

Poemachine - מח(ו)לל הקודש

פרויקט סיום בסדנת NLP בעברית - האוניברסיטה העברית בירושלים

בהנחיית יובל פיינשטיין

חנית חכים

308396480

בר תמיר

305189946

1. הגדרת הבעיה

ייצור של טקסט (text generation) היא משימה ליבתית ב-NLP, הבאה לידי ביטוי באפליקציות שונות כמו: תרגום, סיכום טקסט, צ'אט-בוט, וכן בכתיבה יצירתית. כתיבה יצירתית, בפרט שירה, כוללת אתגרים רבים, אשר הן שיטות NLP קלאסיות והן שיטות מתקדמות המבוססות על למידה עמוקה למשל, טרם התגברו עליהן לחלוטין: שפה פיגורטיבית ומלאה במטאפורות, קושי בביצוע אבלואציה בשל הערכה סובייקטיבית, וכן היעדר פרמטרים מוגדרים לאבלואציה. המבנה הייחודי והמורפולוגיה העשירה של השפה העברית, הופכים את השירה העברית לבעלת רובדים עמוקים ופוטנציאל להגיע לרמה ספרותית גבוהה מאוד. אך אליה וקוץ בה- הפוטנציאל הגבוה ברמת השירה העברית מקשה על למידה של מודל שפה שיודע להבין ולייצר טקסט כמוה. כל אלה בנוסף להיעדר דאטא דיגיטלי זמין, הופכים את ביצועה של משימה זו למורכבת עוד יותר, ומעטים הניסיונות שנעשו בתחום. לאור זאת, החלטנו לנתב את חיבתנו לשירה עברית כדי לנסות להתמודד עם אתגר זה.

מכאן שהגדרת הבעיה עימה מתמודד הפרויקט הינה ייצור אוטומטי של שירה עברית. על מנת להפוך את הבעיה למעניינת אף יותר, ביקשנו לייצר שירה התואמת לרוח תקופות שונות של משוררי שירה עברית. כלומר, אנו מעוניינים לכתוב מספר מכונות המייצרות שירה, כאשר כל אחת מנסה להתחקות אחר סגנון המאפיין משוררים בתקופה שונה - למשל בתקופה המודרנית, בתקופת תחיית העברית, וכו'.

2. איתור וניקוי נתונים מתאימים

הבעיה היא אינה בעיית למידה מפוקחת, ומכאן שאין לנו את הצורך להשתמש במידע מתוויג. זה מאפשר לנו להשתמש בכל סוג של שירה עברית, והופך מקורות שירה רבים לרלוונטיים עבורנו. כלומר, הנתונים הרלוונטיים לפתרון הבעיה יהיו כמה שיותר שירים הכתובים בעברית, המחולקים למשוררים מתקופות שונות, על מנת לאפשר לנו ללמד מספר מודלים שונים אשר יתחקו אחר משוררים בני התקופה.

עקב מגבלות של זכויות יוצרים, השגת הדאטה הייתה אתגר משמעותי. יתר על כן, רוב הדאטה השירי מצוי בעותקים קשים, בעוד אנו היינו זקוקים לעותקים רכים שלו. תחילה השגנו את הדאטה שהייתה זמינה לנו באופן חופשי מהאינטרנט, כזו שלא מוגנת בזכויות יוצרים, בעיקר מפרויקט "בן יהודה". פרויקט "בן יהודה" מאגד תחתיו יצירות שזכויות היוצרים שלהן הפכו להיות נחלת הכלל. כדי לאסוף דאטה זו ביצענו crawling לאתר במטרה לסנן ולחלץ את מה שנכלל בסוגה "שירה". עם זאת, ביקשנו להשיג גם דאטה של משוררים מודרנים על מנת לייצר שירים בסגנון תקופה זו. לכן פנינו למשוררים, הוצאות לאור, ביה"ס האמנותי "מקום לשירה", וכן פרסמנו קולות קוראים בקרב קהילות משוררים על מנת לקבל מהם כמה שיותר שירים. במסגרת מאמץ זה קיבלנו ספרים שלמים של משוררים בודדים וספרים של הוצאה לאור של "מקום לשירה", המאגדים בתוכם שירים של משוררים שונים. לאחר מכן היה עלינו לנקות את הדאטה ולהביא אותו לפורמט אחיד, על מנת לאמן את מודל ייצור השירים שלנו. את הדאטה שהגיע אלינו בפורמט PDF היה עלינו לחלץ ולהמיר לטקסט (פורמט txt) תחילה (מה שלא היה מובן מאליו בשל הניקוד הנלווה תמיד לטקסט מסוג שירה). ניקוי הדאטה כלל הסרת הניקוד, הפיכה של התו '/' (לכנס), המשמש לעיתים כירידת שורה בכתיבת שירים) לתו 'ח' (ירידת שורה), הסרת שורות ריקות משולשות (כלומר 3 שורות ריקות ברצף, או יותר), הסרת תווים שאינם אותיות עבריות (אותיות לטיניות-מילים באנגלית, מספרים, סימנים מיוחדים וכו'). כמו כן, על אף שבמסגרת איסוף הדאטה ביקשנו לקבל אך ורק שירה, בבחננו את הדאטה ראינו שהוא מכיל לא מעט מסות, פואמות או סיפורים קצרים. לכן, כחלק מניקוי הדאטה הורדנו טקסטים המכילים שורות באורכים החורגים מאורך שורה קצר המאפיין שירה, במטרה לסנן את אותן סוגות ספרותיות אחרות. לבסוף נירמלנו את כל השירים להיות בפורמט הבא:

```

שם היצירה: <שם_של_שיר>
מאת: <שם_של_משורר>

<היצירה_עצמה>
...
<|endoftext|>

```

כמו כן, איגדנו קבוצות שירים בעיקר לפי תקופות בהן חיו וכתבו המשוררים. עוד על כך ניתן לקרוא בהמשך בפרק על המודל.

