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

שיצא לפני כ 3 שבועות

תחום מאמר: למידת ייצוג ((representation learning

מאמר הוצג בכנס: הוגש כמאמר כנס ל ICLR 2021

כלים מתמטיים, מושגים וסימונים : גרף סיבתיות של מודל הסתברותי, NCE, InfoNCE,מרחק KL בין התפלגויות, עדון של משימת למידה (task refinement)

תמצית מאמר: המאמר מציע שיטה (הנקראת RELIC) לבנייה של ייצוג 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" אינוריאנטית 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 לפונקצית לוס תורמת. קודם כל שיטות מבוססות(NCE - noise contrastive estimation) בנויות בצורה שגורמת לייצוגים של דוגמאות "קרובות" להיות קרובות גם כן. נזכיר שלמשל עבור דומיין התמונות קירבה מוגדרת כדמיון מבחינה סמנטית/תוכן כאשר פעולות אוגמנטציה כמו הזזה, סיבוב או קרופ אינן משפיעות על קירבה בין ייצוגי תמונות באופן משמעתי. איבר רגולריזציה המוצע במאמר "מאלץ" את הייצוגים, בנוסף לתכונה המתוארת מעלה, להיות אינוריאנטיים 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 שתי תמונות אלו אחרי האוגמנטציה "מאולצות 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 downstream, 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 של Y_t והוא צריך להיות אינוריאנטי (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 מתבססת על על הנחה שייצוג חזק של דאטה בהכרח מסוגל להפריד בין זוגות של הדוגמאות 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 הרעיון של המאמר במלואו אנו צריכים להכניס עוד מושג חשוב, "עדון משימה" (task refinement).

עדון 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 עדון של משימת סיווג בין כלבים לחתולים Y_t. כלומר ייצוג דאטה שמספיק טוב בשביל לבצע את Y_R הוא יכיל מספיק מידע בשביל לבצע את Y_t.

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 משימה "הכי מעודנת" עבור דאטה סט נתון. וזו 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 העדון 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 להעצמה של אי תלות של ייצוגים בסגנון של תמונה. הרי אנו לא רק דורשים שייצוגים של אותה התמונה יהיו קרובים תחת שינויי סגנון שונים (ורחוקים מתמונות עם תוכן שונה) אלא בנוסף אנו רוצים לאלץ את הקרבה של הייצוגים לא להשתנות כאשר דוגמאות עוברות שינויי סגנון שונים.

עכשיו בואו נבין את המבנה של איבר הרגולריזציה:

איבר רגולריזציה - אופן חישוב

- בונים שני סטים של פעולות אוגמנטציה 1A ו 2A, כאשר כל קבוצה מורכבת מזוגות של פעולות אוגמנטציה שונות (a_1i, a_2i).
לכל דוגמא x_i:
- עבור כל זוג שינויי סגנון מ- 1A, משערכים את התפלגות הדמיונות בין ייצוגים של x_i תחת a_1i ו- שאר הדוגמאות ממיניבאטץ' תחת a_2i. בשביל זה מפעילים את a_1i על x_i ומחשבים וקטור דמיונות 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.
- הוקטור מנורמל כדי להפכו למידת הסתברות המסומנת 1p
- מחשבים את וקטור הדמיונות עבור אוגמנטציות מ 2A באותה צורה: 2p
- מחשבים מרחק KL בין 1p ו- 2p (דרך מעניין להחליף את KL במרחק בין מידות הסתברות ולבדוק איך השתנו הייצוגים) וסוכמים אותם עבור כל זוגות הדוגמאות מ 1A ו- 2A.

הישגי מאמר: המאמר הוכיח שייצוגים של RELIC יותר טובים משיטות למידת ייצוג (SOTA) 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 ניתן לסווג אותן ביותר קלות עי" מסווג לינארי (ייצוג חזק יותר)
2. ביצועים על משימות downstream (סיווג)
3. וזה חדש ומגניב: בחנו את עוצמת הייצוג על משימת למידת חיזוק

דאטה סטים: ImageNet ILSVRC-2012

לינק למאמר: https://arxiv.org/pdf/2010.07922.pdf

לינק לקוד: לא מצאתי

נ.ב.: מאמר מציע רעיון מעניין הנותן הסבר מנומק היטב על הסיבה של שיטות ,מבוססות NCE ,מסוגלות להפיק ייצוג חזק של דאטה. בהתבסס על ההסבר הזה הם מציעים שדרוג ללוס של NCE הממנף את המנגנון העומד מאחרי הסבר זה בשביל לבנות ייצוגים יותר טובים משיטות SOTA. הייתי רוצה לראות את שיטה זו מוכללת גם לדומיינים אחרים וגם לסוגים שונים של משימות.