

פרויקט גמר בהנדסה :
דו"ח מספר 1

שם הפרויקט : Gaimification for data collection

מגישים : אלירן לאור ומוטי הולצמן

שם המנחה : ד"ר דפנה שחף – האוניברסיטה העברית
dafna.shahaf@mail.huji.ac.il

1. אז מה בעצם הבעיה שנתמודד איתה?

Gamification – רקע:

כולנו שמענו בהזדמנות זו או אחרת על הנפלאות של ה-"Big Data" ועל התרומה שניתן להפיק ממסדי נתונים גדולים. עם זאת, לא תמיד מובן מאיפה מגיע כל ה-"Big Data" הזה. יש חברות שמחפשות את המידע ברשת או בארכיונים כדי למצוא כל פיסת מידע שתעזור לה בתחום בו היא עוסקת. בעיה נוספת שעולה, היא שכיום כמות המידע שנמצאת ברשת ובכלל, כל כך גדולה כך שקשה לדעת איך לדלות את המידע המהימן אותו אנו צריכים.

הבעיה הגדולה היא שאיסוף מידע מהימן ורלבנטי עלולה לעלות המון – הן בכסף והן בזמן. עם זאת, ישנו ייתרון עצום באיסוף מסד נתונים ייחודי משלך, בניגוד לשימוש במסדי נתונים קיימים. לאור זאת, מה הדרך בה ניתן לאסוף מידע בצורה יעילה? ישנן חברות שאוספות נתונים ע"י סקרים, ומתן מתנות קטנות למשתתפים בסקר. עם זאת, שיטה זו לא מוכיחה את עצמה כיעילה מאוד בטווח הרחוק ובבעיות בהן יש צורך באיסוף מסד נתונים גדול מאוד.

כאן נכנס Gamification, או בעברית משחוק, לתמונה.

משחוק היא שיטה בה משתמשים באלמנטים "משחקיים" בהקשרים לא משחקיים. במשחוק לרוב מכניסים עיצוב משחקי כדי לשפר יעילות משתמש, פעילות משתמש וכו'.

שיטות במשחוק מיועדות למנף את טבעם של האנשים להתחבר, ללמוד, להתחרות, ליצור סטטוס חברתי או אפילו סתם רצון להמשיך לשחק. שיטות מוקדמות של משחוק השתמשו במתן פרסים לשחקנים שהצליחו במשיחה מסויימת; או יצירת תחרות בין אנשים. סוגים שונים של פרסים הם: ניקוד במשחק, תגים, דרגות, מילוי סרגל תהליך או אפילו מתן מטבע וירטואלי למשתמש.

דוגמא למשחוק מהחיים:

אתר הכושר Fitocracy משתמש במשחוק כדי לעודד את משתמשיו להתאמן בצורה יעילה יותר. למשתמשים מוענק מספר משתנה של נקודות על פעילויות שהם מבצעים באימוניהם והם מעפילים לרמות גבוהות יותר בהתבסס על הנקודות שנצברו. משתמשים יכולים גם להשלים משימות (מערכות של פעילויות קשורות) ולהשיג תגי הישג בעבור אבני דרך בתחום הכושר. בהקשר זה, למשל, בתחרות משחוק בשם The Fun Theory שיזמה חברת פולקסווגן, אחד הרעיונות שעלו היה לצבוע מדרגות בתחנת רכבת תחתית כקלידי פסנתר המשמיעים צלילים כשדורכים עליהם. היוזמה העלתה את השימוש במדרגות ב-66%.

בפרוייקט שלנו:

אנחנו נשתמש בעקרון המשחוק בפרוייקט על מנת ליצור מסד נתונים גדול שישמש אלגוריתם ללימוד מכונה אשר המטרה העיקרית שלו היא לאפשר פישוט של משפטים. כלומר, אנחנו נכתוב משחק שמטרתו לאסוף זוגות משפטים (או מילים) בעלי משמעות זהה, האחד מורכב והשני פשוט, כאשר או הפישוט של המשפט או הפיכתו למורכב יותר תעשה ע"י המשתמש, כאשר המשחק גם נותן לו פידבק ויודע לזהות האם הקלט אכן בעל זיקה לוגית למשפט אותו היה צריך לפשט.