3. מיפוי ראשוני - סקירת ספרות

ייצור טקסט בצורה אוטומטית (text generation) נעשה לרוב על ידי למידה של מודל שפה, כלומר מודל שבהינתן מילים קודמות חוזה את המילה הבאה. טיב תוצאותיהם של מודלים אלו, מושפע מאוד מכמות

המילים הקודמות למילה הבאה שנרצה לחזות, אותן לוקח המודל בחשבון. גישות קלאסיות יותר הסתמכו לשם למידת מודל שפה על מודלים סטטיסטיים מרקוביים למשל, כמו Hidden Markov model (בקיצור: HMM) שהינו מודל מרקובי גנרטיבי. מודל זה יכול לעבוד עם n-grams ולהתחשב בח מילים אחורה כדי לחזות את המילה הבאה.

כיום רוב השיטות הנפוצות והמצליחות, מתבססות על למידה עמוקה. המודל הבסיסי של למידה עמוקה למשימות NLP וייצור טקסט בפרט הוא Recurrent-Neural-Network (להלן: RNN), שבזכות המבנה הרקורנטי שלו מאפשר לקחת בחשבון שיקולים והשפעות רחוקות יותר בסדר הלינארי של המילים בטקסט. קיימים שיפורים למודל קלאסי זה - למשל: Long Short Term Memory (LSTM), וכן Gated Recurrent Unit (GRU). יחידות הזכרון ב-LSTM נקראות תאים: הן מקבלות קומבינציה של המצב הקודם והמצב הנוכחי כקלט. התפקיד שלהן הוא להחליט מה לשמור ב"זכרון" ומה לשכוח. ¹GRU הוא הרחבה נוספת של RNNs והוא שונה מארכיטקטורת LSTM ברשת gating, שמייצרת סיגנל ששולט במצב הנוכחי ובזכרון הקודם, כדי לעדכן את האקטיבציה והמצב הנוכחי של הרשת.

ארכיטקטורות גנרטיביות עמוקות רבות השתמשו בGRUs לטובת ייצור טקסט. שיפור נוסף משמעותי היה עם הופעת BiDirectional RNNs ²(להלן: BRNNs). הרעיון מאחורי ארכיטקטורה זו הוא שהפלט בזמן t יכול להיות תלוי לא רק באלמנטים קודמים, אלא גם באלמנטים עתידיים ברצף. המימוש נעשה על ידי שני RNNs כאשר כל אחד מקבל את הקלט מכיוון שונה מהשני. הטכניקות הפופולריות ביותר לייצור טקסט בתקופת הלמידה העמוקה הן Variational Auto-Encoders ³(להלן: VAEs) ו-Generative Adversarial Networks ⁴(להלן: GANs). ארכיטקטורות אלה גם מאפשרות להתמודד עם האתגר שבדאטה לא מתוויג על ידי למידה בלתי מפוקחת. VAEs כולל encoder שמקודד את הדאטה למשתנים חבויים, ו-decoder העושה את הפעולה ההפוכה ומשחזר את הדאטה שהגיע כקלט. פעולת הקידוד לוקחת קלט x ומייצרת אוטופוט ב-latent space $p_{\phi}(z|x)$, כאשר Φ מייצג את הפרמטרים של פעולת הקידוד, בעוד ה-decoder בדיוק את ההפך - הוא מוצא את ההתפלגות $q_{\theta}(x|z)$ של הדאטה בהינתן ה-latent distribution. θ מייצג את הפרמטרים של פעולת ה-decoding.

באשר לGANs, הם מבוססים על גישה שונה ומקורית של אימון בשיטת היריב. ראשית, ה-generator מייצר את דגימות הדאטה, ואז ה-discriminator מסווג את אותן הדגימות כאמיתיות (כלומר דאטה מהתפלגות האימון) או כמזויפות (כלומר יוצרו על ידי generator). המטרה של ה-generator היא לייצר דגימות שהן כמה שיותר קרובות לדאטה אמיתי, כך שיצליחו לשטות ב-discriminator, ומטרת ה-

¹ Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation.

² M. Schuster, K.K. Paliwal. Bidirectional recurrent neural networks.

³ Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes.

⁴ Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets.

discriminator היא לסווג נכון את שני סוגי הדגימות. שיטה זו הוכיחה את עצמה בעיקר עבור משימות של עיבוד תמונה, אולם יש לה פוטנציאל גבוה להצליח גם בטקסט.

בשנת 2017 Google AI הציגו במאמר ⁵Attention is all you need את Transformer, אשר נשען על ארכיטקטורת encoder-decoder וכן מכיל multi-head attention mechanism והביא לשיפור ביצועים משמעותי, בעיקר בזכות יכולות חישוב מקביליות בgpu ומודלים איכותיים רבים שהתבססו עליו (attention הינו מיפוי של query וסט של זוגות key-value לפלט. הפלט מחושב כסכום ממושקל של ה-values, כאשר המשקל המוקצה לכל אחד מחושב על ידי compatibility function של ה-query וה-key התואם).

בשנים האחרונות פותחו Bert⁶ של Google AI וכן GPT-2 של openAI, ולאחרונה גם GPT-3. אלה מודלים הכוללים מספר רב מאוד של פרמטרים, ואומנו על כמות גדולה מאוד של דאטה (כלומר הם pre-trained), ולרוב אפשר להשתמש בהם כמו שהם לייצור טקסט, או לבצע fine-tuning למודל הקיים (למשל אם רוצים לעבוד על קורפוס שונה מאוד מהקורפוסים עליהם אומן המודל. גישה זו נקראת transfer learning).

Bert עושה שימוש בtransformer, אבל את האימון שלו מבצע בצורה דו כיוונית. כמו כן הוא משתמש בגישה חדשה של Masked LM: לפני שרצף המילים מוזן לBert, חמישה עשר אחוזים מהמילים בכל רצף מוחלפים בmask tokens. המודל מנסה לחזות את הערך המקורי של המילים הממוסכות, בהתבסס על הקונטקסט של שאר המילים הלא ממוסכות.

