

מערכת IDA

מגשים: דניאל פרידמן ומתן אלחייני

מנחה: ד"ר עמי האופטמן

רקע

תתארו לעצמם בלש שמחפש מידע על משתמש מסוים בפורום אנונימי כלשהו. נניח שהמשתמש רשם במהלך חייו יותר מ-5000 הודעות. הדבר אומר שהחוקר יאלץ לקרוא את **כל ההודעות** האלה. אלו יכולות להיות הודעות ארוכות, אלו יכולות להיות הודעות לא נעימות. ויתרה מכך, יש להיות מרוכז במהלך כל התהליך וזאת על מנת למצוא כל פרט חיוני לחקירה. הפתרון שאנו מציעים יהיה למעשה, מערכת שעושה זאת בשבילו.

תיאור המערכת

המערכת עושה שימוש בטכנולוגיית Word2Vec שבעצם מהווה מילון מילים לווקטורים. כל מילה מומרת לווקטור שכל מימד בו מהווה ערך לפי ציר מסוים. לדוגמא, ציר יופי שיהווה מדד ליופי וכד'. בכך, לאחר שנתרגם את כל המילים לערכים מספריים שמתארים את משמעותם, נוכל להרכיב מטריצה. מטריצה זו תהווה את השכבה ראשונה במודל למידת המכונה.

המודל יורכב בטקטיקה של שכבות כאשר בכל שכבה יש נוירונים. נוירונים אלה מתחברים לנוירונים משכבות שכנות בצורה כזו שלכל קישור בין נוירון לנוירון יש משקל. משקל זה הוא למעשה מידת החשיבות של הקשר שבא לידי ביטוי בחישוב ההחלטות הסופיות. אם נתייחס לכל נוירון כזה כשלעצמו, נראה כי הוא גם מקבל החלטה לוקאלית שקשורה לקלטים שנכנסים אליו ולפיה הוא קובע אם ואיזה פלט להוציא, דבר שגם כן בא בחשבון החישוב הסופי.

עכשיו כשאנו יודעים כיצד רשת נוירונים מתפקדת, איך מתפעלים אותה?

- מכינים מידע שמכיל משפטים מסווגים לקטגוריות – הסיווג מבוצע ידנית.
- בונים מודל לפי מחקר מעמיק בלמידת מידע טקסטואלי. המודל יורכב מ-8 שכבות מ-4 סוגים שונים, שכבות אלה ימפו, יסננו, ינתחו ויחפשו דפוסים.

- בהתחלה המודל ידע לקבל משפט בייצוג ווקטורי ולתת לו דירוג התאמה לכל קטגוריה, גם אם הוא טועה משמעותית בדירוג.

- לאחר מכן נספק לו דוגמאות רבות ככל האפשר. אחרי כל דוגמא נבחן את דיוק תשובת המודל ביחס לזו האמיתית, ואז "נאמר" למודל איפה הוא צריך להשתפר (תיקונים מתמטיים לערכים המספריים של המשקלים והחלטות הנוירונים). זאת אומרת שהמודל יותאם בחישוביו לדפוס המידע אותו אנו מחפשים.

וכך לאחר מספיק איטרציות המודל יצליח לבצע סיווגים עם אחוזי שגיאה נמוכים. כך שבסופה של הפרוצדורה, נקבל כפלט לכל משפט את מידת התאמתו לכל קטגוריה, בדיוק רב יחסית. תפקיד הקטגוריות הוא לחלק את המידע לפי סוגי ההתקשרות לעולם הפיסי. לשם כך בחרנו ארבעה סוגי התקשרות: כלום (לא חושף שום פרט על האדם), דעה (חושף את דעות, העדפות ותחביבי הפרט), סביבה (מיקומים, תעסוקה, פעילויות פנאי והסביבה החברתית של המושא), עובדות אישיות (תכונות גופניות ונפשיות).

לפי חלוקה זו ועל ידי מיון ההתאמה לקטגוריה, קל להציף את הפרטים הכי אישיים שהנבדק חשף ברשת. המטרה הכללית היא למעשה למצוא פנינות מידע מתוך כמות המידע העצומה. או במילים אחרות, הפרויקט שלנו מוצא "מחט בערימה של שחת".

טכנולוגיות

- Reddit – זהו אתר פורומים ענק שממנו שאבנו מידע לתחקור.
- TensorFlow & Keras - ספריות קוד פתוח ללמידת מכונה ויישום רשתות נוירונים.
- Vue.js – Java Script framework המאפשר יישום UI ואפליקציות עמוד יחיד.
- Word2Vec – מילון המרה של מילים לווקטורים.
- API – כל פונקציונאליות המערכת נגישה דרך API. עימו גם מתקשר ה-API.
- הפיתוח הכללי נעשה ב-Python בתוך סביבת העבודה PyCharm.