

EFFECTIVE USE OF CENSORED DATA FOR NPC RECURRENCE PREDICTION

ORAYA INTEM 4936227 EGBE/M

M.Eng. (BIOMEDICAL ENGINEERING)

THESIS ADVISORY COMMITTEE: PANRASEE RITTHIPRAVAT, D.Eng.,
THONGCHAI BHONGMAKAPAT, M.D.

ABSTRACT

This thesis aims to study various censored data techniques for effective prediction of nasopharyngeal carcinoma (NPC) recurrence. Clinical data and time to recurrence of NPC patients were collected from Ramathibodi Hospital, Thailand. Recurrence factors were then selected by univariate and multivariate analysis. The results showed that only 9 factors related to the NPC recurrence. They were N stage, smoking, alcohol, family history of cancer, neck fibrosis, IgG, IgA, radiation dose, and KPS. These factors were used in predictive model development. In the study, three ANN based censored data techniques are mainly investigated. They are Street, PLANN, and our proposed technique. The results showed that our proposed technique provided the highest predictive performances compared with the other two techniques and Cox regression model. All ANN based techniques outperformed the Cox model. Hosmer-Lemeshow goodness-of-fit test was then applied. The results showed that the chi-square statistic for all models was less than 15.51. This means that every model fitted well.

Survival curves for each predictive model were then generated and compared. The results showed that the curve of the Cox model was obviously different from the others. This is confirmed by the log-rank test in which only the Cox model is significantly different from the Kaplan-Meier model. With the proposed technique, four main problems existing in the previous censored data techniques can be handled. The problems include scalability, generation of a non-monotonic survival curve, specification of unknown recurrence status, and data replication problems.

KEY WORDS: SURVIVAL ANALYSIS/ NPC/ PREDICTION/ NEURAL
NETWORK

45 pages

การนำข้อมูลขาดหายไปใช้ในการทำนายการกลับมาเป็นซ้ำของผู้ป่วยโรคมะเร็งช่องคอหลังโพรง
จมูกอย่างมีประสิทธิภาพ

EFFECTIVE USE OF CENSORED DATA FOR NPC RECURRENCE PREDICTION

อรยา อินเต็ม 4936227 EGBE/M

วศ.ม. (วิศวกรรมชีวการแพทย์)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์ : ปณรสี ฤทธิประวัตติ, D.Eng., ธงชัย พงศ์มณพัฒน์, M.D.

บทคัดย่อ

งานวิจัยนี้ได้ศึกษาเทคนิคในการแก้ปัญหาข้อมูลขาดหายไปใช้ในการทำนายการกลับมาเป็นซ้ำของผู้ป่วยโรคมะเร็งช่องคอหลังโพรงจมูก ซึ่งข้อมูลทางคลินิกและเวลาถูกเก็บรวบรวมมาจากโรงพยาบาลรามารับดี ประเทศไทย โดยตัวแปรเหล่านี้จะถูกเลือกให้เหลือเพียงตัวแปรที่สำคัญด้วยวิธีการวิเคราะห์ตัวแปรเดียว (Univariate analysis) และการวิเคราะห์ตัวแปรเชิงพหุ (Multivariate analysis) ผลจากการวิเคราะห์ทำให้เหลือตัวแปรที่สำคัญ 9 ตัว ได้แก่ N stage, ประวัติการสูบบุหรี่, ประวัติการดื่มสุรา, ประวัติการเป็นมะเร็งในครอบครัว, การปรากฏของพังผืดบริเวณคอหลังการฉายรังสี, IgG, IgA, ปริมาณรังสีรักษาและ KPS งานวิจัยนี้ได้ทำการทดลองเปรียบเทียบประสิทธิภาพการพยากรณ์ของโมเดลพยากรณ์ด้วยเครือข่ายระบบประสาทเทียมสามรูปแบบ ได้แก่ โมเดลพยากรณ์ของ Street โมเดลพยากรณ์ของ PLANN และโมเดลการพยากรณ์ที่เสนอขึ้นมาใหม่ ซึ่งผลการศึกษาพบว่าโมเดลพยากรณ์ที่เสนอขึ้นมาใหม่ให้ประสิทธิภาพการทำนายที่ดีกว่าโมเดลพยากรณ์อื่นๆ รวมทั้งดีกว่าการใช้โมเดล Cox regression ซึ่งเป็นโมเดลทางสถิติ และในการทดลองยังพบว่าการพยากรณ์ของเครือข่ายระบบประสาทเทียมทั้งสามรูปแบบมีประสิทธิภาพสูงกว่าโมเดล Cox regression เมื่อทดสอบความสมบูรณ์ของโมเดลด้วยค่าไคสแควร์จาก Hosmer-Lemeshow พบว่าทุกโมเดลให้ค่าไคสแควร์น้อยกว่า 15.51 ซึ่งแปลว่าทุกโมเดลมีความเหมาะสม

ในการทดสอบประสิทธิภาพของโมเดลเชิงกลุ่มโดยการพล็อตกราฟการรอดชีวิตที่ได้จากแต่ละวิธี พบว่ากราฟของ Cox โมเดลให้การทำนายที่แตกต่างจากกราฟการรอดชีวิตของข้อมูลจริง และเมื่อทดสอบความแตกต่างด้วย Log-rank พบว่า Cox โมเดลให้การทำนายเชิงกลุ่มที่แตกต่างจาก Kaplan-Meier อย่างมีนัยสำคัญทางสถิติ จากงานวิจัยนี้พบว่าเทคนิคที่น่าเสนอสามารถแก้ไข 4 ปัญหาที่มักเกิดขึ้นในวิธีการแก้ปัญหาข้อมูลขาดหายในอดีต โดยปัญหาทั้ง 4 ได้แก่ปัญหาในการ Scalable ของโมเดล, ปัญหาเส้นโค้งการรอดชีวิตที่ไม่ลดลงตามเวลา, ปัญหาข้อมูลขาดหายและปัญหาการ Replication ของข้อมูล