סקירה זו היא חלק מפינה קבועה בה אני סוקר מאמרים חשובים בתחום ה-ML/DL, וכותב גרסה פשוטה וברורה יותר שלהם בעברית. במידה ותרצו לקרוא את המאמרים הנוספים שסיכמתי, אתם מוזמנים לבדוק את העמוד שמרכז אותם תחת השם deepnightlearners.

לילה טוב חברים, היום אנחנו שוב בפינתנו deepnightlearners עם סקירה של מאמר בתחום הלמידה העמוקה. היום בחרתי לסקירה את המאמר שנקרא:

Identifying Mislabeled Data using the Area Under the Margin Ranking

---

פינת הסוקר:

המלצת קריאה ממייק: כמעט חובה – (לא חובה אבל קרוב לזה 😉).

בהירות כתיבה: גבוהה

רמת היכרות עם כלים מתמטיים וטכניקות של ML/DL הנדרשים להבנת מאמר: היכרות בסיסית עם מושגי יסוד של הלמידה העמוקה (בעיקר אלו הקשורות לאימון של רשתות נוירונים).

יישומים פרקטיים אפשריים: אופטימיזציה של תהליך אימון של רשתות נוירונים עי״ זיהוי של דוגמאות מתיוגות תוך כדי האימון.

פרטי מאמר:

לינק למאמר: זמין להורדה.

לינק לקוד: כאן.

פורסם בתאריך: 23.12.2021, בארקיב.

הוצג בכנס: NeurIPS 2020.

תחומי מאמר:

- זיהוי דוגמאות בעלות לייבלים שגויים בתהליך אימון של רשתות נוירונים.

כלים מתמטיים הסימונים:

- לוגיטים (logits): פלט של השכבה האחרונה של רשת סיווג (לפני הנרמול (softmax/sigmoid).

תחומים בהם ניתן להשתמש בגישה המוצעת:

- למידה semi-supervised.
- אוגמנטציה של דאטהסטים.

תמצית מאמר:

אחד הגורמים המרכזיים שמשפיעים על ביצועים של רשתות נוירונים הינו איכות של הדאטה סט שעליו הרשת מאומנת. בחלק לא מבוטל של מדאטה סטים הלייבלים הינם "חלשים" (לא מדויקים) כי התיוג בוצע דרך שימוש במשתני פרוקסי או דרך גירוד דפי האינטרנט. זיהוי דוגמאות עם לייבלים מוטעים עשוי לשפר את יכולת ההכללה של רשת וגם תוריד את רמת הזיכרון שלה (memorization).

מכיוון שרשתות נוירונים העכשוויות הינן בעלות מספר גבוה של פרמטרים, נדרשים דאטה סטים גדולים מאוד בשביל לאמן אותן. עבור רוב הדאטה סטים לא ניתן (או מאוד יקר) לעבור עליהם במטרה לזהות דוגמאות המתויגות בצורה שגויה. עקב כך יש צורך בפיתוח גישות אוטומטיות (ללא התערבות בני אדם) לזיהוי של דוגמאות כאלו.

המאמר מציע שיטה לזיהוי אוטומטי (ללא התערבות אנושית) של דוגמאות עם לייבלים שגויים במהלך אימון של רשתות נוירונים. השיטה מנצלת את המידע על לוגיטים (logits) של דוגמאות לאורך אימון הרשת לזיהוי של דוגמאות עם לייבלים מוטעים. המטריקה שהם משתמשים בה נקראת שטח מתחת השוליים (area under the margin – AUM). יותר קונקרטית, לדוגמא נתונה AUM מודד את ההפרש הממוצע על פני כל האפוקים בתהליך אימון הרשת, בין ערכי הלוגיט של הקטגוריה המתאימה ללייבל שאיתו הדוגמא מתויגת, לבין המקסימום של כל ערכי הלוגיטים של הקטגוריות האחרות.

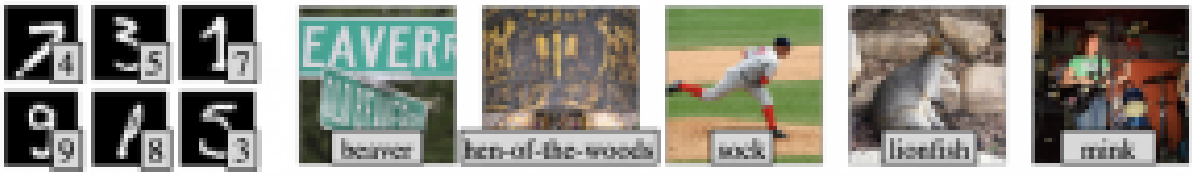

Figure 1: Images from **MNIST** (left) and **ImageNet** (right) with lowest *Area Under the Margin* (AUM) ranking (most likely to be mislabeled). AUMs are computed with LeNet/ResNet-50 models.

