סיכום מאמר - Explainable Shapley-Based Allocation

נכתב ע"י ליעד נג'י 205485618

1) **מבוא:**

הבעיה שמאמר זה בא לפתור היא חלוקה הוגנת בין "שחקנים", אם שחקנים יוצרים קואליציה כדי להשיג מטרה משותפת, כיצד עליהם לפצל את ההכנסות או העלות בצורה הוגנת כך שכל תשלום ששחקן ישלם יהיה פרופורציונאלי ביחס לתרומה שלו לקבוצה, כאן נעזר ב"ערך Shapley" בכדי להשיג זאת, "ערך Shapley" הוא התרומה השולית הממוצעת הצפויה של שחקן אחד לכל תת-הקבוצות האפשריות של שחקנים.

ה-"ערך Shapley" נחשב הוגן מכיוון שהוא הקצאת התמורה היחידה שעומדת באקסיומות הרצויות הבאות: ***יעילות**, **סימטריה** (ערכו של כל שחקן נקבע רק לפי ****העלויות שוליות** שלו, אם לשני שחקנים יש עלויות שוליות זהות ביחס לכל הקבוצות, אז הם צריכים לשלם אותו הדבר.), ******עקרון האפס**, ******עקרון הליניאריות**.

הבעיה שפותרים במאמר זה, זה שלפעמים אנשים לא רואים איך החלוקה לפי "ערך Shapley" היא הוגנת וכאן נייצג את הערכים כחלוקה אחרת שסכומה יהיה ה- "ערך Shapley" ובחלוקה זאת אשר תוצג למשתמש בדרך נוחה לקריאה הוא יבין מדוע החלוקה היא הוגנת וזה יעזור לו להבין למה ה חלוקה לפי "ערך Shapley" היא גם הוגנת.

דוגמה לבעיה - חלוקת תשלום של נסיעה במונית כאשר יש יותר ממקום הורדה אחד וכל השחקנים מתחילים באותה הנקודה השאלה כמה כל נוסע ישלם?

*יעילות: סכום התשלום של כל הסוכנים שווה לערך סה"כ .

** עלות שולית: העלות השולית של שחקן, j ביחס לקבוצת שחקנים, S היא התוספת שהוא מוסיף לעלות כשהוא מצטרף לקבוצה: c (S U {j}) - c(S)

*** עקרון האפס: שחקן שעבורו כל העלויות השוליות הן אפס, משלם 0.
**** עקרון הליניאריות: אם מכפילים את העלויות בקבוע – כל התשלומים נכפלים באותו קבוע, אם מחברים שתי טבלאות-עלויות – כל התשלומים מתחברים.

2) **עבודות קודמות:**

עבודה זאת שייכת לתחום Explainable AI (XAI) המטרה היא להסביר את הפלט של מערכת AI לאדם. הסבר זה חשוב כדי לאפשר לאדם לבטוח במערכת ולהבינה טוב יותר כמו כן לאפשר שקיפות של הפלט של המערכת.
העבודה שהכי קרובה למאמר זה היא המאמר של Cailloux and Endriss (Cailloux and Endriss 2016) הם מציעים את מערכת מבוססת היגיון למתן נימוקים לתוצאה הצבעות. הם גם מפתחים אלגוריתם שגוזר אוטומטית הצדקה לכל תוצאה של * **כלל בורדה**. הרעיון המרכזי של האלגוריתם הוא לפרק את פרופילי העדפות לרצף פרופילי משנה, והאלגוריתם משתמש באחת משש אקסיומות למתן הסברים לתתי הפרופילים ולשילובים שלהם.
הדמיון בין מאמר זה למאמר שלנו הוא הגישה במאמר זה שהקצאת Shapley מבוססת גם על אקסיומות, וגם במאמר זה מפרקים את המשחק הקואליציוני הנתון לתוך קבוצה של משחקי משנה, שמרכיבים יחד הסבר למשחק ה"קואליציוני" הנתון.

* **כלל בורדה:** היא שיטת בחירות שבה נבחר המועמד סך מספר הדירוגים שלו הוא הגבוהה ביותר לדוג':

מספר הבוחרים	מקום ראשון	מקום שני	מקום שלישי
1	A	B	C
7	A	C	B
7	B	C	A
6	C	B	A

(שלישי = 0, שני = 1, ראשון = 2)
כאן נראה כי מועמד C יהיה במקום הראשון כיוון שיש לו 6 הצבעות במקום הראשון ו14 במקום השני סה"כ 26 "נקודות" ויש לו יותר מהשניים האחרים .

3) **<u>הגדרות נוספות:</u>**

* **משחק נקי:** הערכים של כל משחק הם או שליליים או חיובים (אין גם שליליים וגחם חיובים באותו משחק)
* **משחק קל לחילוק:** אם ההתייחסות לכל השחקנים שאינם שחקני 0 היא שווה.
* **משחק קל להסבר:** מקיים גם משחק נקי וגם משחק קל לחילוק.

4) **<u>האלגוריתם (בעברית):</u>**
קלט: קואליציה בעלת N שחקנים וערך v אשר מייצגת את הערך של כל שחקן בכל קבוצה ששייכת לN
פלט: סדרה של פונק' אופייניות x אשר סכומן שווה לi_V אם הסבר עליהן.

