידוע ותלוי ב-K בלבד ללא תלות בתצפיות ובמודל המסוים. החיסרון הוא שאנו עלולים להתעלם ממצבים בעלי מידות שייכות משמעותיים רק בגלל שערכן נמוך מאחרות (לדוגמה כאשר מידות השייכות של כל הביטויים הן כמעט שוות זו לזו). הדרך השנייה היא להתעלם מאותם מצבים שמידת השייכות שלהם נמוכה מערך סף כלשהו. בכך מבטלים את החיסרון של השיטה הקודמת אבל אז החיסרון הוא שאין יכולת בקרה על מספר החישובים (אם כל מידות השייכות מעל ערך הסף לא תתבצע כל אופטימיזציה). בשתי הדרכים הללו חייבים לבצע שלב נוסף של נרמול מידות השייכות של המצבים הנשארים כך שגם לאחר מחיקת המצבים חסרי ההשפעה עדיין סכום מידות השייכות יישאר 1.

c. גישה שלישית לצמצום מספר החישובים היא לאחד מצבים של האב. אמנם מצבי האב מחושבים באופן מלא, אך לפני השימוש בתוצאות לחישוב מצבי הבן החדשים מאוחדים בעלי מידת שייכות קרובה למצב אחד. ההסתברויות של המצב החדש יהיו שקלול של ההסתברויות של המצבים שיצרו אותו במידת השייכות שלהם. מידת השייכות של המצב החדש תהיה סכום מידות השייכות של המצבים שיצרו אותו. אמנם גישה זו דורשת תהליך חישובי יותר מורכב משתי השיטות הקודמות אבל מספר המצבים קטן בצורה משמעותית והיתרון החישובי יהיה בסופו של דבר גדול יותר.

התפשטות אחורה פירושה הסקת מסקנות אודות האב כאשר יש לנו מידע אודות מצב הבן. החישוב מבוסס על משפט בייס המאפשר להסיק מסקנות אודות הסתברות מצבי הגורם כאשר יש מידע אודות מצבי התוצאה וההסתברות המותנית בינם לבין הגורם. בדרך זו אנו יכולים להסיק מסקנות אודות ההסתברות להתרחשות מאורע האב על פי מידע אודות התרחשות הבנים. החישוב נעשה תוך שימוש בטבלאות ההסתברות המותנית ושימוש בחוק בייס. גם עבור רשת בייסיאנית עמומה ניתן להשתמש באותה דרך בדיוק. השינוי היחיד הוא שהמידע שיש לנו אודות הבן הוא עמום, לאמור המידע מכיל ייתכנות של לפחות שתי אפשרויות שונות (עם מידת שייכות שונה עבור כל אפשרות). לכן יש צורך לבצע את החישובים עבור כל אפשרות בנפרד. גם כאן איחוד של מידע של שני בנים יצריך שימוש במכפלה קרטזיאנית של כל האפשרויות של הבנים ושימוש ב- product t-norm כדי לחשב את מידת

השייכות של כל אחת מהאפשרויות של האב. כמובן שאותה בעיה של התפוצצות מספר החישובים קיימת גם פה כמו בהתפשטות קדימה וניתן להשתמש באותן שיטות אופטימיזציה גם במקרה של התפשטות אחורה [40].

5.3.2 ניתוח הבעיות ששיטה זו מועילה לפתרונן

לא נמצאו דוגמאות מעשיות לשימוש ברשת בייסיאנית עמומה כפי שהוצגה בעבודה. קיימת הצעה להשתמש בה לצורך ניתוח רשתות גנטיות (genetic regulatory networks). הטענה היא שניתוח תיאורטי מראה שלעומת הפתרונות הקיימים המאפשרים טיפול של עד כ-100 גנים הרי ששימוש במודל המוצע יאפשר טיפול בכ-1000 גנים כלומר שיפור בסדר גודל אחד. כאמור, בעת כתיבת עבודה זו, עדיין לא התפרסמה עבודה המתארת דוגמה מעשית.