2.מדוע הפרויקט חשוב ומעניין:

ראשית כפי שתיארנו, הפרויקט מתממשק עם מספר נושאים עיקריים, וביניהם הנושא של machine learning שנמצא כיום בחזית המחקר במדעי המחשב. בעזרת הנתונים שיאספו ע"י המשחק אפשר ליצור מסד נתונים רחב ומדויק שממנו ניתן, בעזרת אלג' לומד, ליצור מספר גדול של אפליקציות. למשל, במידה ומחשב ילמד לפשט משפטים בצורה טובה מאוד, ניתן יהיה להפיק תרגום של התנ"ך או להבדיל תרגום של רומיאו ויוליה לשפה פשוטה ובצורה כזו להנגיש טקסטים מסובכים לקהל יעד רחב הרבה יותר. בצורה דומה – אפשר לחשוב על הפעולה ההפוכה, לקחת טקסט שנכתב בשפה פשוטה ולהפוך אותו לטקסט בשפה גבוהה. בנוסף, גם למשחק עצמו יכולות להיות מספר אפליקציות: יהיו כאלו שיוכלו לשחק בו בשביל הכיף, ויהיו כאלו שיכולו למשל להתאמן באמצעותו על אוצר מילים רחב באנגלית לדוגמא לצורך הבחינה הפסיכומטרית וכדומה.

3.מהם אתגרי המפתח בפרויקט:

ראשית, אנחנו נכנסים לתחום אשר יש בו בעיות שאינן פתורות, נציג כמה מהן: בתחום ה-NLP, אין פתרון סגור למציאת שקילות לוגית בין שני משפטים. אחד האתגרים הקשים בפרויקט יהיה להבחין האם הקלט שהוכנס ע"י המשתמש כפשוט למשפט נתון, באמת שקול אליו לוגית. שנית, בתחום הלמידה, חשוב לנו להבין איך האלגוריתם הלומד ינתח את מסד הנתונים שלו, ע"מ שנוכל לאסוף את המידע המתאים. שלישית, האתגר ההנדסי: תחום המשחק דורש היכרות רבה עם בניית אפליקציות לסלולר או פלטפורמת web. בנוסף, יידרש מאתנו לחשוב על משחק מספיק מעניין ומאתגר על מנת שהמשתמש ירצה לשחק בו.

4. מה חדשני בפרויקט שלנו, ומדוע הוא שונה מפתרונות אחרים:

ראשית, תחום המשחק הינו דרך חדשנית ויצירתית להשיג מידע רלוונטי ובעלות מינימלית. במקום לשלם כסף תמורת המידע, יהיה מי שינדב לנו אותו בחינם. בנוסף הפרויקט שלנו שונה שיש לו יכולת לשמר את המשתמשים. אם המשתמש יהנה מהמשחק – הוא ישחק עוד. כתוצאה מכך הוא יתרום עוד מידע ובכך נוכל להגדיל את המאגר באופן תמידי ולא להתבסס על תרומה חד פעמית של משתמש. בנוסף, אנו ננסה להגיע למודל מוצלח דיו בהקשר השקילות הלוגית של המשחקים על מנת שנוכל ליצור מסד נתונים מהימן.

5.מה הם רכיבי המפתח של הגישה שלנו?

רכיב המפתח העיקרי הוא יצירת מנגנון פידבק איכותי. כלומר, נצטרך להכריע בצורה כלשהי האם הקלט שהמשתמש הכניס אכן מתאר את המשפט שכן אחרת נזים את מסד הנתונים שלנו ונפגע במטרת הפרויקט. רכיב מרכזי נוסף הינו הממשק עם המשתמש: כיוון שהמשתמש הוא למעשה מקור מסד הנתונים, אנו חייבים ליצור פלטפורמה נוחה, מפתה וידידותית עבורו. נרצה לחשוב על משחק מהנה ומנגד כזה שניתן לביצוע במסגרת הפרויקט.