גם GPT-2 מבוסס על ארכיטקטורת transformer. שני הבדלים עיקריים בינו לבין Bert: ראשית GPT-2 משתמש ב-decoder blocks בלבד, בעוד Bert עושה שימוש ב-encoder-blocks. בנוסף, הפלט של GPT-2 הוא token אחד כל פעם, כמו מודל שפה קלאסי. זאת בעוד Bert מוציא את הרצף הכולל.

כפי שצויין בהקדמה, ייצור של טקסט ספרותי ובפרט שירה היא משימה קשה יותר, עקב רבדי העומק וסמליות השפה של טקסט מסוג זה. לאחרונה פורסמו טקסטים ספרותיים שנוצרו על ידי GPT-3 (באנגלית), והגיעו לרמה גבוהה. בעבר היותר רחוק נעשו נסיונות ליצור שירה יותר מובנית בעזרת חוקים מוגדרים, כמו למשל שירת הייקו שהינה בעלת מבנה סטט

⁵ Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need.

⁶ Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.

4. המודל

בהתבסס על הסקירה הספרותית, וכן על מורכבות המשימה שניצבת לפנינו, החלטנו לבחור במודל מבוסס למידה עמוקה לפתרון הבעיה.

על מנת להקל על זמן האימון של המודל ובשל כוח חישובי מוגבל שהיה ברשותנו, החלטנו להשתמש בארכיטקטורת רשת נוירונים מסוג GPT-2 (תיאור טכני של המודל ניתן למצוא בסקירה הספרותית) שאומנה מראש על דאטא של ויקיפדיה בעברית, ולבצע לה fine tune על בסיס הדאטה שלנו.

בנוסף, נעשה למודל זה fine tuning עם שירים מהאתר "במה חדשה". אתר זה מכיל מגוון רב של שירים בעברית. הוא מאפשר פרסום שירים חופשי לכל מי שחפץ בכך. מכאן שהוא כולל כמות רבה מאוד של שירים, אך אין בקרה על טיבם, סוגם, רמתם וכו'. עם זאת הוא מהווה בסיס טוב כדי לקחת את פרמטרי המודל לאזור קרוב יותר לזה שבו אנו חפצים.

הבחירה במודל הנ"ל נושאת עמה יתרונות לא מעטים: נראה כי מודל זה מצליח להתמודד עם השפה העברית בצורה מרשימה, וכן מצליח פעמים רבות לתפוס את הסמנטיקה הנדרשת לייצור משפטים קולחים והגיוניים, הנושאים משמעות. היכולות הגבוהות הינן חשובות עוד יותר עבור טקסט הכתוב במשלב שפה גבוה יחסית כמו שירה, והכולל סמנטיקה שאינה straight-forward, כמו למשל מטאפורות ודימויים חדשים שאינם שגורים ונפוצים בשפה. מדובר על מודל מרובה פרמטרים, שהוכיח את עצמו במשימות של ייצור טקסט. בנוסף, המודל המאומן מראש מקל עלינו רבות, ומאפשר לנו להקצות כוח חישובי סביר לביצוע fine tuning בלבד, ולא לאמן את המודל מההתחלה. שימוש בטכניקה זו, אף אפשר לנו את פיצול הדאטא על פי תקופות או משוררים - שכן נראה כי כמות דאטא יחסית קטנה, מספיקה במקרה של ביצוע fine tuning.

על מנת לקודד את השירים שלנו לפני הכנסתם למודל, השתמשנו בtokenizer הנלמד גם הוא על בסיס ויקיפדיה בעברית, בו נעשה שימוש באימון המודל המקורי. שימוש ברכיב זה, אפשר לנו מענה על מספר קריטריונים חשובים מבחינת ה-embedding הרצוי, ובפרט קידוד של מבנה השיר: למשל, התייחסות לירידות שורה ולסימני פיסוק, שהם בעלי ערך ומשמעות בשירה.

על מנת להתאים את הדאטה שלנו אל הדאטה של המודל המאומן, ביצענו שלב מקדים נוסף בעיבוד הדאטה, והפרדנו בין שירים על ידי סימן מוסכם מראש.

כדי לאפשר ביצוע מספר fine-tuning לפי תקופות זמן שונות של משוררים, החלטנו לפצל את הדאטה שאספנו לקטגוריות הבאות:

- משוררים מודרניים
- משוררים מתקופת תחיית העברית
- משוררים מתקופת ימי הביניים

- שירי "ננופואטיקה" - ספרי שירה שקיבלנו מהוצאת "מקום לשירה", המכילים שירה שהיא קצת יותר fringe, כלומר "שירת שוליים", כזו שלא בהכרח עומדת בסטנדרטים "נורמטיביים" של שירה קלאסית.
- שירת ביאליק
- שירת רחל

הפיצול מבוסס על היכרות מוקדמת שלנו עם התחום, וכן על סמך כמות הדאטה-איגדנו תקופה, משורר או סגנון שירה, רק אם היו לנו מספר יחסי גדול של שירים בסט האימון והולידציה. הנחנו, כי מספר גדול יותר של שירים יאפשר להגיע לתוצאות קוהרנטיות וקרובות יותר לסט האימון המקורי. על מנת לאמן את המודל ולהשתמש ב-GPU, עבדנו עם תשתית "google colab" - המספק אפשרות להריץ בחינם קוד על שרתים עם GPU וכוח חישובי גדול יותר מזה שהיה לנו לוקאלית. בעצם על בסיס fine-tuning למודל הבסיסי על בסיס הדאטה-סט מהקטגוריות השונות יצרנו מספר מודלים שונים, במטרה לייצר סגנונות כתיבה שונה. כל דאטה-סט פיצלנו ל-train ו-valid ביחס של 80% : 20% בהתאמה.