הסיבוכיות הנוספת בעבודה זו לצורך טיפול בו זמנית הן בעמימות והן בחוסר הוודאות מצריכה נימוקים ברורים המסבירים את הצורך בעמימות כדי לבנות מודל נכון. יש להראות שאכן הוספת מידע בנוגע לעמימות אכן מוסיף דיוק לתוצאות ניתוח הרשת הבייסיאנית.

לרשת בייסיאנית עמומה, כפי שהוגדרה בעבודה זו, יש גם מספר חסרונות העלולים לגרום לאי דיוק וחוסר התאמה למציאות. כפי שראינו המודל מניח לפחות שלוש הנחות הנדרשות לחישובים השונים. הנחות אלו אינן בהכרח תואמות למציאות. גם הניסיונות להקטין את מספר החישובים יכולים להוביל לחוסר דיוק של המודל. יש אם כן צורך במחקר משווה שיבדוק בצורה מעשית את תוצאות הניתוח של רשת בייסיאנית עמומה לגישות אחרות סטנדרטיות יותר ויוכיח את יתרונה של רשת בייסיאנית עמומה.

ראוי לשים לב שלעיתים למשתנה יכולות להיות מספר תכונות עמומות. אם נשתמש בדוגמאות שבפרק 3.1.5 נראה שנידונו שם שתי תכונות עמומות ביחס למשתנים "מעונן" ו"גשום" ראשית העוצמה: גשם קל או כבד עננות חלקית, מלאה או כבדה וכדומה. תכונה שנייה היא משך הזמן: כל היום גשום, העננים התפזרו אחרי שעתיים וכדומה. המודל המוצע מאפשר לטפל במצבים כאלו על ידי הוספת משתנה עמום עבור כל תכונה וחיבור שלו לכל הבנים של המשתנה המקורי. כמובן שגם גישה זו מגדילה את הסיבוכיות של הרשת.

6 סיכום ומסקנות

בעבודה זו ניתחנו שתי שיטות המשמשות הרבה בבינה מלאכותית. שתי השיטות לכאורה שייכות לשני עולמות שונים לחלוטין. רשתות בייסיאניות נועדו לניתוח מערכות בתנאי אי וודאות בעוד שלוגיקה עמומה פותחה כדי לאפשר תיאור עמום, המתאים יותר לשפה האנושית, של מערכות. נראה אם כן שאין קשר בין הבעיות שרשת בייסיאנית נועדה לפתור לבין אלו שלוגיקה עמומה מתאימה לפתרונן. אולם יש מקרים לא מועטים בהם ניתן להשתמש הן ברשתות בייסיאניות והן בלוגיקה עמומה כדי לפותרן. בעיות סיווג ומידול מערכות, ניתוחן וחיזוי התנהגותן הם תחומים מרכזיים בהם שתי השיטות משמשות לפתרונן. מצב זה מעלה את הצורך למצוא קריטריונים להחלטה באיזו שיטה לבחור לצורך פתרונה של בעיה ספציפית. יש אמנם מספר רב של בעיות שבהן ניתן להכריע בשאלה זו בקלות. אם קיימים מומחים המסוגלים לתאר את המערכת ואת התנהגותה, ברור שהפתרון טמון בלוגיקה העמומה. אולם אם אין מומחים כאלו מצד אחד, וקיים בסיס נתונים המכיל אינפורמציה על המערכת מצד שני, כדאי להשתמש ברשת בייסיאנית בגלל אלגוריתמי הלמידה הקיימים המאפשרים לבנות את הרשת בצורה אוטומטית. אולם קיימים מקרים אחרים ואף הם לא מועטים בהם יהיה קשה להכריע באיזו שיטה לבחור שכן לכל שיטה יש חסרונות ויתרונות משלה ובאותם מקרים גבוליים לא ברור איזו שיטה תוביל לפתרון האופטימאלי.

בעבודה זו ניסינו להציג דרך אחרת להתמודד עם מצבי גבול בהם נראה ששתי השיטות בעלות סיכויים דומים לפתרון טוב של הבעיה. במקום לנסות לבחור את השיטה הטובה ביותר, מציעה עבודה זו גישה המשלבת את שתי השיטות לכלל שיטה חדשה אינטגרטיבית המנצלת את היתרונות של כל גישה כדי לחפות על החסרונות של הגישות האחרות. בדרך זו אין צורך לפסול אף שיטה אלא לנסות להפיק מכל אחת מהן את יתרונותיה לפתרון הבעיה.

