פרויקט תכן בלמידה חישובית

חלק ראשון – תרגום משפטים לא משכנעים לכאלה משכנעים

בחלק הראשון של הפרויקט ניסינו להפוך משפטים לא משכנעים לכאלה משענים. לשם כך מצאנו דאטא של IBM בו היו כ15 אלף זוגות של משפטים, כאשר לכל זוג הוגדר מי קיבל את הציון הגבוה ואת מידת ההסכמה של השופטים (בין 0 ל1). לקחנו אך ורק זוגות בהם מידת ההסכמה עלתה על 0.8, כדי שההחלטה תהיה כמה שיותר מובהקת. בנוסף, היות והמשימה נראתה לנו מאוד קשה, לאחר כמה ניסויים בהם המודל לא הצליח ללמוד, בחרנו להוריד את השונות של הדאטא ולהישאר רק עם משפטים שהנושא שלהם הוא צמחונות. נשארנו עם בערך 1500 זוגות משפטים.

המודל עדיין לא הצליח ללמוד לתרגם בצורה טובה (תוצאות לדוגמה ניתן לראות בטבלה), בין היתר היות ואין בהכרח קשר בין המשפט המשכנע לאחד שפחות. בקובץ שלקחנו כל זוג טיעונים הוא טיעונים שנאמרו כתגובה אחד לשני, כלומר בדרך כלל זה לא אותו הנושא, למשל אחד שידבר על חוקים ואחד שידבר על אקולוגיה.

לכן, החלטנו לשנות את המשימה ליצירת טיעונים מתוך מילות מפתח.