6. מדוע אנו חושבים שהפרויקט ישים?

ראשית בתחום המחקרי: למרות שהתחומים הללו חדשניים, ישנן דריסות רגל של חוקרים באקדמיה ובתעשייה בתחומים האלה, כמו למשל google או קוד פתוח. כלומר, הנושא הזה מעסיק חוקרים רבים ולכן נוכל להישען על מחקרים ופיתוחים בתחום שעוסקים בשאלות דומות. בנוסף, בתחום המעשי: אנו נמצאים בעידן טכנולוגי בו מידע על פיתוח אפלקציות ומשחקים זמין ומונגש באינטרנט, ואנו מקווים לעשות בו שימוש על מנת ליצור פלטפורמה שתתמוך בפרויקט.

7. איזה מידע עלינו לרכוש לפני שנתחיל לעסוק בפרויקט:

לאחר התייעצות עם המנחה שלנו והבנה של הידע אותו עלינו לרכוש כדי שנוכל לעסוק בתחומים האלו, אנו נתמקד בנושאים הבאים:
בתחום ה-NLP נתמקד בנושאים הבאים:

semantics, word embedding's, simplicity and readability scores, preprocessing

בתחום המשחק:

נרצה ללמוד מה נעשה כבר בתחום, אילו רעיונות עובדים ואילו לא. בנוסף נרצה להתעמק קצת באלמנטים שגורמים למשתמש להישאר במשחק כגון תמריצים ומתן מוטיבציה. נוכל להיעזר בעקרונות שמופיעים בספר הבא:

karl m. kapp the gamification of learning and instruction prof.

בתחום הקורסים:

בתחום המחקרי: נרצה ללמוד את הקורסים הבאים על מנת להשיג הבנה טובה יותר של הנושאים גם ברמה הפורמלית: הקורס ב-NLP והקורס ב-IML. בתחום המעשי: על מנת ליצור משחק ישנו צורך בידע תכונתי נוסף מעבר לנלמד במסגרת התואר. את הידע הזה אנו מתכוונים לרכוש באמצעות מספר כלים: עבודה מעשית בתחום התכנות בחברת אינטל בה אנו עובדים כיום וקורסים המוצעים דרכה בנושא.

בנוסף קורסים מקוונים בתחום תכנות המשחקים והאפליקציות. כמו כן, בפגישה שערכנו עם רונן בנושא טרם הגשת הדו"ח, הבנו כי קיימים מוקדי ידע בנושא באוניברסיטה בהם נוכל להיעזר.

8. איזה ידע בתחום התוכנה נדרש מאתנו להשיג או ללמוד כדי ליישם את הפרויקט?

בנוסף לידע אותו העלינו עד כה, בצורה ספציפית יותר:
עבודה עם התוכנה Android studio על מנת ליצור אפלקציה.
עבודה משותפת בסביבת github
התמקצעות בסביבת LINUX

9. מה הוא לוח הזמנים שלנו?

עד התאריך ה-10.6 : למידת git בצורה בסיסית לצורך עבודה שוטפת.

עד התאריך ה-15.6 התמקצעות בסיסית בסביבת LINUX.

עד התאריך ה-20.6 : קביעת פגישה מסודרת עם המנחה להפקת גאנט מעשי לפרויקט.

עד התאריך ה-28.6 : החלטה בסיסית על סוג המשחק אותו נרצה לממש מבין כלל האפשרויות שנחשוב עליהן עד תאריך זה (בשיתוף המנחה).

עד התאריך ה-28.6 : התקנת סביבת Android Studio והיכרות בסיסית עם הממשק שלה.

לאחר הסמסטר – חופשת הקיץ : לימוד ראשוני של הנושאים הנדרשים ב-NLP על מנת שנוכל להגיע מוכנים לסמסטר ולמימוש הפרויקט.