

דוח סיכום - video indexer

זיהוי פרצופים

אלגוריתמים לשיפור זיהוי הפנים של video indexer :

קווים מנחים לכלל האלגוריתמים

שימוש בJSON התובנות של video indexer , מעבר על נתוני זיהוי הפנים שזוהו unknown והחלפת הזיהוי לדמות ידועה .

אלגוריתם 1

איסוף כל מופעי הדמויות ולאחר מכן מיון על פי זמנים והפיכת דמות לא ידועה שכלואה בין שני דמויות זהות לדמות הזו.

אלגוריתם 2

עבור כל דמות לא ידועה נבדק אם הדמות שלפניה ואחריה ברצף הזמן היא אותה הדמות , ובנוסף אם הם באותו השוט , ואם כן נחליף את הזהות בדמות הידועה.

מכיוון שהופעת הדמות הלא ידועה יכולה להתפרס על יותר משוט אחד , ייבדקו שני דברים :

א. תחילת הופעתה של הדמות הלא ידועה וסוף הופעתה של הדמות שלפניה באותו השוט .

ב. סוף הופעתה של הדמות הלא ידועה ותחילת הופעתה של הדמות שאחריה באותו השוט .

אם שני הדברים מתקיימים אז שלושת הפרצופים ייחשבו באותו השוט מבחינת האלגוריתם.

אלגוריתם 3 - שיפור זיהוי הפנים על פי נתוני הדוברים של video indexer

חילוץ הנתונים של הדוברים בסרט מJSON של התובנות : הדוברים מזוהים כ#speaker כאשר # הוא מספר .

מיפוי דובר לדמות בסרט - עבור כל #speaker נמצא מי הדמות הדוברת .

דוגמאות למיפוי : Katherine Heigel - speaker 3 , James Marsden - speaker 5 .

מעבר על נתוני זיהוי הפנים ועבור כל מופע של דמות לא ידועה על רצף זמן של הסרט נבדוק אילו דוברים דיברו במהלך הופעתה . אם היה רק דובר אחד אז נחליף את הזהות הלא ידועה לזהות הדובר לפי המיפוי של דוברים לדמויות.

אלגוריתם 4 (pipeline)

האלגוריתם משלב את שלושת האלגוריתם הנ"ל כ#pipelines בצורה הבאה : הפעלת אלגוריתם 3 , על פלט של אלגוריתם 3 מפעילים את אלגוריתם 2 , ועל הפלט שיוצא מפעילים את אלגוריתם 1.

מדדים לבדיקת שיפור זיהוי הפנים:

עבור תוצאות video indexer ותוצאות לאחר הפעלת אלגוריתם השיפור נבדקו המדדים
הבאים : precision , recall , f - score , jacaard

בבדיקות נעזר במאגר הנתונים MG - התיוג הידני של הסרטים שעליהם נעשו הבדיקות.
תוצאות הבדיקות מצורפות בקובץ excel.

זיהוי דוברים

video indexer נותן נתונים של הדוברים בסרט . כל דובר מזוהה כ- speaker # כאשר # הוא מספר . עבור כל דובר מצויין באיזה זמנים במהלך הסרט הוא דיבר .

הנתונים אינם מהימנים , ישנם המון אי דיוקים בזיהוי של הדובר .
להלן דוגמאות מהסרט 27 Dresses שנבדקו ידנית :

(1) בסצינת הפתיחה של הסרט יש מונולוג של השחקנית קתרין הייגל .
בזמן המונולוג השחקנית הדוברת זוהתה כ-4 דוברים שונים: speaker 2, 3, 9, 10 .

במקרה זה ניתן לשער שהמוזיקה ברקע הפריעה בזיהוי .

(2) במונולוג נוסף של השחקן ג'יימס מארסדן השחקן הדובר זוהה כ-3 דוברים שונים :
speaker 3, 5, 8 .

המונולוג ערך כ-11 שניות בין 1:25:57 ל 1:26:08 , ולא היו שום רעשי רקע / מוסיקה בזמן הזה.

להלן דוגמאות מהסרט crazy,stupid,love שנבדקו ידנית :

(1) במהלך הסרט יש דיאלוג בין השחקן סטיב קארל לשחקנית אנלי טיפטון בין הזמנים 00:09:45 עד 00:10:45 השחקנית זוהתה כ-3 דוברים שונים : speaker 7, 9, 11 בזמן שהשחקן זוהה כ- speaker 6 .

במהלך הזמנים האלה לא היו רעשי רקע / מוסיקה , והדיבור של השחקנית היה רציף .

(2) במהלך הסרט ישנו דיאלוג בין השחקן סטיב קארל לשחקן ריאן גוסלינג בין הזמנים 1:37:00 עד 1:40:44 שבו הדוברים לרוב זוהו כדובר אחד : speaker 3 .

ניתן להניח שהדבר נבע ממוזיקה שהיתה ברקע ועצם זה שזה היה דיאלוג בין שני אנשים מאותו המין .

להלן דוגמאות מהסרט flight :

(1) במהלך המשפט שבו עורכת הדין תחקרה את הדמות שגילם דנזל וושינגטון היא זוהתה כמספר דוברות : 3, 15, 7, 9, 11 speaker כאשר במהלך שאלה ארוכה אחת רציפה ללא רעשי רקע היא זוהתה כ-3 דוברות : speaker 9,11,15 (בין הזמנים 2:03:48 עד 2:4:02).

(2) כאשר השחקן דון צ'ידל מדבר אל השחקן דנזל וושינגטון הוא מזוהה בעיקר על ידי שני דוברים : speaker 3, speaker 5 (בין הזמנים 1:35:40 עד 1:36:40) , כאשר אין רעשי רקע/מוסיקה .

בנוסף יש סרטים ש- video indexer לא מביא נתונים לגבי הדוברים , לדוגמה הסרט פורסט גאמפ.

בדיקת איכות התמלול

על מנת לבדוק את איכות התמלול של video indexer , נעשתה השוואה בין התמלול לבין תכני קבצי srt של הסרטים - אשר מכילים את התסריט המילולי של הסרט.

נעשה שימוש ב-4 אלגוריתמים להשוואת טקסטים :

Levenshtein - אלגוריתם מבוסס edit distance, לעיתים מכונה כedit distances למרות שמדובר במשפחה שלמה של אלגוריתמים .

Jaccard index - אלגוריתם מבוסס token distance , היחס בין החיתוך לאיחוד של שני הטקסטים.

Cosine - אלגוריתם מבוסס token distance , ייצוג הטקסטים בעזרת וקטורים וחישוב קוסינוס הזווית בין שני הווקטורים.

Jaro-Winkler - אלגוריתם נוסף מבוסס edit distance .

תוצאות ההשוואה מצורפות בקובץ אקסל נפרד.

****במהלך המימוש נתקלנו בבאג בספרייה של IMDB, יצרנו קשר איתם והבאג יתוקן בהקדם(הבאג הוא הדפסת שורות רבות של לוגים) .**