בסיום כל fine-tuning שמרנו את המשקלות של המודל וביצענו text-generation (זהו בעצם שלב ה-inference שלנו), על בסיס מודל השפה המאומן. לטובת זאת לקחנו קוד גנרי של מודל GPT-2 המשתמש במודל מאומן לטובת ייצור טקסט, וערכנו אותו בהתאם לצרכינו, כך שנוכל להזין לו בתור מילים התחלתיות את שם השיר, נוכל לשלוט על אורך הטקסט ו/או להזין לו את הtoken ששתלנו בסוף כל שיר בסט האימון ומסמל את סוף השיר - כ-stop-token.

5. הערכת התוצר

כפי שצויין לעיל, אבליואציה של מודלים לייצור טקסט הינה בעייתית. בפרט, בשירה קיימים פרמטרים רבים שאינם מוגדרים בצורה אקספליציטית להערכה של שיר, וכן יש עניין רב של טעם אישי. יש שיחשבו ששירים מסויימים ירודים, בעוד אחרים יתייחסו אל אותם השירים בהערכה רבה. לכן הערכה אוטומטית על בסיס פרמטרים מוגדרים מראש הינה בעייתית לתהליך יצירתי, שמטרתו פעמים רבות היא "לצאת מן הקופסה", וכן הערכה אנושית של איכות השיר יכולה להיות מוטית על בסיס טעם. לכן החלטנו שהדרך הטובה ביותר בהינתן המשאבים שלנו להערכה תהייה סוג של מבחן טיורינג, שיטה נהוגה לבחינת מודלים שאמורים לייצר פלט הדומה לכזה הנוצר על ידי אדם. הצלחה של מכונה במבחן טיורינג מתבססת על חוסר היכולת של אדם לקבוע האם הפלט שהוא רואה נוצר על ידי מכונה או אדם (במקור: להבדיל האם הוא משוחח בציט עם אדם או עם מכונה). כך נוכל לקבל ציון מספרי עבור שירים שהמכונה מייצרת (אחוז האנשים שהצליחו לזהות שהשיר נכתב על ידה). ציון זה יעריך את טיב השיר בכך שייצג את היכולת של אדם לזהות האם השיר נכתב על ידי מכונה או אדם. כמו כן נקבל ציון ממוצע לנבדק, שייצג את היכולת של הנבדקים להבחין בין השירים של המכונה לשירים של משוררים אנושיים.

לטובת שלב האבלואציה כתבנו [שאלון](#) בפלטפורמת google forms. השאלון כולל מספר שירים, הן שהוציא המודל שלנו (6 שירים), והן שירים אמיתיים מסט האימון (5 שירים, ללא ציון שם המחבר). המשתתפים התבקשו לסמן, עבור כל שיר, האם לדעתם מקור השיר הוא מודל אוטומטי, או שהשיר נכתב על ידי בן אדם. כמו כן, ביקשנו מכל משתתף לדווח את מידת ההבנה שלו בשירה עברית, וכן את התדירות בה הוא קורא שירה. צילום מסך של השאלון מצורף להלן:

מכונה v.s בן אדם - מי כתב את השיר?

השירים הבאים נכתבו על ידי משוררים, או על ידי מכונה (מודל אוטומטי) לייצור שירים. עבור כל אחד מהשירים הבאים, סמנו: האם השיר הזה נכתב על ידי משורר כלשהו, או על ידי המודל האוטומטי?
* Required

הלב * 5 points

הלב
הלב שלו הולך ומתרחב.
אני נופל על האדמה
מעל האוקיינוס.
אישה שולחת לו את זרועותי
נושק לרוקן את שמיכת הלילה.

מכונה
 בן אדם

היקום כולו * 5 points

היקום כולו
נולדתי עירם לגמרי
וגם במותי אהיה עירם לגמרי
אז האם כלנו בעצם לבד?
לא
כך היקום כלו בא ונולד
מתוך היקום כלו
וגם הולך ומת
לתוך היקום כלו

מכונה
 בן אדם

קטע מן השאלון לדוגמא - המשתתפים צריכים לסמן עבור כל שיר האם מדובר במכונה או במשורר. בנוסף, נשאלו המשתתפים בסיום השאלון מהי מידת ההבנה שלהם וההכרות עם שירה עברית

6. תוצאות

לאחר ביצוע fine tune למודל הבסיס על פי דאטה-סט מכל קטגוריה, קיבלנו את המשקולות הסופיים של כל מודל וניגשנו לשלב ה-inference, למעשה שלב ייצור השירים. הזנו למודל את שם השיר, והוא המשיך בייצור השיר.
להלן מספר דוגמאות לשירים שקיבלנו מכל קטגוריה:

מודל שירת ביאליק

אהבה

אהבה נכזבת,
אוי,
עברה נתיב,
אל-עיני כל באי,
נמשכת,
עד-ההר.
יחדו,
עגילי זהב,
בים ובקצפני יין;
בטור השגיונות,
משולים גם-יחדו
אהבה

ישראל

עומד ומביט אל- לאחור הר ציון,
פורע לארץ צוואר כאבני
ההולכת עזובה כרמק של-זהר,
פושטת רגליך וזר
ישראל

כנפי השכינה

כנפי השכינה ותבער מחמות,
ונפשי תשמר עלי ותלהט עיני,
ובקשת שדי אל-לבבי : להכנס!
ותמחאי רזי עולם, תשקני כלתה,
ותזל דמעתנו מחיי נפשך,

ואתה בגללך, והיית לצדי המדברה,
וברחוב עיני מלאתי גחלי,
ותרא והיית לי עיני מענקך;
וחמדת אותה בת-עין עד-בוש זיו דמי,
ורוחי חבקי,
ועמדת את כל-חמת חיי בה בשעה זו –
ורוחי נפשי בהרהורי אלהים ובדמות חיי;
ועמדת את-ידך בפוך רב

מודל שירת רחל

אביב

אביב גלעדי עם שיריו
לחייך, בהיתי קולי. מוזר הוא למדרך על
גבעולי קדמתי לתוך לבבי.
מר אשר ילך ויתהלל ויהי

בגני

בגני אשכח
שולט בסתרי
מוטל כמזכרת
על יד מושטת
בודד
שועלים ומובחרים
רך כעלים וטובים
זעפם וחרישו,
בוטחים, מגנים,
נצורים
אחרי הכל
כיומיים
לפני רעל.

