Rizki Rafi fidityo

A11.2020.12822

4703

Rangkuman

***Word embedding*** adalah teknik dalam pemrosesan bahasa alami (NLP) yang mengubah kata-kata ke dalam representasi vektor numerik dalam ruang berdimensi tinggi. Representasi vektor ini memungkinkan komputer untuk memahami hubungan makna antar kata berdasarkan lokasi relatif vektor-vektor tersebut. Teknik word embedding telah membawa terobosan dalam pemrosesan teks dan digunakan dalam berbagai aplikasi NLP seperti analisis sentimen, terjemahan mesin, pengenalan entitas bernama, dan lainnya. Ada berbagai model word embedding yang tersedia, termasuk Word2Vec, GloVe, dan FastText, yang dapat digunakan untuk memperbaiki kinerja model NLP. Word embedding juga memungkinkan penggunaan representasi kata yang sudah ada (pre-trained) atau pengembangan representasi khusus sesuai dengan konteks tugas tertentu. Teknik ini memiliki peran penting dalam pemrosesan bahasa alami dan terus menjadi subjek penelitian intensif untuk meningkatkan kualitas pemahaman komputer terhadap bahasa manusia. Pengumpulan dan pemrosesan dataset yang sesuai untuk berbagai tugas NLP yang akan dievaluasi.Eksperimen dengan berbagai model word embedding yang ada, termasuk model pre-trained dan model custom yang dikembangkan. Evaluasi kinerja model NLP yang menggunakan word embedding berdasarkan metrik yang relevan, seperti akurasi, F1-score, atau metrik tugas khusus. Analisis hasil eksperimen dan kesimpulan untuk menentukan model word embedding terbaik untuk berbagai tugas NLP dan cara memaksimalkan kinerjanya.

**Search engine concept,** atau konsep mesin pencari, merujuk pada prinsip-prinsip dasar dan komponen-komponen sistem yang dirancang untuk mengambil informasi dari kumpulan data besar, seperti yang ada di internet. Beberapa konsep kunci dalam mesin pencari meliputi:

*Pencarian dan Pengindeksan*: Mesin pencari menggunakan perangkat penyelusuran web (web crawlers) untuk mengunjungi dan mengindeks halaman web. Penyelusuran dilakukan dengan mengikuti tautan-tautan untuk menemukan dan menyimpan konten web dalam indeks terstruktur guna pengambilan nantinya.*Pengambilan Informasi:* Mesin pencari menggunakan algoritma untuk mengambil dokumen dokumen yang relevan dari indeks mereka ketika pengguna memasukkan pertanyaan. Faktor-faktor seperti relevansi kata kunci, kualitas konten, dan niat pengguna dipertimbangkan. *Algoritma Peringkat:* Mesin pencari menggunakan algoritma peringkat untuk menentukan urutan tampilan hasil pencarian. Algoritma-algoritma populer meliputi PageRank dan model berbasis pembelajaran mesin. *Antarmuka Pengguna:* Mesin pencari menyediakan antarmuka yang ramah pengguna untuk memasukkan pertanyaan dan menjelajahi hasil pencarian. Fitur-fitur seperti prediksi otomatis dan filter meningkatkan pengalaman pengguna. *Pencarian Semantik:* Mesin pencari modern berusaha memahami niat dan konteks pengguna. Mereka menggunakan pemrosesan bahasa alami dan analisis semantik untuk memberikan hasil yang lebih akurat. *Pencarian Web vs. Pencarian Enterprise:* Mesin pencari dapat disesuaikan untuk pencarian web atau pencarian enterprise. Mesin pencari enterprise dirancang khusus untuk organisasi tertentu guna mencari dokumen dan data internal. *Relevansi dan Kualitas:* Mesin pencari berusaha memberikan hasil yang paling relevan dan berkualitas tinggi. Faktor-faktor seperti otoritas halaman, kesegaran konten, dan interaksi pengguna dipertimbangkan. *Etika Penyelusuran Web:* Perangkat penyelusuran web harus mematuhi panduan etika, seperti menghormati file robots.txt dan tidak memberatkan situs web dengan permintaan berlebihan. *Personalisasi:* Banyak mesin pencari menawarkan hasil yang dipersonalisasi berdasarkan riwayat pencarian pengguna, lokasi, dan preferensi. *Model Pengambilan Informasi:* Mesin pencari menggunakan berbagai model, termasuk model ruang vektor, model Boolean, dan model probabilistik, untuk mengambil dokumen yang relevan berdasarkan pertanyaan pengguna. *Skalabilitas dan Kecepatan:* Mesin pencari harus mampu mengatasi jumlah konten web yang terus bertambah dan memberikan hasil dengan cepat. Secara ringkas, konsep mesin pencari adalah prinsip-prinsip dasar dan teknologi yang mendasari fungsi mesin pencari, memungkinkan pengguna untuk mengakses dan mengambil informasi dari berbagai repositori digital dengan efisien