Bot spammer detection in twitter using tweet similarity and time interval entropy

Rizal Setya Perdana, author

Abstrak

The popularity of Twitter has attracted spammers to disseminate large amount of spam messages. Preliminary studies had shown that most spam messages were produced automatically by bot. Therefore bot spammer detection can reduce the number of spam messages in Twitter significantly. However, to the best of our knowledge, few researches have focused in detecting Twitter bot spam-mer. Thus, this paper proposes a novel approach to differentiate between bot spammer and legitimate user accounts using time interval entropy and tweet similarity. Timestamp collections are utilized to calculate the time interval entropy of each user. Uni-gram matching-based similarity will be used to calculate tweet similarity. Datasets are crawled from Twitter containing both normal and spammer accounts. Experimental results showed that legitimate user may exhibit regular behavior in posting tweet as bot spammer. Several legitimate users are also detected to post similar tweets. Therefore it is less optimal to detect bot spammer using one of those features only. However, combination of both features gives better classification result. Precision, recall, and f-measure of the proposed method reached 85.71%, 94.74% and 90% respectively. It outperforms precision, recall, and f-measure of method which only uses either time interval entropy or tweet similarity. <hr>Ketenaran Twitter mengundang spammer untuk menggunakannya dalam penyebarluasan pesan spam. Penelitian terdahulu menunjukkan bahwa kebanyakan pesan spam dihasilkan secara otomatis oleh bot. Deteksi bot spammer akan dapat mengurangi jumlah pesan spam pada Twitter secara signifikan. Akan tetapi, sejauh yang penulis ketahui, masih sedikit penelitian yang fokus dalam deteksi bot spammer pada Twitter. Sehingga, paper ini mengusulkan pendekatan baru untuk membedakan antara bot spammer dan pengguna sah menggunakan time interval entropy dan kemiripan antar tweet. Kum-pulan timestamp digunakan untuk menghitung time interval entropy dari tiap akun pengguna. Uni-gram matching-based similarity akan digunakan untuk menghitung kemiripan antar tweet. Dataset diambil dari Twitter yang terdiri atas kumpulan akun normal dan akun yang terindikasi sebagai bot spammer. Hasil percobaan menunjukkan beberapa pengguna sah Twitter juga memiliki kebiasaan yang teratur dalam menghasilkan tweet sebagaimana bot spammer. Beberapa pengguna sah juga ter-deteksi menghasilkan tweet yang mirip. Oleh karena itu, deteksi bot spammer menggunakan satu fitur saja akan kurang optimal. Akan tetapi, kombinasi atas kedua fitur tersebut memberikan hasil klasifi-kasi yang lebih baik. Presisi, recall, dan f-measure dari metode yang diusulkan mencapai 85.71%, 94.74% dan 90%. Nilai ini melampaui presisi, recall, dan f-measure dari metode yang hanya meng-gunakan baik time interval entropy maupun kemiripan antar tweet saja.