

Deep Embedded Clustering untuk Pendeteksian Topik Tweet Berita Berbahasa Indonesia = Deep Embedded Clustering for Topic Detection on Indonesian News Tweet

Rilo Chandra Pradana, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509748&lokasi=lokal>

Abstrak

Pendeteksian topik adalah teknik untuk memperoleh topik-topik yang dikandung oleh suatu data tekstual. Salah satu metode untuk pendeteksian topik yaitu dengan menggunakan clustering. Namun, secara umum metode clustering tidak menghasilkan cluster yang efektif bila dilakukan pada data yang berdimensi tinggi. Sehingga untuk memperoleh cluster yang efektif perlu dilakukan reduksi dimensi pada data sebelum dilakukan clustering pada ruang fitur yang berdimensi lebih rendah. Pada penelitian ini, digunakan suatu metode bernama Deep Embedded Clustering (DEC) untuk melakukan pendeteksian topik. Metode DEC bekerja untuk mengoptimasi ruang fitur dan cluster secara simultan. Metode DEC terdiri dari dua tahap. Tahap pertama terdiri dari pembelajaran autoencoder untuk memperoleh bobot dari encoder yang digunakan untuk mereduksi dimensi data dan k-means clustering untuk memperoleh centroid awal. Tahap kedua terdiri dari penghitungan soft assignment, penentuan distribusi bantuan untuk menggambarkan cluster di ruang data, dan dilanjutkan dengan backpropagation untuk memperbarui bobot encoder dan centroid. Dalam penelitian ini, dibangun dua macam model DEC yaitu DEC standar dan DEC without backpropagation. DEC without backpropagation adalah DEC yang menghilangkan proses backpropagation pada tahap kedua. Setiap model DEC pada penelitian ini akan menghasilkan topik-topik. Hasil tersebut dievaluasi dengan menggunakan coherence. Dari penelitian ini dapat dilihat bahwa model DEC without backpropagation lebih baik daripada DEC standar bila dilihat dari waktu komputasi dengan perbedaan coherence antara keduanya yang tidak terlalu jauh.

Topic detection is a technique for obtaining the topics that are contained in a textual data. One of the methods for topic detection is clustering. However, generally clustering does not produce an effective cluster when it is done by using data with high dimension. Therefore, to get an effective cluster, dimensionality reduction is needed before clustering in the lower dimensional feature space. In this research we use DEC method for topic detection. DEC method is used to optimize the feature space and cluster simultaneously. DEC is divided into two stages. The first stage consists of autoencoder learning that obtains the weights of the encoder that used for dimension reduction and k-means clustering to get the initial centroid. The second stage consists of the soft assignment calculation, computing the auxiliary distribution that represents the cluster in the data space, and backpropagation to update the encoder weights and the centroid. In this research, two DEC models are built, namely the standard DEC and DEC without backpropagation. DEC without backpropagation is the DEC which eliminate the backpropagation process in the second stage. Every DEC models will produce topics. The results are evaluated using the coherence measure. From this research, it can be seen that DEC without backpropagation is better than standard DEC in terms of computation time with a slight difference in coherence measure.