שלוש גישות שונות לשילוב שתי השיטות הוצגו בעבודה זו. ראינו כיצד ניתן לבנות את מערך הכללים של מערכת לוגיקה עמומה בעזרת רשת בייסיאנית. ניתחנו את היכולת לבצע דיסקרטיזציה של משתנים רצופים בעזרת לוגיקה עמומה כך שניתן יהיה לבנות רשת בייסיאנית גם כאשר הבעיה אותה מנסים לפתור מכילה משתנים בעלי ערך רציף. לבסוף הראנו כיצד ניתן לבנות רשת בייסיאנית המכילה גם מידע עמום אודות המשתנים כיצד ניתן לבצע ניתוחים ברשת כזו תוך שימוש ושימור המידע אודות עמימות המשתנים.

יש הבדלים גדולים בין שלוש הגישות שהוצגו מבחינת שימושן ומצב המחקר לגבן. הגישה הראשונה הבונה מערך כללים על ידי רשת בייסיאנית נועדה למקרים מאד מוגדרים בהם אין מידע של מומחים על המערכת אולם יש בסיס נתונים אודותיה בעזרתו ניתן לבנות רשת בייסיאנית. אולם, מצד שני נדרש מודל שבני אדם יוכלו להבין ולנתח ולכן נדרשת דווקא מערכת לוגיקה עמומה. האלגוריתם המוצע בעצם מאפשר תרגום של רשת בייסיאנית למערכת לוגיקה עמומה

גם הגישה השלישית מציגה פתרון לבעיה מסוימת והיא הצורך לשמר מידע עמום אודות המשתנים ברשת הבייסיאנית. צורך זה הובהר בניתוח תיאורטי אבל לא הודגם בצורה מעשית. לכן כיוון מרכזי הנדרש בחקר גישה זו הוא למצוא דוגמאות מעשיות המשוות בין הפתרון שמציעה שיטה זו לבעיה לעומת הפתרון של רשת בייסיאנית רגילה שאינה מכילה מידע עמום. על פי השוואות אלו ניתן יהיה להבין את הכדאיות בפיתוח מערכות כאלו שכן כפי שהראנו בעבודה, עלותן החישובית גבוהה למדי. עד כה לא מצאתי בספרות המחקר דוגמה מעשית לשימוש בגישה זו. קושי נוסף הקיים בגישה זו, הוא החוסר בתיאוריה מתמטית המנתחת את ההסתברות של משתנים עמומים,מהן תכונותיה וכיצד ניתן לבצע פעולות בסיסיות כמו חישוב ההסתברות של איחוד או חיתוך מאורעות, הסתברות מותנית ועוד. אמנם נעשו מספר ניסויים לספק כזו תיאוריה [35,41]. אולם עדיין לא נבחנו לעומק איכות התורות הללו ויכולתן. נראה שרק אם פיתוח תורה כזו המוסכמת על הכול ניתן יהיה לבחון גישה זו יותר לעומק.

