סיכום מאמר בנושא: אלגוריתם יעיל לסיווג סרטוני ווידאו.

סרטוני הוידאו כיום, הינם כוח חזק בתעשיית האינטרנט וככל שהשימוש בהם גדל, אנשים מעוניינים בסיווג הסרטונים ובידיעה מוקדמת לגבי תוכנם, כשם שבעולם הטקסט קיימים חיפושים וסיווגים לתכנים מסויימים, סיווגים שהחלו בסימניות בספריות בעידן הספרים והמשיכו לסיווגים ע"י מילות מפתח באינטרנט, כך יש למצוא מענה להבנת עיקרי הסרטון וסיווגו לפני שהמשתמש צופה בו, פתרון לבעייה זו יחסוך בזבוז זמן מיותר של משתמש בצפייה בסרטונים שאינם רלוונטיים לו וגם ייתן למשתמש הגנה מפני סרטונים שאינם מטיבים עימו או אף פוגעים בו (כגון תוכני אלימות ועירום לדוג'), עד היום הוצעו וכבר בשימוש שיטות שונות של סיווג, אשר כולן מתבססות בעיקרן על עיקרון אחד חשוב, הוא 'עיקרון הדיוק', השיטה האינטואיטיבית והבטוחה: א. חילוץ כל הפריימים מהסרטון (על סמך CNN-Convolutional Neural Netwark, רשת נוירונים המשמשת לזיהוי הפריים)
ב. סיווג הסרטון על ידי התכנים שנמצאו בסך הפריימים על ידי אחד משתי שיטות האב:
1. Pooling methods
2. RNN-recurrent neural Networks
כאשר הראשון מסווג את הפריימים לא על פי סדר ולכן יעיל יותר אך פחות מדוייק, והשני משתמש בסדר הפריימים על מנת לסווג את תוכן הסרטון, יותר מדוייק אך בעל מחיר גבוה יותר)

שיטות אלו, הנזכרות לעיל, הוכחו כמדוייקות מאוד, אך החיסרון שלהן הוא היעילות, את היעילות ניתן לשפר בעזרת שני כיווני חשיבה:
<u>הראשון</u>, שיפור הזמן והיעילות של הCNN אשר תפקידו הוא לחלץ את הקבצים מהסרטון, שיטה זו שופרה בכמה דרכים אך השיפורים אינם משנים את העובדה שזמן הסיווג יהיה תלוי ליניארית באורך הסרט, כיון שהסיווג עצמו יקרה רק לאחר סיום חילוץ הפריימים.
<u>השני</u>, צמצום כמות הפריימים הנחלצים מהסרטון עצמו, וממילא התחלת סיווג מהירה יותר.

במאמר שלנו אנו מתמקדים בדרך השניה- צמצום הפריימים אותם נחלץ, כאשר אנחנו רוצים לשמור על רמת הדיוק של הבנת תכני הסרטון וסיווגו נעשה זאת בשלושה מישורים מקבילים:

א. <u>fast forward</u> – סוכן המבין על פי פריימים קודמים אלו פריימים פחות רלוונטיים בסרטון מבחינת המידע שאנו מחפשים לפתירת השאילתא, ועל פי הבנה זו הוא מדלג על הפריימים הללו. ערך המידע שנוציא מפריים כלשהו שונה מערך המידע מפריים אחר, לדוג': אם ניקח לדוג' שני פריימים סמוכים בסרטון כלשהו, נקרא לראשון A ולשני B כמובן שאחרי שלקחנו את פריים A, פריים B כמעט בוודאות יוסיף לנו מעט מאוד מידע כיון שבסרטונים מתקיים 'עיקרון המקומיות', ככל שפריימים צמודים יותר קיימת הסתברות גדולה מאוד לדמיון ביניהם ולכן אפשר לקחת פריים מייצג מכל כמה פריימים צמודים על פי אלגוריתם, עיקרון זה שאנו מלמדים את ה'סוכן' שלנו נלמד מדרך הפעולה האנושית, כאשר אדם רוצה לבדוק מה תוכן סרטון בלי לראות את כולו הוא מריץ חלקים ממנו שאינם רלוונטיים, וכאשר הוא רואה פריימים בהרצה שהם יותר רלוונטיים אליו הוא מריץ את הסרטון בקפיצות קטנות יותר הסוכן יכול לבחור בכל שלב אחד מהפתרונות הקיימים בקבוצה A
A{ -2s , -1s , +1s , +2s , +4s , +8s , +16s } כאשר '-' משמעו חזרה אחורה בסרטון ו'+' משמעו התקדמות בסרטון (קיימים מקרים בהם הסוכן יצטרך לחזור אחורה בפריימים במידה וחסר לו מידע בעקבות קפיצה גדולה מידי קדימה, אך מקרים אלו מעטים כיון שהקפיצה המקסימלית שלנו היא של 16 שניות, וגם זה נעשה בדרך חכמה שתוסבר בהמשך ולכן הקפיצה אחורה היא עד 2 שניות).

