Project EPSILON
Automatic JIRA triage

Yonatan Feleke, Gurkanwal Brar & Ashok Poothiyot

Stanford University

PROBLEM STATEMENT

The corporate world deals with task management in a variety of ways with each having some form of triaging process to correctly assign tickets to developers. Automation of this task has proven elusive with less than 60% accuracy of latest ML solutions. Web and SaaS companies handle high volumes of tickets in the form of exceptions, support requests, user-reported bugs, and crash reports. Effective automation is essential to improve productivity and obviate the tedious work of manually triaging tickets. JIRA and Asana are the most widely used task and ticketing systems to tame this beast with dedicated teams that work on aggregating, triaging and assigning these tickets to the right individual or team.

Project aims to eliminate this overhead by experimenting with supervised-learning classifiers to assign tickets to a developer. We aim to deliver higher accuracy for predicting assignee based on past tickets. In the future, we see abundant applications: automatically setting priority, EHR classification with modified featurization, automated customer support, and prioritized exception alerts.

DATASET: EXPIUM and LINKEDIN

The project utilized generated jumble.expium.com for developing the algorithm and then evaluated methods by training and testing on a private LinkedIn dataset of Foundation team support tickets. Our implementation used a bag of words multinomial event model to represent a JRIA ticket. The text components of JIRA ticket namely: the subject, body and comments sections, were concatenated and featurized with a frequency threshold of >= 5.

A particular JIRA Ticket has the following JSON structure:


```
{
  "summary": "Update success. 3.0 USB Card",
  "description": "OR memory.dmp folder, ...",
  "priority": "Minor",
  "reporter": "chantal.colman",
  "labels": [
    "Communication"
  ],
  "worklogs": [],
  "status": "Open",
  "issueType": "Epic",
  "created": "2018-09-13T14:22:15-07:00",
  "updated": "2018-11-16T16:00:00-08:00",
  "affectedVersions": [],
  "fixedVersions": [],
  "watchers": [
    "kevin.mcwhorter"
  ],
  "components": [],
  "externalId": "OLDMOB-167",
  "comments": [
    {
      "body": "No results. 3. Replacing the next
would like help someone has efficient cooling.
Crashes only hardware problems. Even something or
removing the buzz is supposed to do anything.",
      "author": "sarah.clark",
      "created": "2018-10-11T17:00:00-07:00"
    },
  "customFieldValues": [
    {
      "fieldName": "Epic Name",
      "fieldType": "com.opper.jira:gh-epic-label",
      "value": "Update success. 3.0 USB Card"
    }
  ],
  "history": [
    {
      "author": "chantal.colman",
      "created": "2018-11-06T16:00:00-08:00",
      "items": [
        {
          "fieldType": "jira",
          "field": "status",
          "from": 1,
          "fromString": "Open",
          "to": 3,
          "toString": "In Progress"
        }
      ]
    },
  ]
```


Features Models, Results and Discussion

Method

The corporate world deals with task management in a variety of ways with each having some form of triaging process to correctly assign tickets to developers. Automation of this task has proven elusive with less than 60% accuracy of latest ML solutions. Web and SaaS companies handle high volumes of tickets in the form of exceptions, support requests, user-reported bugs, and crash reports. Effective automation is essential to improve productivity and obviate the tedious work of manually triaging tickets. JIRA and Asana are the most widely used task and ticketing systems to tame this beast with dedicated teams that work on aggregating, triaging and assigning these tickets to the right individual or team.

Project aims to eliminate this overhead by experimenting with supervised-learning classifiers to assign tickets to a developer. We aim to deliver higher accuracy for predicting assignee based on past tickets. In the future, we see abundant applications: automatically setting priority, EHR classification with modified featurization, automated customer support, and prioritized exception alerts.

Results

The corporate world deals with task management in a variety of ways with each having some form of triaging process to correctly assign tickets to developers. Automation of this task has proven elusive with less than 60% accuracy of latest ML solutions. Web and SaaS companies handle high volumes of tickets in the form of exceptions, support requests, user-reported bugs, and crash reports. Effective automation is essential to improve productivity and obviate the tedious work of manually triaging tickets. JIRA and Asana are the most widely used task and ticketing systems to tame this beast with dedicated teams that work on aggregating, triaging and assigning these tickets to the right individual or team.

Project aims to eliminate this overhead by experimenting with supervised-learning classifiers to assign tickets to a developer. We aim to deliver higher accuracy for predicting assignee based on past tickets. In the future, we see abundant applications: automatically setting priority, EHR classification with modified featurization, automated customer support, and prioritized exception alerts.

Conclusions

The corporate world deals with task management in a variety of ways with each having some form of triaging process to correctly assign tickets to developers. Automation of this task has proven elusive with less than 60% accuracy of latest ML solutions. Web and SaaS companies handle high volumes of tickets in the form of exceptions, support requests, user-reported bugs, and crash reports. Effective automation is essential to improve productivity and obviate the tedious work of manually triaging tickets. JIRA and Asana are the most widely used task and ticketing systems to tame this beast with dedicated teams that work on aggregating, triaging and assigning these tickets to the right individual or team.

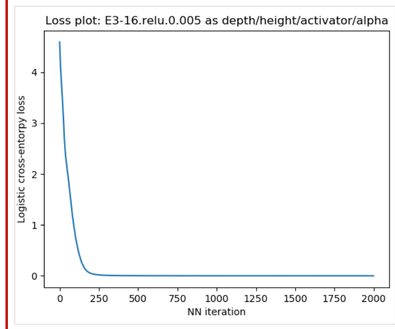
Project aims to eliminate this overhead by experimenting with supervised-learning classifiers to assign tickets to a developer. We aim to deliver higher accuracy for predicting assignee based on past tickets. In the future, we see abundant applications: automatically setting priority, EHR classification with modified featurization, automated customer support, and prioritized exception alerts.

Neural Network Design

The Neural network parameters were chosen by searching through parameters using the Expium dataset. The dimensions investigated in the images on the left are:

- Activator Functions: [tanh, relu]
- Learning rates: [0.5, 0.05, 0.005]
- Hidden layers: [8, 16, 32]
- Depth : [3, 5]

The final architecture chosen is: a 3 deep, 16 high neural, tanh activated network with a learning rate of 0.005 and 1,000 backpropagation iterations. At the bottom we list other higher accuracy networks that we voided.



The architecture chosen due to the smooth descent and relative simlicity compared to complex ones listed below

Concluding Remarks

Concluding remarks

Selecting parameters using the expium dataset proved to create bias when vectorization of the JIRA tickets leaves a lot of room t

Future Work

The parameter were calculated using Expium but doing the same grid search on the LinkedIn dataset though computationally expensive could improve our model capabilities

The feature selection and vectorization of JIRA tickets leaves a lot of room for improvement

- The current algorithm does not take advantage of NLP concepts, a great extension of this work would be to implement stemming, lemmatization and word embeddings.
- Word2Vec also promises to be an interesting pattern to test in the featurization state
- Utilize additional features such as: watchers, labels, reporter, hashed exceptions and so on.

References

[] –

[] –

[] –

[] –

[] –

[] –

[] –

[] –

[] –

[] – WUYUNTANA, D. and WANG, S. (2018). Distributed Representations of Mongolian Words and Its Efficient Estimation. DEStech Transactions on Computer Science and Engineering, (iceit).