استخراج الخصائص لكشف التغاريد غير المرغوب فيها باللغة العربية باستخدام خوارزمية متجه الدعم الآلي

إعداد

عمر اياد اسماعيل

اشراف

الاستاذ الدكتور رياض الشلبي

الملخص

أدى الانتشار الواسع للشبكات الاجتماعية في الآونة الأخيرة إلى النظر في مخاطر استخدام الإنترنت من الذين يهدفون إلى توزيع محتويات ضارة مثل الرسائل غير مرغوب بها. تلك الرسائل غير المرغوب بها تتم من خلال إنشاء حسابات وهمية على خدمات الشبكات الاجتماعية المختلفة لإرسال رسائل ضارة (SPAMS) إلى المستخدمين الشرعيين. قد يستخدم مجرمي الإنترنت أيضا ملفات شخصية مسروقة للمستخدمين الشرعيين لإرسال رسائلهم العشوائية والرسائل الضارة.

تركز معظم البحوث الموجودة في الكشف عن الرسائل غير المرغوب فيها على تويتر على تعريف مرسلي الرسائل غير المرغوب فيها لحجب الحسابات. هذه الالية لا تعمل بشكل جيد مع مرسلي الرسائل غير المرغوب فيها الذين قد ينشرون محتوى غير مرغوب فيه بشكل منتظم لكي يعملوا كمستخدمين قانونيين. قد تؤدي هذه الآلية إلى إلحاق الضرر بالمستخدم القانوني الذي يمنح الإذن لطرف ثالث يستخدم اسم المستخدم لإرسال مشاركات غير مرغوب فيها.

ويهدف هذا البحث إلى تقديم مراجعة شاملة لوسائل التواصل الاجتماعي، والرسائل القصيرة وتقنياتها، والتقنيات السابقة التي تستخدم للكشف عن الرسائل الضارة في تويتر، تصفية مجموعات التغاريد الضارة على أساس الاستدلال ، وتصنيف مجموعات التغاريد التي هي غير مرغوبة. وبناء خوارزمية التنبؤ من أجل

تصنيف مجموعة التغاريد التي تم التغريد إلى ضارة أو غير ضارة. وتظهر النتيجة نسبة عالية من الدقة تقدر ب (95)% عندما قمنا بتقييم الأداة المقترحة على أساس 1000 تويت (كمجموعة بيانات لبناء الخوارزمية)، و 100، 500، 1000، 2000 تغريدة (كمجموعة بيانات للاختبار) حيث تمت مقارنة مع خوارزميات اخرى , التقنية المقترحة تفوقت على (Naïve Bayes و Random Forest). متوسط دقة للنظام المقترح هو (0.9525) متبوعًا بتقنية Random Forest ، و Naïve Bayes بـ (0.70) ، (0.50) على التوالي

Feature Extraction to Detect Arabic Tweets Spam Using Support Vector Machine Algorithm

Prepared by

Omar Ayad Ismail

Supervised by

Prof. Riyad Alshalabi

Abstract

The wide publicity of Social Networks lately has brought the consideration of *Cybercriminals* which aim to distribute spams and harm contents. Those spammer create FAKE accounts over different Social Networks services to send their spams to legitimate users. *Cybercriminals* may also use stolen profiles of legitimate users to send their spams and malicious messages.

Most existing researches for detecting spams on Twitter focus on defining of spammers for account blocking. This mechanism is not work well with spammers who may regularly post non-spam content in order to act as legal users. This mechanism may hurt a legal user who grant permission to a 3rd party that uses his username to post spam tweets. This research aims to build an automated tool to detect the spams in the Twitter and control the spammer profiles using SVM and feature selection algorithms. The result shows high percent of accuracy (95%) when we evaluated the proposed tool based on 1000 tweets (as training data set), and 100, 500, 1000, 2000 tweets (as testing data set). has been compared with algorithms the proposed technique outperforms other existing detection techniques including (Naïve Bayes and Random Forest). The average of accuracy of the

proposed technique is (0.9525) followed by Random Forest technique, and Naïve Bayes with (0.70), (0.50) respectively