ב. Adaptive stop – סוכן זה בוחר מתוך קבוצה בוליאנית C כאשר: {continue(1) , stop(0)} C סוכן זה בא לפתור חילוץ של פריימים שכלל אינם רלוונטיים, כלומר: לאחר שמצאנו שסרטון מסויים משדר תכנים של 'ריקוד' או של 'כדורסל' אנו כבר יודעים שהסרטון זה מדבר על ריקוד או כדורסל ולכן אין צורך לבדוק אם גם שאר הפריימים הם כאלה כי השאילתא היא שאילתא של כן ולא, וכאן התשובה היא כן, במקרה כזה הסוכן ישדר 0 ובכך יסיים את תהליך החילוץ ויקדים את המועד של תחילת תהליך הסיווג.

כיון שאנו רוצים לחלץ כמה שפחות פריימים, אנו חייבים לדעת בשלב החילוץ כמה שניות לדלג (fast forward), וכן לדעת האם לעצור את פעולת החילוץ באמצע התהליך (Adaptive stop), אימון שני הסוכנים הנזכרים לעיל, מתבצע על ידי - reinforcement learning למידה באמצעות חיזוק, כלומר אימון המערכת על ידי מערך נתונים קיים, ובכך בכל איטרציה המערכת משפרת את עצמה בזמן ריצה, ככל שהמערכת תבדוק יותר פריימים כך היא תהיה מדוייקת יותר, וזה נעשה על ידי: reword function, פונקציה שמחזירה ציון לסוכן. הציון מבוסס על שני פרמטרים:
א. אחוז הדיוק בזיהוי (על פי הנתונים במערך המידע הנתון).
ב. מספר הפריימים שנחלצו על מנת לסווג את הסרטון (כאשר המקרה העדיף הוא זה שנצרך לפחות פריימים).
התוצאה שהפונקציה תחזיר תהווה חיזוק חיובי של המערכת על ידי ציון גבוה, כאשר הסוכן דייק בתוצאה וחיזוק שלילי על ידי ציון נמוך.
על פי האמור לעיל, היה על מחברי המאמר למצוא מערך נתונים גדול בכדי שיוכלו לבצע בו את האימונים למערכת, הם בחרו ב: youtube 8M, שהוא מערך הנתונים הגדול ביותר לסיווג סרטונים, ובעזרתו בהמשך הניסוי הם גם בודקים את דיוק הסיווג (אחוזי הצלחה) ואת יעילות הסיווג (מספר פריימים שחולצו מהסרטון) ביחס לשיטות הסיווג האחרות שהיו עד למאמר זה, אלו המבוססות -RNN recurrent neural Networks, ואלו המבוססות Pooling methods, ותוצאת הניסוי היא צמצום כמות הפריימים הנחלצים, ללא הפסד באיכות דיוק הסיווג.

סיכום: מחברי המאמר מציעים שיטה לסיווג סרטוני וידאו בעזרת שני מנגנונים: "fast forward " ו-"Adaptive stop ". הם משתמשים ב"reinforcement learning " כדי לאמן את שני המנגנונים הללו. מערכת זו תקטין באופן משמעותי את העלות החישובית לסיווג הוידאו תוך שמירה על רמת הדיוק.

לדעתי האלגוריתם של שיטתינו עדיף מקודמיו בשני מישורים: הראשון הוא המישור הפרקטי אשר בפועל יותר יעיל ובעל עלות זולה יותר, והשני הוא המישור התפיסתי, האלגוריתם שלנו עובד בצורה יותר הגיונית, הוא לא "מעתיק" את ידיעותינו לגבי תמונות לסרטוני וידאו אלא לומד מידע שיש לנו על סרטונים (למשל: עיקרון המקומיות) ומשתמש בו. קיים חיסרון מזערי שהוא הוספת שלב על האלגוריתמים הקודמים, שלב ה- reinforcement learning. כותב המאמר לא התייחס לזה בכלל במאמר שלו.

כותב המאמר לא הציע הצעות התפתחויות לעתיד, אך לדעתי, על מנת לשפר את האלגוריתם עוד יותר בהמשך, עלינו להפסיק לראות סרטון כאוסף תמונות שאין יחס ביניהן, אלא לראות אותו כרצף של פעולות בעלות משמעות, נביא דוג' להבדל בין דרכי החשיבה: ישנו סרט פעולה, שבו מצויה גיטרה בהרבה מאוד חלקים מהסרטון כאביזר בביתו של השחקן, וודאי שבמקרה זה אין תוכנו של הסרט שייך ל'גיטרות', אך על פי השיטות שהובאו במאמר הוא ייתפס כאחד כזה, כאשר נשנה את דרך החשיבה, נוכל להבין שהימצאות הגיטרה אינה מגדירה אותה כחלק מהגדרת תוכן הסרט.

מגיש: יחזקאל כדורי
ת.ז: 303106157