

INTISARI

MODEL DETEKSI AKTOR DOMINAN PENYEBAR UJARAN KEBENCIAN MENGUNAKAN METODE WEIGHTED GRAPH COLLECTIVE INFLUENCE PADA TWITTER

Oleh

Galih Hendro Martono

16/405308/SPA/00565

Deteksi aktor dominan pada Twitter memiliki peran penting dalam menjelaskan distribusi informasi dari setiap aktor. Identifikasi aktor penyebaran ujaran kebencian bermanfaat untuk mengetahui aktor dominan dalam distribusi informasi serta memutuskan informasi agar tidak menyebar lebih luas. Penelitian ini mengusulkan suatu metode untuk mengidentifikasi aktor dominan dalam penyebaran ujaran kebencian berdasarkan keterhubungannya dengan aktor lain pada Twitter. Keterhubungan aktor pada Twitter dapat dilihat berdasarkan relasi dan interaksi antara pengguna Twitter. Metode ini selanjutnya disebut sebagai metode WGCI.

Metode WGCI dikembangkan berdasarkan metode CI. Berbeda dari metode CI, metode WGCI mempertimbangkan besarnya pengaruh dari setiap aktor. Besarnya pengaruh ini direpresentasikan sebagai *weight*. *Weight* diperoleh dari intensitas komunikasi pengguna twitter seperti *follower*, *mention*, dan *retweet*. Penggunaan *weight* pada metode yang diusulkan karena setiap aktor memiliki kontribusi yang berbeda dalam penyebaran ujaran kebencian pada Twitter.

Tahapan pada penelitian ini terdiri dari tiga tahapan yaitu tahap persiapan data *graph*, tahap perhitungan aktor dominan dengan metode WGCI, dan tahap evaluasi. Tahap pertama yaitu persiapan data *graph* yang diawali dengan proses pengumpulan data. Data yang digunakan dalam penelitian ini merupakan data *tweet*. Data tersebut merupakan hasil ekstraksi data pada Twitter dengan menggunakan beberapa kamus kata ujaran kebencian untuk kasus Bahasa Indonesia. Data tersebut selanjutnya di olah dan di transformasi ke dalam bentuk data *graph*. Tahap selanjutnya adalah perhitungan aktor dominan dengan metode WGCI. Data yang digunakan pada tahapan ini adalah data *graph* yang merepresentasikan hubungan setiap aktor beserta nilai pengaruhnya.

Tahap akhir yaitu evaluasi kinerja metode yang diusulkan dengan menggunakan dataset ujaran kebencian. Terdapat dua evaluasi yang digunakan yaitu evaluasi pertama dengan menggunakan *Precision@k*, *AP*, dan *Jaccard Similarity Coefficient*. Data yang digunakan sebagai *ground-truth* adalah *list* ranking hasil dari simulasi dengan SIR Model. Evaluasi kedua menggunakan *Monotonicity M(R) Index* untuk melihat ketepatan dalam perankingan. Hasil evaluasi metode WGCI dengan *Precision@k* sebesar sebesar 0,3-0,56, *AP* sebesar 0,472-0,488, dan *Jaccard Similarity Coefficient* sebesar 0,312-0,324. Sedangkan untuk pengukuran metode WGCI dengan *Monotonicity M(R) Index* diperoleh nilai sebesar 0,99.

Kata kunci: deteksi aktor dominan, *centrality measure*, *Collective Influence*, ujaran kebencian.

ABSTRACT

DETECTION MODEL OF DOMINANT ACTOR IN THE SPREADING OF HATE SPEECH USING THE WEIGHTED GRAPH COLLECTIVE INFLUENCE METHOD IN TWITTER

By

Galih Hendro Martono

16/405308/SPA/00565

Dominant actor detection on Twitter is vital in explaining the distribution of information from each actor. Identifying the actors in spreading hate speech helps know the dominant actors in distributing information and stopping information from spreading more widely. This study proposes a method to identify dominant actors in spreading hate speech based on their relationship with other actors on Twitter. The connectedness of actors on Twitter can be seen based on the relationships and interactions between Twitter users. This method is, from now on, referred to as the WGCi method.

The WGCi method was developed based on the CI method. Unlike the CI method, the WGCi method considers the magnitude of the influence of each actor. The magnitude of this influence is represented as weight. Weight is obtained from the intensity of Twitter user communication, such as *follower*, *mention*, dan *retweet*. The use of weight in the proposed method is because each actor contributes differently to spreading hate speech on Twitter.

This study consisted of three stages: the data graph preparation stage, the dominant actor calculation stage using the WGCi method, and the evaluation stage. The first stage is preparing the data graph, which begins with data collection. The data used in this research is tweet data. The data results from data extraction on Twitter using several hate speech word dictionaries for the Indonesian language case. The data is then processed and transformed into a data graph. The next stage is the calculation of dominant actors using the WGCi method. The data used at this stage is a data graph representing each actor's relationship and influence value.

The final stage is evaluating the performance of the proposed method using the hate speech dataset. There are two evaluations used, namely the first evaluation using Precision@k, AP, and Jaccard Similarity Coefficient. The data used as ground-truth is a ranking list resulting from a simulation with the SIR Model. The second evaluation uses the Monotonicity M(R) Index to see the accuracy in ranking. The evaluation results for the WGCi method with a Precision@k of 0.3-0.56, an AP of 0.472-0.488, and a Jaccard Similarity Coefficient of 0.312-0.324. Meanwhile, for the measurement of the WGCi method with the Monotonicity M(R) Index, a value of 0.99 was obtained.

Keywords: detection of dominant actors, *centrality measure*, *Collective Influence*, hate speech.