

■ 논문요약문5

논문제목	De-noising documents with a novelty detection method utilizing class vectors in customer-voice classification
게재정보	Intelligent Data Analysis, 22(4),2018
개요	<p>The classification of customer-voice data is an important matter in real business since it is necessary for customer-voice data to be delivered to relevant departments and responsible individuals. Additionally, customer-voice data typically includes several novel words, such as typo's, informal terms, or exceedingly general words to discriminate between categories of customer-voice data. Furthermore, noisy data often has a negative effect on the classification task.</p> <p>In this study, advanced novelty detection method is proposed to utilize class vector that possessed high cosine similarity with words to effectively discriminate between classes. The class vector is considered as the centroid or the mean of each word vector distribution as derived from the neural embedding model, and the novelty score of each word is calculated and novel words are effectively detected. Each novelty score is calculated by improvements of GMM and KMC methods utilizing a class vector. The experiments verify the propriety of the proposed method with qualitative observations, and the application of the proposed method with quantitative experiments verifies the representational effectiveness and classification performance of customer-voice data.</p>
연구결과	<p>The experiment results indicate that the performance of a classification of customer-voice data improved with the application of the newly proposed novelty detection method in this study</p>
활용분야 및 기대효과	<p>Advanced novelty detection method is proposed to utilize class vector that possessed high cosine similarity with words to effectively discriminate between classes.</p>