Survival Analysis of Questions Posted on the iFixit Answers Forum

Lisa Oshita[a], Anthony Pileggi, Shannon Pileggi

Department of Statistics, California Polytechnic State University

[a]Frost Research Fellow, recipient of the Frost Undergraduate Student Research Award

Overview

➤ iFixit's online question and answer forum, *Answers*, features over 120,000 user-asked questions related specifically to device repair. Analysis of question response times can reveal factors that affect how quickly questions receive answers, which can lead to suggestions for how users can ask better questions to minimize response times and for how forum design can be improved.

➤ Objective Develop a Cox proportional hazards model to predict the survival probability, defined as the probability that a question remains unanswered beyond a certain time t, of questions on the forum, with the goal of identifying variables signficantly associated with response time.

Materials

The data analyzed contained 8,025 questions posted from April 8, 2017 to July 7, 2017 (date of data download). Variables in the data included: device name, device category, question title, text, tags, new user status, date and time the question was posted, and date and time the first answer was received. Fourteen variables, capturing textual and user information of each question, were derived from the data.

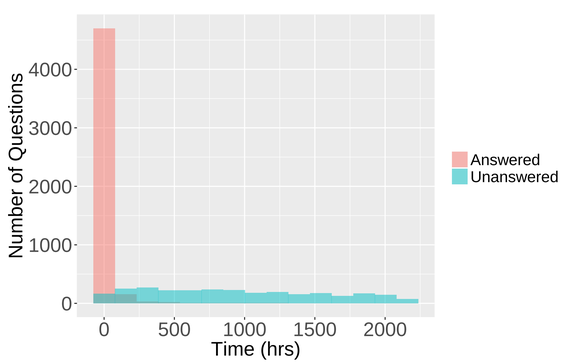

Figure 1: Distribution of response times

Model Building

➤ Univariate analysis was used to identify which variables, as well as the optimal form of continuous variables (if transformations and splines should be used), to include in the final Cox model for five fold cross-validation.

➤ In each iteration of cross-validation, the model was built on the training set and used to generate predicted hazard ratios on the corresponding test set. To assess performance, predicted hazard ratios were entered into separate Cox models as the single quantitative predictor with response times as the survival time. The resulting Nagelkerke's R^2 statistic, concordance statistic, Somers' *Dxy*, partial likelihood ratio and p-value, were averaged over each iteration and assessed as indicators of the model's performance.

➤ The final model was fit to the full data and the proportional hazards assumption was assessed.

Results

Of 7,760 questions in English, 4,951 (63.8%) received an answer by the download date. The shortest response time was 0.5 hours. The longest was 2,159 hours (90 days). Figure 1 shows the distribution of response times for all questions analyzed.

Univariate analysis determined that all categorical predictors were signficantly associated with response time. The following transformations were made to continous variables:

➤ $\sqrt{\text{average tag frequency}}$

➤ $\log(\text{average tag length} + 1)$, spline of 4 knots

➤ $\log(\text{device length} + 1)$, spline of 5 knots

➤ $\sqrt{\text{line breaks / text length}}$, spline of 3 knots

➤ $\log(\text{text length})$, spline of 5 knots

Average performance metrics for test sets, training sets, as well as the full model are in Table 1. Partial likelihood ratios and p-values indicate that the model as a whole is significantly associated with response time. However, R^2 statistics and discrimination indexes are considerably low.

Results for training, test, and the full data did not change significantly, indicating that the model was not over fit.

	HR	LR	p-value	R^2	*Dxy*	C
Training	2.03	937.39	<0.0001	0.14	0.27	0.63
Test	1.99	220.83	<0.0001	0.14	0.26	0.63
Full	2.03	1165.03	<0.0001	0.14	0.28	0.63

Table 1: Performance metrics. (HR: Hazard Ratio, LR: Partial Likelihood Ratio, C: Concordance)

Table 2 displays the coefficients for predictors in the final model fit to the full data. As a whole, the final model was significant with a partial likelihood ratio of 1265.29 (p-value <0.0001). Its R^2 statistic was 0.15, and Somers' *Dxy* was 0.27.

Assessing the proportional hazards assumption indicated that several levels of the device categorization variable, specifically Apple Product, Camera, Game Console, and Other, violated the assumption.

The following are interpretations of select hazard coefficients in Table 2. Hazard is approximately equivalent to the conditional probability that a question will receive an answer within the next moment in time, given that it has not already received an answer.

➤ The estimated hazard of receiving an answer is 154% higher (95% CI (132%, 179%)) for questions pertaining to Apple products than the hazard for questions about Android and Other phones, controlling for all other predictors.

➤ The estimated hazard of receiving an answer is 25% lower (95% CI (19%, 29%)) for questions with titles that contain at least one word considered to be frequently-used among unanswered questions, than the hazard for questions with titles that do not, controlling for all other predictors.

Discussion

The model's weak predictive accuracy, which may be explained by limitations in the data, lead to suggestions for changes in CQA design. Many users incorrectly specified the device names and tags. It is likely that these inconsistencies contributed to the final model's low predictive power. As findings indicate that questions with correctly defined tags and device names may lead to quicker response times, the CQA can provide a stricter framework for asking questions by restricting the tags or devices that users can include to a drop-down list. The CQA can also include more tips to guide users asking questions. Implementing a more structured framework for asking questions can help users create understandable and clear questions and in turn decrease response time, as well as create a set of consistent questions for improved analysis.

Conclusion

Predictors found to be significant in the final Cox regression model included: device category, if the question was posted on a weekend or a weekday, whether or not the question's title contained at least one word considered frequently-used among unanswered questions, and whether or not the question's title ends in a question mark. While overall the model was significant, its predictive performance was considerably low.

Acknowledgements

This research was supported by the Bill and Linda Frost fund of the California Polytechnic State University of San Luis Obispo. We also thank iFixit for providing access to the CQA data.