הצעת פרויקט- שלומה טימסיט, ספיר בקשי אטיאס

תז 332376813, 212125660

נושא הפרויקט- סיווג כוונות משתמשים בפניות שירות לקוחות פיננסיות באמצעות מודלי שפה

תיאור כללי

Intent Detection היא משימה מרכזית במערכות Conversational AI וצ'אטבוטים, שבה המערכת צריכה להבין מהי מטרת הפנייה של המשתמש מתוך טקסט חופשי. בפרויקט זה נתמקד בפניות שירות לקוחות בתחום הבנקאות והפיננסים, שבו קיימות 77 כוונות שונות ולעיתים דומות מאוד זו לזו מבחינה סמנטית.

תיאור הדאטה

נשתמש ב־ BANKING77, דאטה-סט אקדמי. המאגר כולל 13,083 פניות שירות לקוחות בתחום הבנקאות, המסווגות ל־77 כוונות שונות. הדאטה מתמקד ב־ fine-grained single-domain intent detection, כלומר סיווג כוונות עדין בתוך תחום אחד מוגדר -בנקאות ושירותים פיננסיים .

כל דוגמה כוללת טקסט קצר של פנייה מצד לקוח ותווית intent מתאימה. העובדה שיש 77 מחלקות הופכת את המשימה למאתגרת יותר מסיווג רגיל למספר קטן של קטגוריות, משום שחלק מהכוונות קרובות זו לזו מבחינה סמנטית.

שאלת המחקר

ביצד שיטות שונות מבוססות מודלי שפה מתמודדות עם משימת Intent Detection רב-מחלקתית בתחום פיננסי, ובפרט עם הבחנה בין כוונות קרובות סמנטית?

מתוך שאלה זו נבחן כמה שאלות משנה:

1. האם baseline קלאסי כמו TF-IDF + Logistic Regression מצליח להתמודד עם 77 intents, או שהוא מתקשה כאשר המחלקות קרובות סמנטי?
2. האם שימוש ב־ sentence embeddings משפר את איכות הסיווג לעומת ייצוג מילים קלאסי?
3. מהו הערך המוסף של Fine-tuning מלא למודלי Encoder ייעודיים וקטנים (כגון BERT, DistilBERT, RoBERTa)?
4. האם שימוש בטכניקות כוונון יעיל בפרמטרים (PEFT) באמצעות LoRA/QLoRA על מודל שפה גנרטיבי בינוני (בטווח של 12B3 - B פרמטרים, כגון Llama-3-8B) מציג עליונות על פני שיטות ה-Fine-tuning המסורתיות של מודלים קטנים?
5. האם LLM prompting ב־ zero-shot או few-shot מצליח להתחרות במודל שעבר אימון ייעודי על הדאטה?
6. האם שילוב ואגרגציה של מודלים או פרומפטים שונים באמצעות מודל אנסמבל (Ensemble Model) המבוסס על רוב קולות (Majority Voting) מייצב את הניבויים ומשפר את הדיוק הסופי?
7. אילו קונספטים או מאפיינים (Features) מיוצגים בתוך המודלים ומנחים את הניבויים שלהם, ומה ניתן ללמוד מניתוח שגיאות סיסטמטי (Interpretability) על המגבלות הסמנטיות של המודלים השונים?

כיוונים מתודולוגיים אפשריים

נשווה בין מספר גישות:

1. **Baseline קלאסי** –ייצוג הטקסט באמצעות TF-IDF ואימון Logistic Regression או Linear SVM .
2. **Sentence Embeddings**- המרת כל פנייה לוקטור סמנטי באמצעות Sentence-BERT ואימון מסווג פשוט.
3. **Fine-Tuning**- התאמת מודל Transformer קטן כגון BERT או DistilBERT למשימת סיווג רב־מחלקתי.
4. **LLM Prompting**- שימוש במודל GPT לסיווג הכוונה באמצעות Zero-Shot ו־ Few-Shot Prompting.

5. **PEFT-** הרצת כוונון יעיל בפרמטרים באמצעות טכניקות LoRA / QLoRA על גבי מודל שפה פתוח בגודל בינוני (כגון Llama-3-8B או Mistral-7B).
6. **Ensemble Modeling**- בניית מנגנון אגרגציה המחבר את התחזיות ממודלים שונים או מפרומפטים שונים באמצעות שיטות הצבעה (Majority Voting) לטובת שיפור היציבות (Robustness) והדיוק.
7. **ניתוח אנליטי ו-Interpretability-** מעבר לבחינת מדדי הביצועים היבשים (Accuracy, F1-Score), נבצע ניתוח שגיאות סטטיסטי וסיסטמטי מעמיק (לרבות שימוש במטריצות מבוכה מורחבות). כמו כן, נבצע בחינה של הייצוגים הנלמדים (Embeddings) ומרחב הווקטורים של המודל המאומן, במטרה לאפיין אילו קונספטים המודל מקודד ועל אילו פיצ'רים הוא מסתמך בזמן הניבוי.