* נבנה את N המשחקים כך שסכום כל המשחקים יהיה שווה v באותו מקום
* נקבל מערך שמכיל כמה כל אחד צריך לשלם בכל "משחק" אשר ייצג תרחיש מסויים שיעזור לנו להסביר מדוע החלוקה הזו הוגנת.
* נבנה הודעה מתאימה לכל תרחיש ע"י בדיקה כמה כל אחד משלם בכל תרחיש ומדוע היא הוגנת.

5) **<u>הוכחת נכונות:</u>**

(**P:** תת קבוצה של N , **S :** תרחיש , **i :** אינדקס, **x :** פונק של תרחיש)

* **סכום כל הX_i שווה לv :** לפי תכונת שאפלי של עקרון הליניאריות.
* **כל משחק קואליציוני הוא קל להסבר:**

+ראשית האלגו' מקצה ערך שאינו 0 לכל תרחיש שייך לקבוצה, אחרת הוא מקצה 0, בנוסף עבור כל שחקן בתרחיש וכל תת קבוצה P(i\N) מתקיים x(P with {i})=x(P) כך שכל שחקן i שלא שייך לS הוא שחקן 0 מנגד מתקיים גם ההיפך i ששייך ל S הוא אינו שחקן 0 כאשר x(S\{i})=0 אבל x(S) = val != 0 לכן עבור כל 2 שחקנים {i, j} ששייכים לS כך של- P שמוכל ב N\{I, j} מתקיים x(P with {j})= x(P with {i}) לבסוף עבור שתי קבוצות אלו יש 2 אפשרויות :
1) S לא מוכל בשניהם וכן הערך שלהם בv יהיה שווה ל0 ומתקיים ש- v(P1 with P2)>=v(P1) + v(P2) וגם v(P1 with P2)<=v(P1) + v(P2)
2) בלי הגבלת הכלליות, S שמוכל ב P1 אבל לא מוכל ב P2 נקבל v(P1)!=0 וגם v(P2) =0 בנוסף מכיוון שS מכיל את P1 with P2 אז v(P1 with P2) = v(P1) + v(P2) = val.
לכן אם val הוא חיובי אז x הוא super-additive אחרת אם val שלילי אז x הוא sub-additive כלומר (N, x) הוא "נקי" ומכיוון שהוא גם "קל לחילוק" אז הוא "קל להסבר"

6) **נסויים:**

על מנת להאריך כמה ההסבר שנוצר ע"י האלגוריתם של המאמר לקחו קבוצה של אנשים, המשתתפים קיבלו תחילה רקע מתאים על משחקים קואליציוניים בכלל והנחיות ספציפיות לסקר. כדי לוודא שהמשתתפים קראו והבינו את ההוראות, כל משתתף נדרש לענות נכון על חידון קצר עם ארבע שאלות כדי להמשיך. לאחר מכן הוצג בפני המשתתפים משחק קואליציוני שבו "השחקנים" כונו גופים, וערכי הפונקציה האופיינית כונו הכנסות. המשחק הורכב של שלוש ישויות, מסומנות כ-a, b, c, ולמשתתפים הוצגה טבלת ההכנסות של הגופים כאשר הם נמצאים לבד וכאשר הם משתפים פעולה אחד עם השני. למשתתפים הוצגה גם הקצאת ה- "ערך Shapley" כהצעה לחלוקת ההכנסות בין שלושת הגופים כאשר כולם משתפים פעולה.
לאחר מכן, כל משתתף קיבל הסבר ספציפי מהאלגוריתם של המאמר למשחק או הסבר כללי

(מצבי + התאמות של "ערך Shapley", ששימש כקו בסיס).
התוצאה הייתה שהמשתתפים אכן ראו באלגו' של המאמר חלוקה הוגנת יותר מאשר ראו רק את "ערך Shapley"

7) **סיכום ועבודות עתידיות:**
לסיכום המאמר מראה את הבעיה שלפעמים אנשים אינם רואים איך לפי "ערך Shapley" ומכיוון שאנו אנשי מדעי המחשב יודעים כי חלוקה בשיטה טובה לכן כדאי לנו להסביר את ההגיון מאחורי חלוקה זאת ובכך נוכל לפתור לאנשים בעיות חלוקה כאלה בקלות, במאמר זה מוצאים חלוקה אשר יכולה לסייע לאנשים להבין את ההגיון מאחורי האלגוריתם אולם הישטה הזאת עדיין לא נבדקה עם פסיכולוגים ואוליי זה משהו שהכותבים ירצו לעשות בעתיד על מנת לדעת איך לגשת לאנשים בכתיבת ההסבר על החלוקות השונות, בנוסף אוליי כדאי גם לותר על הצגת החלוקות של "משחקי" המשנה עם תמורה של אפס, אוליי אפילו כדאי לתת הסברים בתוך תהליך האלגוריתם על מנת לעזור למשתמש להבין את כיוון החשיבה.

קישור למאמר:

http://azariaa.com/Content/Publications/Explainable_allocation_SA.pdf

חומר עזר:

שיעור 8 אלגוריתם כלכליים עם אראל סגל-הלוי אוני' אריאל:

https://github.com/erelsgl-at-ariel/algorithms-5782/blob/master/08-cost-sharing/slides-1-shapley.pdf