

## 学位論文の要旨

専攻名	環境工学	ふりがな氏	名	おおえもとき 大江基貴	
学位論文題目	Modeling covariates effects on ROC curves with smoothing (ROC 曲線に及ぼす共変量効果の平滑化によるモデリング)				
<p>In modern medicine, biomarkers, clinical data and measured values in molecular biology analysis are routinely used for classification or prediction with respect to clinical diagnoses. Scientific assessment of such measurements or measurement methods involves assessing the accuracy of classification and prediction of outcomes of interest based on the measurements. A receiver operating characteristic (ROC) curve is used to evaluate classification and prediction accuracy based on such measurement. Recently, the curve is often applied to the fields of medicine and epidemiology, where it is used to summarize the potential performance of diagnostic testing (in broad terms, classification and diagnostic accuracy).</p> <p>It is commonly assumed that a test performance summarized by the ROC curve does not depend on covariates such as characteristics of subjects and methods of examinations. In fact, however, as argued by Begg (1987) and Kraemer (1992), many factors (covariates) may affect the performance of a medical test. Therefore, in order to summarize the diagnostic performance appropriately, it is necessary to evaluate influences of the covariates on the test.</p> <p>Various ROC regression models have been proposed to evaluate covariates effects on the diagnostic performance of medical tests. An approach for estimating the ROC regression models focused in this dissertation is referred to as “direct approach” (Pepe, 1997, 2000). Alternatively, there is an approach which is called “induced approach” (Tosteson &amp; Begg, 1988).</p> <p>In the direct approach, it may lead to wrong conclusions to use the ROC regression model based on Generalized Linear Model (GLM) in the case where we cannot assume linearity assumptions with respect to covariates effects. To deal with this problem, Rodríguez-Álvarez <i>et al.</i> (2011b) extended the ROC regression model based on GLM and suggested an inference process in the framework of Generalized Additive Model (GAM) (Hastie &amp; Tibshirani, 1984). They proposed to use a local linear kernel smoother (Wand &amp; Jones, 1995) with a bandwidth selected by Cross Validation (CV) in order to estimate smoothing functions. In the direct approach, however, the responses are empirical estimates of the ROC curve and are correlated. In such case, Diggle &amp; Hutchinson (1989) and Wang (1998) have pointed out that the bandwidth selected by CV may result in under-smoothing. We, here, propose to use a spline regression model in order to estimate the smoothing functions. Since the estimation of the smoothing functions based on the spline regression model is reduced to fitting a linear mixed model, we can estimate the smoothing parameter based on the restricted maximum likelihood (REML) framework. In this approach, even if responses are correlated, smoothing parameters can be selected</p>					

appropriately (Wang, 1998; Krivobokoba & Kauermann, 2007). We compare the fitting performance of the proposed approach with their approach by some simulations.

The above approaches mainly focus on smoothing influences of continuous covariates such as age or weight on the accuracy of a diagnosis. Discrete covariates such as sex or smoking, however, may also have a significant influence on the diagnostic performance. We further extend the approach so that we can model discrete and continuous covariates simultaneously. The formulation and inference procedure are described, and the inference performance of this approach is studied through several simulations. Furthermore, we apply our approach to the neonatal hearing impairment screening data and the diagnostic data of coronary artery disease (CAD).

For the first proposal, when we compared the proposed approach and the existing one (Rodríguez-Álvarez *et al.*, 2011b) by some simulations, we found that our approach outperformed theirs in terms of the precision of estimating, especially when the sample size was not large.

For the second proposal, the simulation study indicated that the influence of the discrete covariates was appropriately estimated, and that smoothing of the continuous covariates effects on the ROC curve gave reliable model estimates. Additionally, we found that our approach was more efficient than the approach of Rodríguez-Álvarez *et al.* (2011b).

In the application to the neonatal hearing screening data, we focused on the evaluation of the diagnostic performance of the auditory brainstem response (ABR) and considered the ROC regression model based on our approach, which included discrete covariates (sex, left/right of ears) and a continuous covariate (Transient Evoked OtoAcoustic Emission, TEOAE) at the same time. As a result of the analysis, it was suggested that sex and TEOAE influenced on the diagnosis performance of ABR while ears had little influence.

In the analysis of the diagnostic data of CAD, the diagnostic performances of three indices (apoB48, HbA1c and adiponectin) to diagnose CAD are compared, considering risk factors (age, sex) of CAD as covariates. As a result, although the overall diagnostic performance of apoB48 was the best of these, adiponectin might be better than the other depending on sex and age.

【774 語】

(注) 和文 2,000 字又は英文 800 語以内

続紙 有  無

### 学位論文審査結果の要旨

専攻	環境工学 専攻	氏名	大江 基貴
論文題目	Modeling covariates effects on ROC curves with smoothing		
主査	中島 誠		
審査委員	古家 賢一		
審査委員	西野 浩明		
審査委員	越智 義道		
審査委員			
審査結果の要旨 (1000字以内)			
<p>本研究は、連続的な値を取る診断変量(生理学的な特性値等)を用いて疾病の診断を行う際に、その診断変量の診断能力の検討や閾値の決定などに使われる ROC(Receiver Operating Characteristic:受動者動作特性)曲線の導出に関わる研究である。一般に、ROC は診断変数に関連する他の要因(共変量)の値に基づいて異なる様相を示すと考えることが自然である。</p> <p>そのような共変量効果を反映した ROC 曲線の構成については、診断変数と共変量による回帰構造表現から間接的に構築する方法が Tosteson &amp; Begg(1988)他によって提案されている。一方で、ROC 曲線に対して関数回帰の形態で直接的に共変量効果を表現する方法が Pepe(1997)によって提案され、その後いくつかの拡張が提案されてきている。本論文は、後者の立場をとり、その共変量効果の構造表現については、Pepe(1997)他の一連の研究のような線形性を前提とする硬い数理モデルではなく、一般化加法モデルを基礎として、平滑化によってデータ適応的に柔軟な効果の確認が可能な分析モデルを提案し、その推測手順を示している。同様の研究として Rodríguez-Álvarez ら(2011)の研究があるが、彼らは局所線形カーネル平滑化法を行い、モデルのチューニングにはクロスバリデーションを利用している。この論文ではスプラインを用い、チューニングを含めたモデル推定の方法として制限付き最尤法を提案している。このために、先行研究では実現不能であった、離散量・連続量が含まれる場合での一貫した推測の枠組みが可能となり、モデルの推測性能も向上している。実データへの適用事例においても、有用な新たな知見を導いている。この点から、本研究の新規性、有用性が確認できる。また、モデル推測性能の向上については、シミュレーションによる検討に加えて、理論的な根拠を提示しており、得られた結果の信頼性も高い。</p> <p>論文審査会や公聴会における質問に対しても明確かつ的確に回答が行なわれた。また、予備審査、本審査で指摘された論文の表現上の明確化についても精査が行われ、適切な配慮と修正が行われており、論文の完成度は十分である。</p> <p>以上の審査結果から、本論文は博士(工学)の学位に値するものと審査委員会全員一致して判定した。</p>			