Literature Review of Online Question Forums

Lisa Oshita

1 Introduction

Community-driven online question and answer forums (CQA) are becoming widley used sources of information. These online platforms allow for anyone to ask or answer a question. It provides a place for people of all backgrounds and levels of expertise to provide input on a certain question. One popular CQA is Yahoo! Answers. This platform features questions of all topics, from politics, to food and drink, and even mathematics. On the other hand of the spectrum is Stack Overflow, the CQA that features strictly computer programming question and answers. This forum reiceves around 40 million visits every month. This platforms ultimately provide people with a place to share their knowlege, and clearly are becoming increasingly popular sources of information.

The CQA discussed in this paper is iFixit's Answers forum. iFixit is a company founded in 2003 by two engineering students of the California Polytechnic State University in San Luis Obispo. What began as a small business running out of the college dorms, is now a company that helps thousands of people fix their broken devices everyday, providing over 30,000 free repair guides with pictures and step-by-step instructions. iFixit also sells the specialized tools and parts used in their repair guides, specialized tools like an iPhone 7 batter adhesive strip or a Nexus 9 LCD screen and digitizer. The main goal of this company is to teach people how to fix their own devices, and in extending the lifetime of their own devices, they save money and cut down on electronic waste. As such, this company helps thousands of people fix their devices everyday.

Along with repair guides and tool sales, another important component of this company is their online question and answer forum called Answers. This platform features questions related to device repair, featuring over 9,000 devices and over 100,000 solutions and answers. Questions on this forum range from broken devices like a jammed zipper on a Patagonia jacket, to a shattered iPhone 6 plus screen. This forum is an integral part in teaching people how to fix their own devices. If people run into problems, they can always trust that they have a large community to turn to for help. Thus it's important, for both the individuals as well as iFixit, that askers get prompt answers. Fast response times will drive up user engagement and generate more traffic, which in turn is great for the reputation and longevity of Answers and iFixit. As such, investigation of response times would be both beneficial and informative. Analysis can reveal factors within the forum that affect how quickly questions get answers.

Suggestions for how users can ask questions to minimize answer times, as well as suggestions for how the forum design can be improved, can be derived from such analysis. However, to the best of our knowledge, analysis and prediction of answer times on forums has not been investigated by many researchers. A majority of the research focuses on assessing and predicting the question and answer quality. As such, there is need for further analysis of response times in these forums. This paper presents a survival analysis on the time until a question is answered on iFixit's Answers forum. It attempts to determine the factors that are significantly related to answer time, and create a cox proportional hazards model that can accurately predict the survival probability of a question.

2 Related Work

In regards to investigation of forum response times, (Bhat et al., 2014) developed a classification model (? is that what it's called) to analyze response times of questions posted on Stack Overflow, and found that tag-based features like the number of tags included or the number of subscribers a certain tag has, were the best predictors of answer time.

(Mamykina et al., 2011) found that the swift answer times of Stack Overflow's community is a result of the reputation system and the strict emphasis on factual and informative questions and answers, rather than discussion-based.

(Asaduzzaman et al., 2013) analyzed unanswered questions on Stack Overflow to determine common characteristics and found that questions that went unanswered shared certain characteristics in that they were too short and vague, or utilized the tagging system incorrectly.

3 Materials and Methods

The data set analyzed in this paper contained 7,760 questions posted from April 8, 2017 to July 7, 2017. Variables within the data set included information about the questions, like the title, the body text, device, date it was posted, and information about the user like whether or not the user was considered new (or a member for less than 24 hours) at the time the question was posted.

Variables created in the model:

The time until event variable used in the cox regression model is defined as the time until a question receives its first answer. If a question received an answer, this was calculated as the time from the date the question was posted to the date the first answer was received. If the question did not receive an answer, this was calculated as the time from the date the question was posted, to the date the date set was downloaded. The censoring variable needed in a cox regression model, is 0 if the question remained unanswered by the time the data set was downloaded, and 1 if the question received an answer.

Newcategory

This variable categorized the devices of the questions based off of the existing category variable as well as several other factors. The original category variable had NAs, which were a result of the asker creating a question for a device not already featured on the website, and were coded to Other. All Apple products (e.g., iPhones, iMacs, Apple watches) were given their own category. After pulling out iPhones from the original Phones category, the remaining phones were categorized as Android/Other Phone. Appliance and Household categories of the original variable were merged into Home. Car and Truck were added to the Vehicle. Lastly, any category that contained less than 2

Containanswered and contain$_{u}nansweredwerecreatedbasedoffoftheideathatcertainkeywordswouldbemo$

Title$_{q}uestionmarkandtitle_{b}eginwhThesetwovariableswerecreatedtodetermineifstatingthetitleinthefor$

Text$_{c}ontain_{p}unct.Thisisalogicalvariableindicatingtrueifthequestionstextcontainsanyendpunctuationm onsentences(sentenceswithnoendpunctuationmarks)takelongertoreceiveananswer.$

Text$_{a}ll_{l}owerThisisalogicalvariableindicatingtrueifthetextofthequestionisinalllowercase.$

Update This is a logical variable indicating true if the asker updated the question (edited or added information about the question) after posting it.