הגישה בעלת הפוטנציאל הגדול ביותר מבין אלו שהוצגו היא הגישה המשתמשת בלוגיקה עמומה לצורך דיסקרטיזציה של משתנים רציפים. גישה זו עונה על בעיה אקוטית הקיימת בשיטות הבניה, הלמידה והניתוח של רשת בייסיאנית שאינן מסוגלות לטפל במשתנים רציפים. אמנם הוצעו מספר דרכים להתמודד עם בעיה זו אולם לכל אחת מהן יש חסרונות רבים שהופכים אותה ללא יעילה במרבית המקרים. לעומת זאת הפתרון שהלוגיקה העמומה מציעה הוא פתרון יעיל וטוב המאפשר התמודדות פשוטה ויעילה במרבית המקרים. ואכן מספר המחקרים המטפלים בגישה זו הוא רב יחסית. גישה זו מציעה שימוש לא טריוויאלי בלוגיקה עמומה. על פי גישה זו הלוגיקה העמומה משמשת דיסקרטיזציה ואין משמעות ל"הבנה אנושית" של ביטויים עמומים (אם כי הבנה זה אפשרית). אחד הקשיים שאיתו מתמודדים החוקרים הוא הצורך לתרגם את מידת השייכות של המשתנה העמום למושגי ההסתברות המותנה המשמשים כבסיס לרשת הבייסיאנית. כפי שראינו, קיימות מספר גישות לפתרון בעיה ואין ניסיון מחקרי להשוותן זו לזו. לדעתי אין משמעות רבה לדרך החישוב העיקרון ההכרחי הוא שמידת השייכות תכתיב את המשקל של הנתון בחישוב ההסתברות ולכן השיקול המרכזי צריך להיות הקטנת הסיבוכיות שכן בנית וניתוח רשת בייסיאנית צורך משאבי חישוב רבים והדיסקרטיזציה רק מגדילה את הסיבוכיות. משקל רב יותר יש לקביעת מספר הקבוצות העמומות אליהן מתבצעת דיסקרטיזציה ולמרות שגם כאן הגדלת מספר הקבוצות מגדילה את הסיבוכיות, הרי שהגדלת הדיוק הכרוכה בכך שווה את המחיר. כמובן שיש לקבוע את מספר הקבוצות בהתאם לבעיה ולאיכות הפתרון לה מצפים.

במהלך השנים האחרונות פותחו שיטות רבות ומגוונות המרחיבות כל אחת בדרכה את יכולת הבינה המלאכותית. למרות יכולותיהן של השיטות השונות, עדיין לא נמצאה השיטה האולטימטיבית לפתרון כל הבעיות בעזרת הבינה המלאכותית. נראה שגם לא תימצא שיטה כזו בעתיד הנראה לעין. מציאות זו הובילה, מלבד הניסיון למצוא שיטות חדשות, גם לניסיונות שונים לשילוב של השיטות הקיימות לפתרון בעיה אחת. בדרך זו כל שיטה יכולה לתרום את יתרונותיה ולחפות על חסרונותיה של השיטה האחרת. דוגמה לכך

הקשורה לנושא עבודה זו מתוארת ב-[46]. במאמר זה מתוארת מערכת להערכת מהימנות המבוססת על שיטות בייסיאנית תוך שימוש בלוגיקה עמומה לדיסקרטיזציה של הנתונים לקבוצות עמומות. בנוסף נעשה שימוש ברשת עצבית מלאכותית להערכת מאפיינים, ניבוי אמינות והערכת עליות תחזוקה צפויים. ולבסוף, אלגוריתם גנטי הופעל כדי לקבוע את נקודות הגבול של פונקציות השייכות. כך לצורך בנית מודל אחד הופעלו ארבע שיטות שונות. אינטגרציה כזו של שיטות שונות מאפשרת להגיע ליכולות גבוהות יותר של פתרון בעיות. יכולות שכל שיטה כשלעצמה אינה מסוגלת להגיע אליהן. נראה שכיוון זה יכול להוביל את תחום הבינה המלאכותית להישגים חדשים שהשיטות הקיימות לא מסוגלות להשיג. הקושי המרכזי בשילוב שיטות שונות ביחד הוא הצורך במומחיות בתחומים רבים מה שמקשה על פיתוח שיטות אינטגרטיביות כאלו.

7 ביבליוגרפיה

[1] Zadeh, L.A., 1965. Fuzzy sets. Information and Control 8, 338–353.

[2] Bayes T. An Essay towards solving a Problem in the Doctrine of Chances Philosophical Transactions of the Royal Society of London 53 (1763), 370–418.

[3] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.

[4] Dubois, D., and Prade, H., Fuzzy Sets and Systems, 1980 (Academic Press, Orlando)

[5] Xingui He, Fuzzy Theories and Fuzzy Techniques in Knowledge Processing, National Defence Industry Press, Beijing, 1998.

[6] Mamdani, E. H. Application of fuzzy algorithms for control of simple dynamic plant. *Proc. Inst. Elect. Engrs*, 1974, **121**(12), 1585–1588.

