קירה 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:

Representation Learning via Invariant Causal Mechanisms

פינת הסוקר:

המלצת קריאה ממייק: מומלץ לאוהבי למידת ייצוג, בעלי ידע בסיסי בתורת הסיבתיות.

בהירות כתיבה: בינונית פלוס.

רמת היכרות עם כלים מתמטיים וטכניקות של ML/DL הנדרשים להבנת מאמר: היכרות בסיסית עם כלים מלמידת ייצוג ומתורת הסיבתיות.

יישומים פרקטיים אפשריים: שיפור ביצועים לכל שיטת למידת ייצוג המבוססת NCE.

פרטי מאמר:

לינק למאמר: זמין להורדה.

לינק לקוד: לא נמצא בארקיב.

פורסם בתאריך: 15.10.20, בארקיב.

הוצג בכנס: ICLR 2021 Poster.

תחום מאמר:

- למידת ייצוג (representation learning).
- תורת הסיבתיות.

כלים מתמטיים, מושגים וסימונים:

- גרף סיבתיות של מודל הסתברותי.

- InfoNCE - Contrastive Predictive Coding.
- לוס ניגוד - NCE.
- מרחק KL בין התפלגויות.
- עידון של משימת למידה (task refinement).

תמצית מאמר:

המאמר מציע שיטה (הנקראת RELIC) לבנייה של ייצוג של דאטה במרחב ממימד נמוך. הרעיון מהווה הכללה של InfoNCE ומתבטא בהוספת איבר רגולריזציה לפונקצית הלוס שלה. איבר רגולריזציה זה נועד "לוודא שהתפלגות הדמיונות בין הייצוגים" אינווריאנטית תחת אוגמנטציות שונות המופעלות על הדוגמאות האלו" (במאמר זה גם נקרא שינוי סגנון ואשתמש בשני המושגים האלה בהמשך הסקירה) ארחיב על כך בהמשך.

אז בואו נבין מה התוספת הזו תורמת לפונקצית לוס. קודם כל השיטות מבוססות, NCE - noise contrastive estimation, בנויות בצורה שגורמת לייצוגים של דוגמאות "קרובות" להיות קרובים גם כן (במרחב הייצוג). עבור דומיין התמונות קירבה מוגדרת כדמיון מבחינה סמנטית/תוכן. פעולות אוגמנטציה כמו הזזה, סיבוב או קרופ אינן משפיעות על דמיון (קירבה) בין ייצוגים של תמונות באופן משמעתי. איבר רגולריזציה המוצע במאמר "מאלץ" את הייצוגים, בנוסף לתכונה המתוארת מעלה, להיות אינווריאנטיים לשינויים לא סמנטיים "שאין להם השפעה על הקירבה" (קרי שינוי סגנון). במילים אחרות בהינתן הייצוגים של תמונות בעלות קירבה מסוימת ביניהם (הקירבה יכולה להיות גבוהה או נמוכה), הייצוגים של תמונות אלו לאחר האוגמנטציה "מאולצים לשמור על אותה הקירבה כמו התמונות המקוריות". זו תוספת משמעותית ללוס הרגיל של שיטות מבוססות NCE כי היא "מאלצת" את הייצוגים "לייצג את התוכן של התמונה בלבד (!!)" עם כמה שפחות תלות בסגנון של תמונה. זה מוביל לייצוג יותר רלוונטי וקורלטיבי למשימות downstream (הקשורות לתוכן) -זו בעצם הנחת יסוד של המאמר.

רעיון בסיסי:

הרעיון הבסיסי של המאמר בנוי על 3 הנחות יסוד שמאפשרות להציג את תהליך של יצירת תמונה כגרף סיבתי.

תהליך יצירת תמונה:

1. התמונה נוצרת ממשתנה לטנטי של תוכן C ומשתנה לטנטי של סגנון S.
2. המשתנים S ו- C הינם בלתי תלויים (התוכן לא תלוי בסגנון).
3. רק תוכן של תמונה רלוונטי למשימות downstream שעבורם הייצוג נבנה. סגנון של תמונה אינו רלוונטי למשימות אלו כלומר שינויי סגנון לא משפיעות על תוצאת משימה Y_t. לדוגמא במשימת סיווג עם שני קלאסים (נגיד כלבים וחתולים), איברי גוף שונים של כלבים ושל חתולים מהווים תוכן כאשר רקע, תנאי תאורה, אופיינים של עדשת מצלמה וכדומה מיוחסים לסגנון.

תחת הנחות אלו תוכן של תמונה מהווה ייצוג טוב שלה עבור משימות downstream וכתוצאה מכך המטרה של למידת ייצוג זה שערוך תוכן של תמונה. במילים אחרות, משתנה תוכן של תמונה X מכיל את כל המידע הרלוונטי לחיזוי, המבוצע במסגרת משימה Y, והוא צריך להיות אינווריאנטי (לא משתנה) תחת כל שינויים כלשהם של סגנון.

