URN etd-0407109-183725-2081
Author zih-yang Dai
Author's Email Address No Public.
Statistics This thesis had been viewed 65 times. Download 0 times.
Department
Year 2007
Semester 1
Degree Master
Type of Document Master's Thesis
Language zh-TW.Big5 Chinese
Title Predicting Relapse of Breast Cancer with Classification Analysis
Date of Defense 0000-00-00
Page Count 61
Keyword
Survival Analysis
Inductin Decision Tree
Cox Regression Model
Classification Analysis.

Abstract
Machine learning techniques have recently received considerable attention, especially when used for the construction of prediction models from data. However, their applications to survival analysis are very rare, because of the difficulty to handle censored data. The censored cases are often discarded when applying classification techniques to survival analysis. Therefore, Zupan et al. proposed a method that enables the use of classification methods for survival analysis. To appropriately consider the follow-up time and censoring, Zupan's method estimates their probability of event and assigns them appropriate outcome probabilities for the patients for which the event did not occur and have short follow-up times.
To investigate the utility of the censored data, we applied the C4.5 learning alogirhtm, a k-NN method, and a cost learning strategy to investigate a particular problem of building prognostic models for breast cancer recurrence, where the sole prediction of result of relapse event (i.e., whether there is relapse or not) is of interest. Since most classification techniques do not deal with outcome distributions, we borrowed Zupan's idea and estimated occurrence probability of replase for censored data with a k-NN method. Then, we applied a cost learning strategy to build a prognostic committee machine based on C4.5. According to our evaluation results, both the C4.5 model and the C4.5 committee machine perform better than the Cox's proportional hazards model does. Moreover, the C4.5 committee machine also improves the effectiveness of the C4.5 model. The results of this study show that by incorporating censored data handled by k-NN method we can build a better prognostic model to support prognosis of breast canser with the classification techniques.
Advisory Committee - advisor

Files No Any Full Text File.
Date of Submission 2008-08-08