[7] Jensen F.V., An Introduction to Bayesian Networks, Springer, New York, 1996.

[8] Heckerman D., A tutorial on learning with Bayesian networks, Technical report, Microsoft Research, Redmond, Washington, 1995.

[9] Friedman N., Geiger D., Goldszmidt M., Bayesian Network Classifiers, Machine Learning, v.29 n.2-3, p.131-163, Nov./Dec., 1997.

[10] Cheng J., Greiner R., Kelly J., Bell D., and Liu W., Learning Bayesian network from data: An information-theory based approached, Artificial Intelligence, vol. 137, p. 43-90, 2002.

[11] Hruschka E. R., Camargo, H. A., Cintra, M. E., Nicoletti M. C., BayesFuzzy: using a Bayesian classifier to induce a fuzzy rule base, The IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2007), pp. 1788-1793, 2007.

[12] Cooper G., Herskovitz E., A Bayesian method for the induction of probabilistic networks from data, Machine Learning, 9, pp. 309-347, 1992.

[13] Heckerman D., Mamdani A., Wellman M. P., Real-world applications of Bayesian networks, Commun. ACM 38(3), pp. 24–68, 1995.

[14] Chickering, D.M., Learning Bayesian networks is NP-Complete. In: Fisher, D., Lenz, H.J. (eds.) Learning from Data: Artificial Intelligence and Statistics, pp. 121–130, 1996.

[15] Jensen, F.V., Lauritzen, S.L., Olesen, K.G., Bayesian Updating in Causal Probabilistic Networks by Local Computations, Computational Statistics Quarterly. 4, pp. 269 – 282, 1990.

[16] Heckermann D., Geiger D,. Learning Bayesian networks: a unification for discrete and gaussian domains, In UAI ’95, pp. 274–284, 1995.

[17] Driver, E., Morrell, D., Implementation of Continuous Bayesian Networks Using Sums of Weighted Gaussians. In Besnard and Hanks, (eds). Proc. 11th Conf. Uncertainty in Artificial Intelligence, 1995.

[18] Lauritzen, S.L., Jensen F., Stable Local Computation with Conditional Gaussian Distributions. Technical Report R-99-2014. Department of Mathematical Sciences. Aalborg University DK, 1997.

[19] Baldwin J. F., Tomaso E. D., Inference and learning in fuzzy Bayesian networks, Fuzzy Systems, volume 1, pp. 630-635, 2003.

[20] Friedman, N.; Goldszmidt, M., Discretizing continuous attributes while learning Bayesian networks. Proc. of the Thirteenth Int. Conference on Machine Learning (ICML), Bari, Italy, pp. 157-165, 1996.

[21] Brookes C.J. et al., A comparison of Fuzzy, Bayesian and Weighted Average formulations of a habitat model, International Environmental Modelling and Software Society (iEMSs), 2010.

[22] Jiejun H., Heping P., Youchuan W, An Algorithm for Cooperative Learning of Bayesian Network Structure from Data, In: Shen W. et al. (Eds.): CSCWD2004, LNCS 3168, pp. 86 – 94, 2005.

[23] Fruhwirth-Schnatter, S. On fuzzy Bayesian inference, Fuzzy Sets and Systems, 60(1), pp. 41–58, 1993.

[24] Fisher, P., Fuzzy Modeling. In: Openshaw,S. and Abrahart, R., Geocomputation, London, pp. 161-186, 2000.

[25] Fogelberg, C., Belief propagation in fuzzy Bayesian networks: A worked example, In: Faily, S., Zivny, S. (eds.) Proc. Comlab. Student Conference, 2008.

[26] Cintra, M. E., Hruschka Jr., E. R., Camargo, H. A., Nicoletti M. C., Fuzzy rule base generation through genetic algorithms and Bayesian classifiers - a comparative approach, The 7th International Conference on Intelligent Systems Design and Applications (ISDA07), 2007.

[27] Wang, L., The WM method completed: a flexible fuzzy system approach to data mining, IEEE Trans. on Fuzzy Systems, vol. 11, n. 6, pp. 768-782, 2003.