הסבר קצר על מושגי יסוד במאמר:

אחד ממושגי היסוד במאמר זה שיטות ללמידת הייצוג מבוססות NCE - בואו מרענן בקצרה את הנושא הזה:

שיטות NCE: הנחת היסוד ב- NCE מתבסס על ההנחה שייצוג חזק של דאטה בהכרח מסוגל להפריד בין זוגות של דוגמאות דומות לבין זוגות דוגמאות רנדומלית. בין השימושים של טכניקה זו אפשר להזכיר negative sampling שהשתמשו בו למשל ב- word2vec. ניתן להוכיח שעבור צורה מסוימת של NCE לוס (הנקראת InfoNCE) כי ככל שלוס זה קטן יותר המידע הדדי בין הדוגמא במרחב המקורי לבין ייצוגה במרחב ממימד נמוך עולה (צריך לציין שהמאמר הנסקר טוען שיש

עבודות שטוענות שהביצועים של ייצוגים על משימות downstream יותר תלויה בארכיטקטורה של האנקודר ופחות קשורה למידע הדדי). זה כמובן מצביע על אובדן פחות אינפורמציה בין הדאטה המקורי לבין ייצוגה כלומר הייצוג יהיה פחות לוסי ומייצג את הדאטה בצורה יותר מלאה. חשוב לציין שהאימון מתבצע במרחב הייצוג ולא במרחב המקורי כלומר הלוס מחושב על הייצוגים במרחב ממימד נמוך. לוס NCE זה בעצם לוקח זוג דוגמאות קרובות והרבה דוגמאות רנדומליות ומנסה למקסם את המנה בין דמיון של זוג הקרוב לסכום הדמיונות בינו לבין דוגמאות רנדומליות.

תקציר מאמר:

בשביל להבין את רעיון המאמר במלואו אנו צריכים להכניס עוד מושג חשוב, "עידון משימה" (task refinement).

עידון משימה: הגדרה ריגורוזית של מושג זה נלקחת מתורת הסיבתיות, אבל לצורך פשטות אסביר זאת עי" דוגמא. משימת סיווג Y_R בין זנים שונים של כלבים (או זנים שונים של חתולים) הינה עידון של משימת סיווג בין כלבים לחתולים Y_t. כלומר, אם ייצוג הדאטה מספיק טוב בשביל לבצע את Y_R, הוא יכיל מספיק מידע גם בשביל לבצע את Y_t בצורה טובה.

ולמה בעצם כל זה חשוב, אתם שואלים? קודם כל, נשים לב כי משימת ההבחנה (דיסקרימינציה) בין תכנים שונים בתמונות, כמו שנעשה בשיטות המבוססות NCE, הינה המשימה "הכי מעודנת" עבור דאטה סט נתון. זו הסיבה הנוספת (קיימים הסברים המקשרים שיטה זו למקסום מידע הדדי בין ייצוג דאטה ודאטה עצמו) לכך שהייצוגים שנלמדו בדרך זו, הוכחו כשימושיים למשימות downstream שונות. בעצם המאמר מוכיח טענה שלפיה ייצוג אינווריאנטי תחת שינויי סגנון עבור משימה Y_R נותר אינווריאנטי לכל משימה Y_t ש-Y_R הינה העידון שלה. כלומר, אם הצלחנו ללמוד ייצוג המסוגל לבצע דיסקרימינציה בין תכנים שונים ללא קשר לסגנון, ייצוג זה יעבוד טוב גם במשימות downstream שמהותן מבוססת על תוכן.

בעצם הוספת איבר רגולריזציה ללוס הרגיל של InfoNCE תורם להעצמת אי התלות של ייצוגי התמונה בסגנון שלה. כלומר תמונות קרובות יישארו קרובות גם לאחר שינוי סגנון ותמונות רחוקות יישארו כאלו אחרי שינוי סגנון גם כן.

עכשיו בואו נבין את המבנה של איבר הרגולריזציה:

איבר רגולריזציה - אופן חישוב:

- בונים שני סטים של פעולות אוגמנטציה (שינויי סגנון) A_1 ו- A_2, כאשר כל קבוצה מורכבת מזוגות של פעולות אוגמנטציה שונות (a_{1i}, a_{2i}).

לכל דוגמא x_i:

- עבור כל זוג שינויי סגנון מ-A_1, משערכים את התפלגות הדמיונות בין ייצוגים שלxi תחת a_{1i} ושאר הדוגמאות ממיני-באטץ' תחת a_{2i}. בשביל זה מפעילים את a_{1i} על x_i ומחשבים וקטור דמיונות d שלו עם הייצוגים של שאר הדוגמאות תחת a_{2i}. הדמיון מחושב כאקספוננט של מכפלה פנימית של הייצוגיים אחרי ששניהם מועברים דרך רשת נוירונים רדודה בעלת שכבה אחת או שתיים.
- וקטור d מנורמל כדי להפכו למידת הסתברות המסומנת p_1.
- מחשבים את וקטור הדמיונות עבור אוגמנטציות מ-A_2 באותה צורה: p_2.
- מחשבים מרחק KL בין p_1 ו- p_2 (דרך מעניינת להחליף את KL במרחק בין מידות הסתברות ולבדוק איך השתנו הייצוגים) וסוכמים אותם עבור כל זוגות הדוגמאות מ-A_1 ו- A_2.