מודל שירה מודרנית

אהבה

אהבה, איזו שמחה.
אומרת. ומה את צריכה?
השנים עוברות לאט,
זמן, זמן קצר.
אחת, זמן, זמן קצר.
ידך פוקחת אותן.
אני מרגיש איך בכלל הבנתי
שמבינים
מבינים, ממתינים, מבינים
שיש גם איזו בחורה
יש בחורה כזאת טיפשית.
רק המבט הזה שווה לנשום
לחשב מה צריך לקרות כשאינן חפץ.
פשוט לא לאבד את הראש
ברגע הלא נכונה –
שאלה אם יש לי בעיה.
נניח לך שאני טועה, די.

גשם

גשם, קדש, מאיר
ואיתו רב המכתבך, בנך שלי,
ראי, היית גם אבי. עולם ומלואו.
אם אסלח:
דרכו של עולם לא תשכח.
אם אסלח : עכשיו אני עלם.
אם אסלח:
ילדי, ילדי, אם אסלח!

ירוק

דרכו אדמה,
רן הבליחו
את הסתו.
לא נגרם נזק.
הכביש נדם,
נגה ממושב האספלט.

הכביש, בלם מסתו.
וגשם ירד וגוע בבגד מר
מש ואריק, כאבר רך
על רע לדרכו.

מודל שירת ימי הביניים

אהבה

ופני אבכה רגלי
חסדיך נפשי שבי בי
ובם תלבשי תרעני
אדני ושמי יצרי
כבוד אל מחלתו
בצדקתו ובמותו יחמלו
מחריבך והעמידה
לא אתבונן ולא הם בנו
רגליו וצוף עטרותיו
בשוררי היו לסתוריו
תשא נאוניו
ולא יהגה מעמד עמם והחלו
יגיעי תמיד בשמיהם
ואולם פני אל ישמחו
נבערו בקול אדני והנואל
אלי ידיד רוח תשמעו ראשי
לבש חזון ולחקר

שירה

הבה נשבתה כמו תמיד
במלים קיומיות,
כאל החי הנפל
במותנים קצובים היא נזמתה
מתפתלת מחול
חרוזי נוצותיה,
שפע אהבה
אל קללתה הנבוכה
מתפנקת אל אסופת חייה
אחותי המנוחה

מודל תקופת תחיית העברית

אהבה

אהבה, לטבע וכרוב
התמר רב אותם ונהנים ונעימים,
בשתיקה ובמוסיפות חן וצליל שבת.
הבריקו עסיס עתידים ופרחו,
הלבנה אם גזו עפר השכינה.
פרשת סתיו אלוהיו בציפורניו,
משה ורבקה בצפצפו כליון,
אזנו צלח כולו ויעזוב את ביתו.
פארות כל-ניר וכרמי החתומים
אחד איש מלמד מנדודים המלבינים,
עלי-גמרא ונופם בחוכמות וישישים,
בחן וקנין ותרגמן ולב ולב ער.

מודל ננופואטיקה

הלב

הלב שלו הולך ומתרחב.
אני נופל על האדמה
מעל האוקיינוס.
אישה שולחת לו את זרועותי
נושק לרוקן את שמיכת הלילה.

אהבה

ההתגלות התאומה של התבגרות הידוענית של הגוף.
אין לדעת אם הגוף נברא עם זכרון מוזר או לא.
ההתגלות התאומה של הגוף מושגת ללא ביטוי ממשי או בלתי
בלתי ממשי וללא קשר איתי. אם הגוף נברא עם זכרון מוזר ונשגב זה
מונוני ומרענן ללא שיבוט ושיבוטים ביניהם
אהבה לא פותרים.
אהבה משותפת.

אהבה

את
ניסיתי
לדבר בשפתוטייך
ובפני קהות יופיה
למות מצמא כרות יומית
ואת, את
כותבת שירה

שירה

אני נבהל ממשאלה
שכמותן מבעבעת
רגשי נעלים
מתרוצצים בחוטים

בהתבוננות ראשונית על השירים שהפיקו המודלים השונים, נראה כי מבנה השיר בא לידי ביטוי בצורה נאותה, שכן מרבית השירים שהפיקה המכונה מכילים ירידות שורה מרובות ושורות קצרות, כנהוג בסוגה ספרותית זו. בנוסף, ניתן לראות הבדל בין המודלים השונים, וכן התאמה למחלקה אותה מנסה המודל לחקות: במודל הננו פואטיקה למשל, שירים רבים הכלו שורות ארוכות יותר מאשר בשירים הנוצרו על ידי מודלים אחרים, ובהתבוננות קצרה על סט האימון זה אכן מאפיין יותר מחלקה זו. כמו כן, בשירת ימי הביניים ניתן היה להבחין בשירים ארוכים יותר מהשאר, וגם כאן זה מאפיין שניתן לזהות בסט האימון (כאשר אנו מריצים את קוד inference עד ייצור stop-token, ולא לפי מספר מילים מוגדר מראש, ניתן לבחון מאפיין זה).