[28] Jim F. Baldwin J. F., Di Tomaso E., Bayesian networks for continuous values and uncertainty in the learning process, Proceedings of the 3rd Conference of the European Society for Fuzzy Logic and Technology, Zittau, Germany, p. pp. 679-682 , 2003.

[29] Yang C. C., “Fuzzy Bayesian Inference”, 0-7803-4053- 1/97/ 1997 IEEE, pp. 2707-2712, 1997.

[30] Park H. S., et al.: A context-aware music recommendation system using fuzzy Bayesian networks with utility theory. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS, vol. 4223, pp. 970–979, 2006.

[31] Tang H., Liu S., Basic Theory of Fuzzy Bayesian Networks and Its Application in Machinery Fault Diagnosis, Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery, volume 4, pp. 132-137, 2007.

[32] Yang, C. C., Cheung K. M., Fuzzy Bayesian analysis with continuous-valued evidence, IEEE 0-7803-2559 1/95, 1995.

[33] Chou K. C., Yuan j., "Fuzzy-Bayesian Approach to Reliability of Existing Structures, Journal of Structural Engineering, Vol. 1 19, No. 1 1, November, 1993.

[34] Chiu, C.Y., Lo, C.C., Hsu, Y.X., Integrating Bayesian theory and Fuzzy logics with Case-Based Reasoning for Car-diagnosing Problems. In: 4th Fuzzy System and Knowledge Discovery, pp. 344–348, 2007.

[35] Huang, H.Z., Zuo, M.J. and Sun, Z.Q., Bayesian reliability analysis for fuzzy lifetime data. Fuzzy Sets and Systems, 157(12), 1674–1686, 2006.

[36] Eleye-Datubo A. G., Wall A., and Wang J., Marine and Offshore Safety Assessment by Incorporative Risk Modeling in a Fuzzy-Bayesian Network of an

Induced Mass Assignment Paradigm, Expert Systems with Applications, Volume 36 , Issue 6, pp. 9879-9890, 2009.

[37] Shijun Q., Agogino A. M., Song S., Wu J. and Sitarama S., A Fusion of Bayesian and Fuzzy Analysis for Print Faults Diagnosis, Proceedings of 16th Conference on Computers and Their Applications, 2001.

[38] Viertl, R., Fuzzy Bayesian Inference, D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 10–15, 2008.

[39] Fogelberg, C., et al.: Belief propagation in fuzzy bayesian networks. In: Hatzilygeroudis, I. (ed.) 1st Int'l Workshop on Combinations of Intelligent Methods and Applications(CIMA) at ECAI 2008, University of Patras, Greece, July 21–22, 2008.

[40] Pan H., Liu L., Fuzzy Bayesian networks - a general formalism for representation, inference and learning with hybrid Bayesian networks. IJPRAI 14(7), pp. 941–962, 2000.

[41] Heng, X.-C., Qin, Z.: Fpbn: A new formalism for evaluating hybrid Bayesian networks using fuzzy sets and partial least-squares. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3645, pp. 209–217, 2005.

[42] Horvitz E., Breese J., Heckerman D., Hovel D., Rommelse K., The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–98), pages 256–265, 1998.

[43] Horvitz E., Barry M., Display of information for time-critical decision making. Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI–95), pages 296–305, 1995.

[44] Shwe M.A., Middleton B., Heckerman D.E., Henrion M., Horvitz E.J., Lehmann H.P., Cooper G.F., Probabilistic diagnosis using a reformulation of the INTERNIST–1/QMR knowledge base: I. The probabilistic model and inference algorithms. Methods of Information in Medicine, 30(4):241–255, 1991.

[45] Arvestad L, Berglund A, Lagergren J, Sennblad B., Gene tree reconstruction and orthology analysis based on an integrated model for duplications and sequence

evolution. Proceedings of the Eighth Annual International Conference on Computational Molecular Biology 326–335, 2004.

[46] Yu L., Yanfeng L., Hong-Zhong H., Ming J. Z., Zhanquan S., Optimal preventive maintenance policy under fuzzy Bayesian reliability assessment environments, IIE Transactions, Volume 42, Issue 10, pages 734 – 745, October 2010.