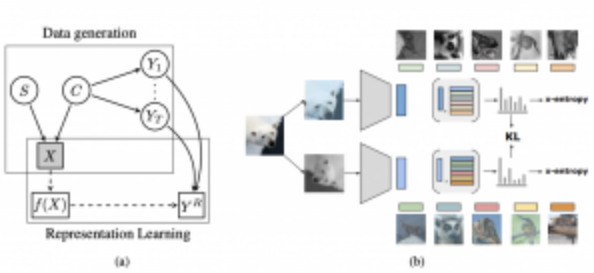

Figure 1: (a) Causal graph formalizing assumptions about content and style of the data and the relationship between targets and proxy tasks. (b) RELIC objective. KL refers to the Kullback-Leibler divergence, while x-entropy denotes cross entropy.

הישגי מאמר:

המאמר הוכיח שייצוגים של RELIC יותר חזקים מאלו של שיטות למידת ייצוג (BYOL, AMDIM, SimCLR) ב-3 היבטים שונים:

1. יחס דיסקרמינטיבי לינארי של פישר (LDR - linear discriminant ratio) המודד מרחק בין הייצוגים של הקלאסים השונים. ככל שהמרחקים בין מרכזי הקלאסטרים של ייצוגים בין הקלאסים השונים רחוקים יותר והדיאמטירים של הקלאסטרים קטנים יותר, נקבל LDR יותר גבוה. LDR גבוה של ייצוג הדאטהסט מצביע על כך שניתן להבחין בין דוגמאות מהקטגוריות השונות ביותר קלות עי" מסווג לינארי (ייצוג חזק יותר).

2. ביצועים על משימות downstream שונות (סיווג).

3. וזה חדש ומגניב: בחנו את עוצמת הייצוג על משימת למידת באמצעות חיזוקים (reinforcement learning) וראו ש- RELIC מצליח לשפר את הביצועים.

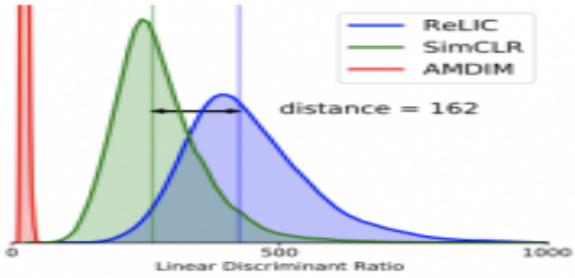

Figure 2: Distribution of the linear discriminant ratio (F_{LDA}, see text) of f for RELIC, SimCLR and AMDIM (y-axis clipped to aid visualization).

Table 1: Accuracy (in %) under linear evaluation on ImageNet for different self-supervised representation learning methods. Methods with * use SimCLR augmentations. Methods with † use custom, stronger augmentations.

Method		Top-1	Top-5
ResNet-50 architecture			
PIRL (Misra & Maaten, 2020)		63.6	-
CPC v2 (Hénaff et al., 2019)		63.8	85.3
CMC (Tian et al., 2019)		66.2	87.0
SimCLR (Chen et al., 2020a)	*	69.3	89.0
SwAV (Caron et al., 2020)	*	70.1	-
RELIC (ours)	*	70.3	89.5
InfoMin Aug. (Tian et al., 2020)	†	73.0	91.1
SwAV (Caron et al., 2020)	†	75.3	-
ResNet-50 with target network			
MoCo v2 (Chen et al., 2020b)		71.1	-
BYOL (Grill et al., 2020)	*	74.3	91.6
RELIC (ours)	*	74.8	92.2

דאטהסטים: ILSVRC-2012, ImageNet.

נ.ב.

המאמר מציע רעיון מעניין לשיפור ביצועים של שיטות ללמידת הייצוג ,המבוססות NCE. הם מציעים להוסיף איבר רגולריזציה לפונקצית לוס הסטנדרטית של NCE. מטרתו של איבר זה הינו לגרום ליחסים בין ייצוגי תמונות להיות אינוריאנטיים לשינויי סגנון בתמונות. המאמר מראה שהשיטה המוצעת מצליחה לבנות ייצוגים יותר טובים חזקים ממספר שיטות SOTA. הייתי רוצה לראות שיטה זו מוכללת גם לדומיינים אחרים וגם לסוגים שונים של משימות.

#deepnightlearners

הפוסט נכתב על ידי מיכאל (מייק) ארליכסון, Michael Erlihson, 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.