פינת האינטואיציה: עבור דוגמאות מתויגות נכון הפרשים אלו אמורים לעלות לאורך האימון כי יכולת ההכללה (הנבנית על סמך דוגמאות עם אותו לייבל) של הרשת עולה ככל שהאימון מתקדם. לעומת

זאת בדוגמאות המתויגות בצורה שגויה הרשת לא מצליחה לנצל את העלייה ביכולת הכללה שלה ו- AUM לא "מתרומם" ככל שהאימון מתקדם. הסיבה לכך טמונה בעובדה שהרשת "רואה" שדוגמא מתויגת עם לייבל l_wrong, דומה לדוגמאות הנושאות לייבל אחר (הנכון) l_cor ומנסה לדחוף את הלוגיט המתאים ל- l_cor למעלה שגורם לירידה במרג'ין של הדוגמא. אז באופן אינטואיטיבי AUM עבור דוגמאות "נכונות" אמור להיות יותר גבוה מזה של הדוגמאות "הלא נכונות".

המאמר מציע לנצל את האינטואיציה הזו ולזהות דוגמאות שגויות על בסיס ה- AUM שלהם. מכיוון שאנו לא יודעים מה האחוז של דוגמאות "לא נכונות" בדאטה סט נשאלת השאלה: איך נבחר את ערך הסף של AUM המבדיל בין דוגמאות נכונות ללא נכונות. המאמר מציע לבנות קטגוריה מלאכותית מתוך הדאטה סט שעל בסיס ה- AUM שלה הסף הזה נבחר.

הסבר של רעיונות בסיסיים:

כמו שכבר אמרנו עבור דוגמא x ואפוק t, השול (margin) מחושב כהפרש בין הלוגיט של הלייבל שאיתו הדוגמא מתויגת לבין המקסימום בין הלוגיטים של כל הלייבלים האחרים. AUM עבור דוגמה x מוגדר כממוצע של הפרשים אלו על פני כל האפוקים. אתם תשאלו – מה הקשר של הממוצע הזה לשטח מתחת לגרף של מרג'ינים: כדי להבין זאת מספיק להביט באיור המצורף: שטח מתחת לשול משוערך ע" הממוצע של השולים על פני האפוקים של אימון (אלו שעדיין לא הספיקו לשכוח חדו"א 1 יכולים לראות שהממוצע זה הינו סכום דרבו של הפונקציה מוגדרת על האפוקים והערכים שלה הם המרג'ינים). קל לראות שאם המרג'ין של דוגמא הינו מספר חיובי גבוה אז הרשת מצליחה לחזות נכון את הלייבל של דוגמא זאת. לעומת זאת השול שלילי גבוה מצביע על כך שהרשת רואה את הדוגמא כדומה לדוגמאות המתויגות עם לייבל אחר (החיזוי שלה יהיה כמובן לא נכון עבור דוגמא זו).

Figure 2: Illustration of the *Area Under the Margin* (AUM) metric. The graphs display logit trajectories for easy-to-learn dogs (left), hard-to-learn dogs (middle), and BIRDS mislabeled as DOGS (right). (Each plot's logits are averaged from 50 CIFAR10 training samples, 40% label noise.) AUM is the shaded region between the DOG logit and the largest other logit. Green/red regions represent positive/negative AUM. Correctly-labeled samples have larger AUMs than mislabeled samples.

הערה: מספר עבודות הוכיחו שגודל ממוצע של מרג'ין מהווה אינדיקציה ליכולת הכללה של הרשת: כלומר ככל שהשול הממוצע גבוה יותר, יכולת הכללה של הרשת עשויה להיות טסקירה זו היא חלק מפינה קבועה בה אני סוקר מאמרים חשובים בתחום ה-ML/DL, וכותב גרסה פשוטה וברורה יותר שלהם בעברית. במידה ותרצו לקרוא את המאמרים הנוספים שסיכמתי, אתם מוזמנים לבדוק את העמוד שמרכז אותם תחת השם deepnightlearners.

#deepnightlearners

הפוסט נכתב על ידי מיכאל (מייק) ארליכסון, PhD.

מיכאל עובד בחברת סייבר Salt Security בתור Principal Data Scientist. מיכאל חוקר ופועל בתחום הלמידה העמוקה, ולצד זאת מרצה ומנגיש את החומרים המדעיים לקהל הרחב