Greeting Logical variable indicating true if the asker included a greeting (e.g. Hello, Hi) as the first word of the text Gratitude Logical variable indicating true if the asker used manners in the question of the text (e.g. Thank you, please, appreciate)

Prior$_{e}ffortLogicalvariableindicatingtrueiftheaskerusedlanguagethatindicatesthatprioreffortorresea$

Weekday The day of the week the question was posted

Ampm The time of day the question was posted. If the question was posted between 5am and 12pm, it was categorieze das morning. If it was posted between 12pm and 5pm, it was categorized as afternoon, between 5pm and 8am- evening, and 8pm and 5 am, night

Avg$_{t}ag_{s}coreThefrequencyorscoreofatagisdefinedastheproportionoftimesthattagappearsinthedataset.T$

Text$_{l}engthThenumberofcharactersinthetext$

Device$_{l}engthThenumberofcharactersintheuser-defineddevicevariable.Thisvariablewasincludedtocapt -TurtleBeachEarForceXmygrandsonchewedthroughthewirewhilewewasplayingit'sbrand- newisthereanywayIcanhaveitfixedOOne.11questionsalsodidntincludeanydevicename.thisvariablewasc$

Avg$_{t}ag_{l}engthTheaveragenumberofcharactersintheuser-definedtags.Thisvariable,likethedevice- lengthvariable,wascreatedtocapturewhenauserincorrectlyusedthetaggingsystem.Anexampleofanincorr ineedtorepairtheheadsetbecauseicannotfindthebluetootandcomesot64charahcters.Ifaquestiondidnotinc$

Newline$_{r}atioThisvariableiscalculatedastheratioofthenumberofnewlinesaquestionincludes,overthenum$

Variables included in the final model

The above variables were all entered into the final model. The final model was stratified on ampm, as prior analysis indicated that this variable violated the proproitnal hazard assumption. The square root of the avg$_tag_scorewasentered.Viewingtheplotofanswertimesagainsttheavg_tag_screo$

Methods

Univariate analysis for each of the variables of interest was carried out on one of the training data sets. Variables with a p-value of less than 0.01 were entered into the final model. Since this is a predictive model, it was decided that all variables would be kept in the model, to maintain predictive power. After building the model on one of the training data sets, assumptions and residuals were checked.

Initally, ampm was put in the model without any stratification. Checking the proportional hazards assumption, by both examinging the schoenfeld residuals as well as the cox.zph function in r, indicated that ampm violated the proportional hazard assumption. It was decided that this variable would be stratified on.

Checking the plot of martingale residuals shows that there doesnt appear to be any obvious pattern, indicating that the functional form seems adequate (how to check residuals for splines??)

Checking the deviance residuals shows that a considerable amount of questions have residuals with an absolute value of well over 2.5. Taking those questions out of the data set and refitting the model and performing cross validation, showed that fitting the model without these questions resulted in worse predictive performance.

Looking at the score resiudals for each predictor indicated that there may be a couple points that can be considered influential. Experimentation with those questions, and refitting the model without those certain points, along with performing cross validation, showed that refitting the model without those question resulted in worse predictive performance, similar to how it did for the deivcance resiudals. After determining that the model did not grossly violate any assumptions or have any weird residuals, a cross-validation scheme was carried out.

A cross-validation scheme was used for this analysis. The data set was split into 5 training and test data sets. Training data sets had 6,208 questions and test data sets had 1,552 questions. The model was built on the training data sets, and tested on the corresponding test data sets. To assess the prediction performance of this model, metrics similar to /citeChen were computed. The model and its coefficients, built on the training data sets, were then use to calculate the predicted hazard ratios for questions in the test data sets. Those predicted hazard ratios were then entered into another cox regression model with the time until answer as the survival time variable. Metrics calculated were the hazard ratio for this model, along with the partial loglikelihood score and respective p-value. The AIC and R-square was also evaluated (explain what this is in survival analysis). Somers Dxy along with a concordance statistic were also calculated as performance metrics. This process was repeated each time a model built on a training data set was fit and predicted on the test data set. The metrics taken into consideration was the average of all of the metrics computed on the test data sets. The same metrics were also computed for predictions on the training data set itself. Although the metrics were not considerably high or impressive, they did not change from training data set to test data set.

Lastly, the model was fit to the full data set. The same process and metrics were calculated. Metrics for the full data set are not different from the results of our cross validation plan.

Results and Discussion (or separate the two).

The final model gave an AIC value of 69945.32. The partial loglikelihood ratio test statistic was 1307.15 with a p-value of 0. The R-squared value for this model is 0.1555. Variables that proved to be significant predictors with a p-value of less than 0.01 are: new$_{c}ategory(appleproduct, camera, gameconsole, home, pc, andvehical), contain_{u}nanswered, title_{q}uesito$

While the model itself overall is significant, the metrics for the models predictive accuracy are not impressive. This indicates that perhaps a different model might perform better as a model to predict the survival time for questions. Or, a model with more quantitative variables might preform better as well.

References

Asaduzzaman, M., Mashiyat, A. S., Roy, C. K., and Schneider, K. A. (2013), "Answering questions about unanswered questions of Stack Overflow," *2013 10th Working Conference on Mining Software Repositories (MSR)*, pp. 97–100.
URL: *http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6624015*

Bhat, V., Gokhale, A., Jadhav, R., Pudipeddi, R., and Akoglu, L. (2014), "Min (e) d Your Tags : Analysis of Question Response Time in StackOverflow," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, (Asonam), 328–335.

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., and Hartmann, B. (2011), "Design Lessons from the Fastest Q&A Site in the West," *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, p. 2857.
URL: *http://dl.acm.org/citation.cfm?doid=1978942.1979366*