אם נתבונן על תוכן השירים, אפשר לראות מאפיינים נוספים, חלקם בולטים לעין, המשתנים בין מחלקות שונות. השפה למשל שבה משתמשים בשירים מתפתחת עם ההתפתחות הכרונולוגית של קבוצות המודלים: שירת ימי הביניים כוללת שפה יותר ארכאית, דתית, וכן התוכן נראה עוסק יותר בשמים, חסדים, אלוהים, חורבן וכו' - כל אלה מאפיינים את הכתיבה בתקופה זו. בשירת התחייה העברית ניתן לזהות מילים ותכנים האופייניים למשוררים בתקופת ראשית הציונות ותחיית העברית: שדות (ניר), עסיס, גידולים חקלאיים (תמר, כרמים), ואף "חדי עין" יכולים להבחין ב"מאבק" הפנימי ששרר בקרב משוררי תקופה זו, בין העיסוק בתחומים הני"ל, שסימנו את היותם חלוצים חילונים, לבין השמירה על התרבות והמסורת (גמרא, צליל שבת, שכינה וכו'). כך גם בקרב ביאליק ורחל, ניתן לזהות את סגנונם ושפתם הייחודית, לדוגמה אצל רחל בולטים מאוד המלנכוליה, הקושי והתסכול שליוו את חייה ושירתה ("בודד", "יומיים לפני הרעל" וכו'). לבסוף בשירה המודרנית ובננופואטיקה, אנו רואים שימוש באוצר מילים ותכנים

יותר יומיומיים (כבישים, בחורה טיפשית, לאבד את הראש וכו'). בשירת ננופואטיקה גם ניתן היה לזהות טקסטים שקצת שוברים את הצורה של שיר קלאסי, לדוגמה בשיר "אהבה" שנשמע כמו הגיגים על נושאים מוזרים כמו שיבוטים.

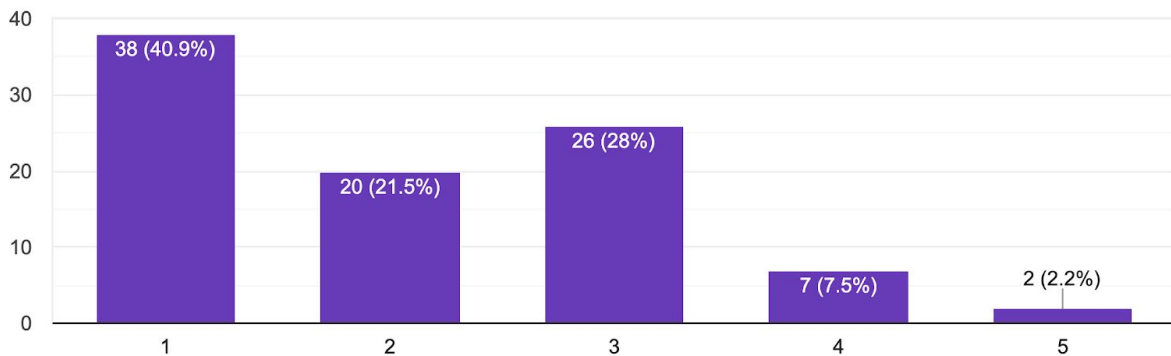
יחד עם האמור לעיל, ניתן לזהות גם מגבלות בתוצרים שקיבלנו. קיימים מקרים של חוסר קוהרנטיות, חוסר התאמה בין מילים שונות (ברמה הסמנטית או הסינטקטית), או קושי לזהות רעיון כולל, קו מחבר ורבדי עומק.

תוצאות סקר ההערכה:

בסקר השתתפו כ-95 נבדקים בעלי הבנה ממוצעת בשירה ושקראים שירה בתדירות נמוכה עד בינונית. להלן הנתונים אודות אוכלוסיית הנבדקים (1=תדירות/הבנה נמוכה, 5=תדירות/הבנה גבוהה):

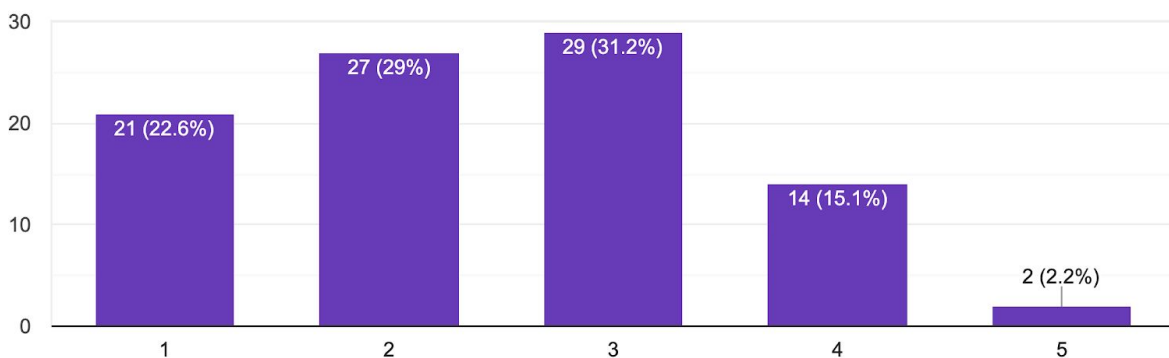
מהי התדירות בה אתה קורא שירה?

93 responses



מה מידת הבנתך בשירה עברית?

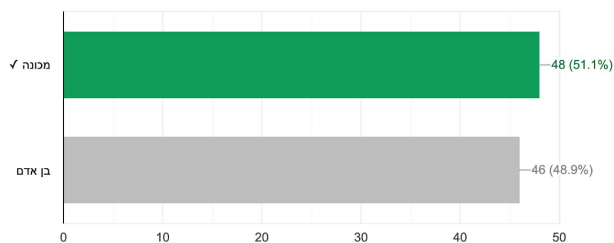
93 responses



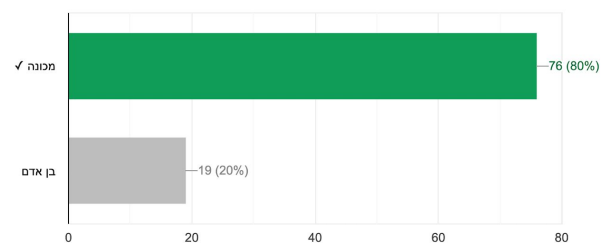
הציון הממוצע של הנבדקים במבחן היה 34.89 מתוך ציון מקסימלי של 55 נקודות (כל תשובה נכונה עבור 11 השירים מזכה ב-5 נקודות). הציון החציוני הינו 35 מתוך 55, הציון המינימלי שהתקבל הינו 15 והמקסימלי הינו 55.

