Sentiment Analysis for Jordanian Dialect Using Memetic Algorithm and Support Vector Machine Classifier

Prepared by

Ahmad Muflih Alhadab

Supervised by

Dr. Ghaith Jaradat

Abstract

Sentiment analysis, also referred to as opinion mining, is a computational technique utilized to detect sentiments and personal information conveyed within a text. Its objective is to extract opinions, emotions, attitudes, and personal information from textual data and classify them into positive, negative, or neutral categories. However, one of the main challenges in sentiment analysis lies in accurately discerning between various emotions. Textual statements can be intricate, incorporating elements of sarcasm, ambiguity, or complexity, thereby posing difficulties in accurately identifying the underlying emotion being expressed.

The Arabic language presents additional challenges in sentiment analysis due to the limited availability of language resources and tools compared to some widely spoken foreign languages. This scarcity of resources hinders the development of robust sentiment analysis models specifically designed for Arabic script. Moreover, there is limited research focusing on sentiment analysis in the Jordanian dialect, which highlights the significance of exploring sentiment analysis in this specific context. By studying sentiment analysis in the Jordanian dialect, valuable insights into regional patterns of feelings and expressions can be gained. In this study,

the impact of using the Memetic algorithm on the performance of classifiers will be investigated, aiming to achieve different results and compare them with traditional classification approaches. Support Vector Machine, Multinomial Naïve Bayes, and Logistic Regression were chosen as the classifiers based on their proven success in SA in previous studies.

The memetic algorithm strategy used in this study included using Pearson correlation as a local search strategy and integrating the genetic algorithm as a global search strategy. By integrating the memetic algorithm into the classification process, it was applied to the Arab-Jordanian Public Tweets (AJGT) dataset. This led to positive results across all classifiers and performance measures. Specifically, the accuracy scores achieved by each classifier using the memetic algorithm were as follows: (SVM): 82.22% , (MNB): 85.61% and (LR): 84.56%.

These results indicate the effectiveness of the MEM algorithm in enhancing the performance of sentiment analysis classifiers in a given context .

Keywords : sentiment analysis, support vector machine, Memetic algorithm, local search, global search, feature selection, fitness function, genetic algorithm, Pearson correlation.

تحليل المشاعر في اللهجة الاردنية باستخدام خوارزمية الذاكرة ومصنف آلات ناقلات الدعم

إعداد

أحمد مفلح سالم الحدب

إشراف

الدكتور غيث جرادات

الملخص

تحليل المشاعر، المعروف أيضًا باسم التنقيب عن الرأي، هو أسلوب حسابي يستخدم لاكتشاف المشاعر والمعلومات الشخصية المعبّر عنها في النصوص. يهدف هذا الأسلوب إلى استخلاص الآراء والمشاعر والمواقف والمعلومات الشخصية من البيانات النصية وتصنيفها إلى فئات إيجابية أو سلبية أو محايدة. ومع ذلك، يواجه تحليل المشاعر تحديات رئيسية في التمييز الدقيق بين المشاعر المختلفة. قد تكون العبارات النصية معقدة وتحتوي على عناصر من السخرية أو الغموض أو التعقيد، مما يصعّب تحديد المشاعر الأساسية بدقة.

تشكل اللغة العربية تحديات إضافية في تحليل المشاعر بسبب قلة الموارد والأدوات اللغوية مقارنةً ببعض اللغات الأجنبية المستخدمة على نطاق واسع. تقيّد ندرة الموارد تطوير نماذج قوية لتحليل المشاعر مصممة خصيصًا للنصوص العربية. بالإضافة إلى ذلك، يوجد بحث محدود يركز على تحليل المشاعر في اللهجة الأردنية، مما يبرز أهمية استكشاف تحليل المشاعر في هذا السياق المحدد. من خلال دراسة تحليل المشاعر في اللهجة الأردنية، يمكن الحصول على رؤى قيمة حول الأنماط الإقليمية للمشاعر والتعبيرات. في هذه الدراسة، تم دراسة تأثير استخدام خوارزمية الذاكرة على أداء المصنفات بهدف تحقيق نتائج مختلفة ومقارنتها مع أساليب التصنيف التقليدية. تم

اختيار آلة دعم المتجهات ، ومتعدد الحدود، والانحدار اللوجستي كمصنفات بناءً على نجاحهم المؤكد في تحليل المشاعر في الدراسات السابقة.

استراتيجية خوارزمية الذاكرة المستخدمة في هذه الدراسة تشمل استخدام ارتباط بيرسون كاستراتيجية بحث محلية ودمج الخوارزمية الجينية كاستراتيجية بحث عالمية. من خلال دمج خوارزمية الذاكرة في عملية التصنيف، تم تطبيقها على مجموعة بيانات التغريدات العربية الأردنية (AJGT) . أدى ذلك إلى نتائج إيجابية على جميع المصنفات ومقاييس الأداء. بالتحديد، حققت المصنفات باستخدام خوارزمية الذاكرة درجات دقة كالتالي:

.٪LR: 84.56 ،٪MNB: 85.61 ،٪SVM: 82.22

تشير هذه النتائج إلى فعالية خوارزمية الذاكرة في تحسين أداء مصنفات تحليل المشاعر في سياق محدد.

الكلمات الرئيسية: تحليل المشاعر، آلة الدعم القطاعي، خوارزمية الذاكرة، البحث المحلي، البحث العالمي، اختيار الميزة، وظيفة اللياقة، الخوارزمية الجينية، ارتباط بيرسون