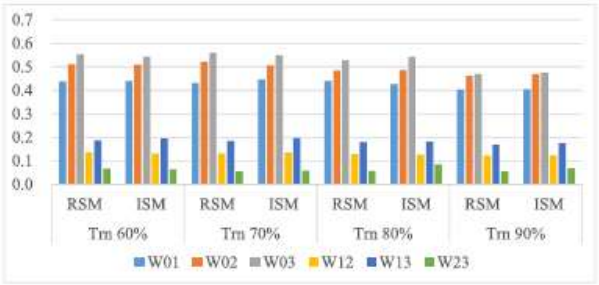
■ 연구논문 요약문

논문제목	Box-office forecasting based on sentiments of movie reviews and Independent subspace method
게재정보	Information Sciences
개요	Box-office forecasting is a challenging but important task for movie distributors in their decision making process. Many previous studies have tried to determine a way to accu- rately predict the box-office, but the results reported have not been satisfactory for two main reasons: (1) lack of variable diversity and (2) simplicity of forecasting algorithms. Although the importance of word-of-mouth (WOM) has consistently emphasized in past studies, only summarized information, such as volume or valence of user ratings is com- monly used. In forecasting algorithms, multiple linear regression is the most popular al- gorithm because it generates not only predicted values but also variable significances. In this study, new box-office forecasting models are presented to enhance the forecasting ac- curacy by utilizing review sentiments and employing non-linear machine learning algo- rithms. Viewer sentiments from review texts are used as input variables in addition to conventional predictors, whereas three machine learning-based algorithms, i.e., classifica- tion and regression tree (CART), artificial neural network (ANN), and support vector regres- sion (SVR), are employed to capture non-linear relationship between the box-office and its predictors. In order to provide variable importance for machine learning-based forecasting algorithms, an independent subspace method (ISM) is applied. Forecasting results from six different forecasting periods show that the presented methods can make accurate and ro- bust forecasts.
연구결과	
활용분야 및 기대효과	In this study, we developed box-office forecasting models based on machine learning-based algorithms with ISM. Three different factors, motion picture, external, and audience were considered as the source of input variables. Particularly for the audience factors, user sentiments in a review text, not just a simple rating, were also considered by constructing a domain- specific sentiment dictionary based on movie review data. In addition, input variables in these factors were assessed for each forecasting model using ISM, which was proved to be robust to a correlated structure. For a prediction task, one linear model, the MLR, and three machine-learning based algorithms, the CART, ANN, and SVR, were employed. The forecasting re- sults showed that machine learning-based algorithms could make more accurate forecasts when available data is insufficient or for long-term forecasting.