ברוב המקרים הזיהוי לא היה חד משמעי. בשיר הכי פחות טוב (מבחינת אחוז המזהים אותו ככזה שנכתב על ידי מכונה), "הלב", כ-80% מהנבדקים (76 מתוך 95) זיהו שהוא נכתב בידי מכונה. בשיר הכי טוב לפי מדד זה "אהבה" (השיר "אהבה" השני שמופיע במקבץ השירים של מודל "ננופואטיקה") כ-51% (48 מתוך 95) מהנבדקים זיהו שהוא נכתב ע"י מכונה, כלומר הנבדקים לא הצליחו להבחין האם הוא נכתב בידי אדם או מכונה. להלן הגרפים עם התוצאות הני"ל:

אהבה
48 / 94 correct responses

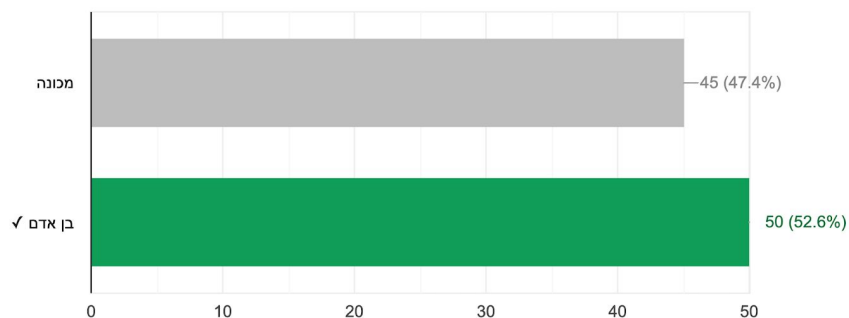


הלב
76 / 95 correct responses



ניתן לראות שגם עבור שירים שנכתבו על ידי משוררים אנושיים לעיתים היה קושי בזיהוי, לדוגמה עבור השיר "אבא" שכתב משה בסוק, כמעט חצי מהנבדקים חשבו שהוא נכתב בידי מכונה:

אבא
50 / 95 correct responses



עובדה זו מדגימה בין השאר את התופעה לפיה הנבדקים "החמירו" בבחינה שלהם את התוצרים, וניתן להניח שעצם הידיעה שחלק מהשירים נכתבו על ידי מכונות יצרה מראש bias בהערכה. אנו מעריכים

שהנבדקים דקדקו יותר בבואם להעריך את השיר כאשר המשימה הייתה לזהות נכונה בידי מי הוא נכתב ולזכות בניקוד בהתאם, מאשר אילו הם היו ניגשים לקרוא אותו באופן נאיבי.

עבור שיר שנכתב על ידי המכונה כ-67% מהנבדקים זיהו אותו ככזה בממוצע (האחוז הממוצע עבור 6 השירים בסקר שנכתבו על ידי מכונה). עבור השירים שנכתבו על ידי משורר אנושי התקבל 62% זיהוי ממוצע (ממוצע של אחוז הזיהוי כמשורר אנושי עבור 5 שירי משוררים שפורסמו בסקר).

אם אנחנו מצמצמים את האוכלוסיה ומנתחים רק לפי תוצאות הנבדקים שמבינים וקוראים פחות שירה (נבדקים שענו תשובות 1-2 בשאלות מידת ההבנה בשירה ותדירות קריאת שירה. נבדקים אלה מהווים כ-44% מהנבדקים), אנו מקבלים תוצאות מעט יותר טובות: לדוגמה עבור השיר "אהבה", כ-54% מאוכלוסיה זו חשבו שהוא נכתב בידי משורר אנושי, כך שהוא "ניצח" במבחן טיורינג. ראוי לציין שמתוך רצון לבצע הערכה כמה שיותר אמינה לטיב השירה המיוצרת, השתדלנו להפיץ את הסקר לנבדקים רבים אשר מבינים בתחום (ביניהם משוררים, מורים לספרות, אקדמאים בתחומי מדעי הרוח ועוד).

7. שלבים במהלך הפרויקט

כבר בשלב הראשוני לפרויקט זה, ידענו כי ברצוננו לבנות מכונה לייצור שירים. רצינו לעשות זאת בצורה המשלבת "פילטרים", כלומר לאפשר למודל לייצר שירים בהתאם לפרמטרים מסוימים, כמו למשל סגנון שירה, משורר או נושא השיר.

חשבנו על האפשרות לבנות מודל שיוכל לקבל כקלט אחד מפרמטרים אלו, וכך מודל אחד יהיה מסוגל לייצר מגוון גדול של שירים המתאימים למגוון קטגוריות.

מספר התמודדויות לא פשוטות במהלך הפרויקט הובילו אותנו לשנות מעט את הכיוון הראשוני: ראשית, איסוף הדאטה היה שלב מאתגר, ובפרט ייצור דאטה יחד עם תיוג המסמן את נושא השיר היווה אתגר בפני עצמו. לא היו תיוגים כאלו מראש בדאטה שאספנו, ועל מנת להשיג כאלו, היינו צריכים להשתמש במודלים אחרים, שתהינו לגבי מידת הצלחתם בזיהוי נושאים עבור טקסט מסוג שירה. העדפנו להתמקד בשאר הקטגוריות, כמו משוררים או סגנון השירה, ולהמשיך לשלב בניית המודל. בשלב הבא, היינו צריכים לעבד את הדאטה שאספנו - ובתור התחלה לבצע טוקניזציה.

שקלנו דרכים שונות לביצוע הטוקניזציה, על בסיס כלים שלמדנו בכיתה, ונראה היה כי YAP המתאים ביותר למשימה זו, עבור השפה העברית, והוא התחשב ברכיבים שונים הייחודיים לשפה, וכן במורכבות שלה. על מנת להשתמש בכלי זה, הורדנו אותו אל המחשב האישי והרצנו אותו על מספר רב של שירים. במהרה הבנו כי YAP הוא אמנם כלי מצויין, אך איטי יחסית, ובשל הזמן המוגבל שהיה ברשותנו וכמות הדאטה שהיינו צריכים לעבד לשם האימון, החלטנו לנסות למצוא כלי אחר.

