

Metode Improved Deep Embedded Clustering untuk Pendeteksian Topik = Improved Deep Embedded Clustering Method for Topic Detection

Naufal Farhan, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509854&lokasi=lokal>

Abstrak

Pendeteksian topik adalah suatu proses yang digunakan untuk menganalisis kata-kata pada suatu koleksi data tekstual untuk menentukan topik-topik yang ada pada koleksi tersebut. Salah satu metode standar yang digunakan untuk pendeteksian topik adalah metode clustering. Deep embedded clustering (DEC) adalah algoritma clustering dengan pendekatan deep learning yang menyatukan pembelajaran fitur dan clustering menjadi satu kerangka kerja sehingga dapat menghasilkan kinerja yang lebih baik. Namun metode DEC memiliki kelemahan, yaitu terjadinya penyimpangan ruang embedded ketika melakukan pembelajaran yang didapat ketika membuang decoder. Kelemahan tersebut diatasi dengan tidak membuang decoder, sehingga diperoleh metode yang lebih baik lagi yaitu Improved Deep Embedded Clustering (IDEC). Proses mempertahankan decoder disebut sebagai pelestarian struktur lokal. Pada penelitian ini, metode IDEC diadaptasi untuk masalah pendeteksian topik data tekstual berbahasa Indonesia. Selanjutnya kinerja metode IDEC dibandingkan dengan metode penelitian lain yang menggunakan DEC untuk masalah pendeteksian topik yaitu dengan cara membandingkan nilai dari coherence. Nilai coherence yang dihasilkan menunjukkan bahwa metode DEC lebih cocok jika dibandingkan dengan metode IDEC untuk permasalahan pendeteksian topik. Hal tersebut terjadi karena bagian decoder pada metode IDEC diperbarui sehingga parameter decoder sudah tidak sesuai untuk mengembalikan data ke dimensi semula. Sedangkan pada metode DEC bagian decoder dibuang sehingga parameter tidak diperbarui.

Topic detection is a process that is used to analyze words in a textual data collection to determine the topics within that collection. One of this standard topic detection method is clustering method. Deep embedded clustering (DEC) is a clustering algorithm with a deep learning approach that combines feature learning and clustering into one framework to obtain a better performance. However, DEC method has a weakness namely the distortion of embedded space that is caused by removing the decoder during the learning process. This weakness can be overcome by preserving the decoder, hence a better method is acquired, namely Improved Deep Embedded Clustering (IDEC). The process of preserving the decoder is called local structure preservation. In this research we adapt IDEC method for topic detection problem in Indonesian textual dataset. Furthermore, we compare the performance of IDEC method and other research using DEC by comparing the coherence value. The acquired coherence value shows that DEC method is more suitable compared to IDEC method for topic detection problems. This happens because of the decoder part in IDEC method is updated, so that the decoder parameters are no longer suitable to return the data into the original dimension. While in the DEC method the decoder was removed, therefore the parameters are not updated.