שלב נוסף היה למצוא ייצוג ראוי לדאטה (embedding). תחילה, חשבנו על שימוש במודל fast-text שאומן מראש על דאטה בעברית. במהרה הבנו כי שימוש בכלי זה עבור שירה עברית לא יניב הרבה: הכלי אומן על דאטה במשלב שפה בינוני, ובפרט לא ספרותי וכן מבנה הטקסט היה שונה לחלוטין מזה של שירה עברית. לאחר סקירת הכלים הקיימים כיום הבנו ש embeddings מוכנים מראש בעברית יתקשו לייצג טקסטים מסוג שירה, וכן לייצג אלמנטים כמו ירידת שורה וסימני פיסוק. עובדה זו הניעה בנו את המחשבה לאמן embedding בעצמנו - אך הבנו כי אין לנו מספיק כוח חישובי לביצוע אפשרות זו. הבנה זו הובילה אותנו לכיוון של transfer learning. כל אלה, יחד עם ההנחה שקשה יהיה לייצר שירה בעזרת חוקים או פיצרים מוגדרים מראש, שכן בבסיסה היא תהליך יצירתי, חדשני, שבכל תקופה מנסה לפרוץ גבולות טקסטואליים וסגנוניים, וכן היעדר תיוגים לדאטה הנדרשים לטובת למידה מפוקחת, הביאו אותנו לבחור במודל של רשת נוירונים. הנחנו שלמידה אימפליציטית של כתיבת שירה יאפשרו לנו להגיע לתוצאות טובות יותר. ידענו שבשימוש במודל כמו למשל GPT-2 נוכל גם "לעקוף" את בעיית ה-embedding, משום שגם היא נלמדת בצורה אימפליציטית על הדאטה עצמו, שמוזן כטקסט. כמו כן, הייתה לנו כאן את האפשרות להשתמש במודל pre-trained, כך שייחסך מאיתנו האימון from scratch למיליוני הפרמטרים של המודל. בצורה זו, הבנו שנוכל להגיע לתוצאות גם בעזרת כוח החישוב המוגבל שעמד לרשותנו.

8. מסקנות

מהתוצאות נראה שאפשרי ליצור בעזרת מכונה טקסט מסוג שירה, ואף לשלוט על הסגנון שלו כך שידמה לסגנון תקופה מסויימת בשירה העברית, או אפילו לסגנון של משורר ספציפי. נראה שהרמה של השירים המיוצרים על ידי מכונה טרם הגיעה לרמה של שירה הנכתבת על ידי אדם (ואכן כפי שפרטנו בהתחלה, זו תת-משימה קשה בתוך המשימה של ייצור טקסט, ואפילו בתוך תת-המשימה של כתיבה יצירתית), אך כבר ניתן להגיע לרמה גבוהה יחסית, ובמקרים מסוימים גם לכזו שאינה ניתנת להבחנה ברורה האם השיר נכתב בידי אדם או מכונה (ראינו דוגמה לשיר כזה בעבודה, "אהבה" השני של "ננופואטיקה"). הקוהרנטיות של השירים הרגישה כנקודת תורפה במבחן שלנו להערכת התוצר, אשר טומנת בחובה פוטנציאל לשיפור השירים (ואכן, כששירים זהו ככאלה שנכתבו בידי מכונה, הקוהרנטיות הייתה אחת מדרכי הזיהוי העונים על הסקר ציינו שהשתמשו בהם כששאלנו אותם על כך). מכאן ששיפור מימד זה יוכל להביא לשיפור משמעותי בתוצאות ובמדדי ההערכה. לבסוף, מעניין היה לראות בתוצרים שקיבלנו שימוש בביטויים מוכרים של המשוררים מסט האימון (אלו אמנם מאפיינים את המשורר, אך עלולים להוות גם "חיקוי"). דוגמה לכך ניתן היה לראות בביטוי "יד מושטת" בשיר שייצר המודל לשירת רחל, המופיע גם בשיר "גן נעול". נקודה זו מעלה את השאלה האם המכונה יודעת רק לחקות, או שיש ביכולתה להיות יצירתית וחדשנית. זו אף נקודה שיכולה להיות פתח למחקר עתידי בכתיבה יצירתית.

9. רעיונות לעתיד

בהינתן זמן וכוח אדם נוספים ראשית ניתן היה לאסוף יותר דאטא, ואף לתייגו, ובכך הן לטייב את התוצאות וליצור פילטרים מסוגים שונים ומעניינים כמו למשל נושא השיר. כיוון מעניין נוסף שניתן לנסות הוא שימוש בGAN לטובת המשימה. כפי שביקשנו מאנשים לזהות האם השיר נכתב על ידי אדם או מכונה, ניתן היה להגדיר שאלה זו כמטרת הסיווג של discriminator בארכיטקטורת הGAN, וכך אולי להגיע לתוצאות "אמינות" יותר, או כאלה שמרגישות "טבעיות" יותר מבחינת היכולת לזהות אם הכותב הוא אדם או מכונה.

שיפור נוסף יכול להיות לבנות מודל שמלבד קלט התחלתי של שיר (במקרה שלנו: שם השיר), מקבל גם פילטרים נבחרים, ויכול להוציא מספר שירים לפי פילטרים אלה. כלומר לא להפריד למודלים נפרדים, אלא ליצור מודל אחד שיתאמן על כל הדאטא וידע להבחין וליצור סוגים שונים של שירים בהאם לפילטרים איתם בוחרים להתמודד. זו הייתה גישה שחשבנו עליה בתחילת הפרויקט, אך וויתרנו עליה משום שלא היו לנו מספיק משאבים וזמן לאמן מודל גדול from scratch.