Fuzzy Prediction of QoS for IT Maintenance Tickets

[Extended Abstract]

Suman Roy[1], Dipanjan Dutta[1], Durga Prasad Muni[1] and Adrija Bhattacharya[2]

[1]Infosys Ltd., #44 Electronic City, Hosur Road, Bangalore 560100, India

[2]Dept of Computer Science and Engineering, University of Calcutta, Kolkata, India

1. INTRODUCTION

Ticketing system, an example of Service Systems (SS), is used as one of the inputs for Information Technology Infrastructure Library (ITIL) services such as problem management and configuration management. Normally a ticket is recorded on the system with a summary which describes the concerned issues. The maintenance team examines the ticket, responds to the user with an acknowledgment with possibly seeking an explanation of the issue and indicating the possible solution, and subsequently, provides remedial actions on the ticket by writing a short note on the resolution steps, and closes the ticket in the end. The maintenance team also records other details about the ticket like, the opening date, the response date, the resolution date and the closing date for the ticket. These parameters could be used to compute different Quality of service (QoS) parameters like response time (difference between response date and open date), resolution time (difference between resolution date and open date), closure time (difference between closure date and open date) etc.

Fuzzy sets seems to be an ideal choice for modeling different QoS parameters arising in ticketing application services. Resolution time for example, could vary from person to person who are handling the same ticket. These QoS parameters cannot be expected to be recorded precisely. A more appropriate way to describe the resolution time for an incoming ticket is to say that the resolution time is high, or medium, or low. These phrases ‘high’, ‘medium’ and ‘low’ can be regarded as fuzzy quantities. This kind of categorization of ticketing services also helps people especially those with almost no technical experience, to understand reasonable QoS values as provided by the service.

Motivated by this we use a two-stage procedure for predicting QoS values for new tickets as depicted in Figure 1. First we choose Fuzzy C-means algorithm to cluster ticket data using their different QoS values and compute their membership values with respect to each cluster. We use these membership values as dependent variables and feature vector of tickets using TF*IDF values as independent variables to set up a linear regression model which helps us predict QoS values for incoming tickets.

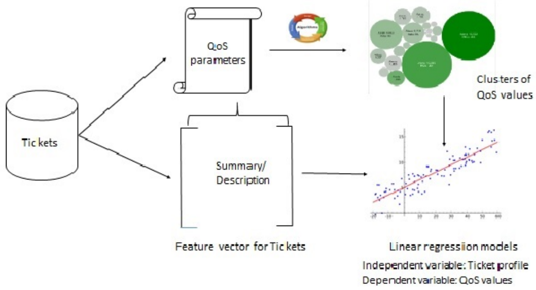

Figure 1: A two-stage procedure for categorization and prediction of QoS parameters for tickets

2. TICKET DATASET

We consider incident tickets with similar schema which are frequent in IT maintenance. We conduct experiments on IT maintenance tickets originating from 3 domains to develop QoS quality models. The domains are Travel request (Travel), Application Maintenance and Development (AMD) and Local Area Network Management (LANM). These tickets usually consist of two fields, fixed and free form. Fixed fields are customized and inserted in a menu-driven fashion. Example of such fields are the ticket’s identifier, the employee number of the user raising the ticket, priority, the time the ticket is raised, responded to or closed on the system. Various other important information are also captured through these fixed fields such as category of a ticket, sub-category, application name (application under which the ticket was raised) etc. We assume a free field of ticket to contain a succinct problem description associated with a ticket in the form of a summary (or call description).

We extract features from the collection of summaries of tickets. We use light weight natural language processing for such feature vector generation of a ticket. We can model a ticket T_i as a row vector $\mathbf{x}_i = [x_{i,1} \ \ldots \ x_{i,p}]$, where each element $x_{i,j}$ represents the importance or the weight of a term/keyword w_j in the form of TF*IDF with respect to the ticket T_i, each such feature vector contains p keywords. We shall divide the ticket data sets into few subsets based on fixed field entries which we call tuples. An example of a tuple is: (application name, category and sub-category).

3. QOS CLUSTERING USING FCM

Let us assume a QoS parameter Q modeled on ticket data, *e.g.*, response time or resolution time or closure time. The Fuzzy C-means (FCM) algorithm [2] partitions the ticket data of each tuple with respect to Q into κ clusters with $c_i, 1 \leq i \leq \kappa$, as cluster centers. Let us take a domain for Q as the collection of cluster centers, that is $\Gamma = \{c_1, \ldots, c_\kappa\}$. Then for a ticket T_i we can find the membership value of QoS parameter Q as $\mu_{ij}, 1 \leq j \leq \kappa$. That is, for each ticket we can define a fuzzy set for the QoS Q as $\mu_{i1}/c_1, \cdots, \mu_{i\kappa}/c_\kappa$ on the set Γ with the constraint $\sum_{j=1}^{\kappa} \mu_{ij} = 1$.

We run FCM clustering algorithm on each tuple for all three data sets. We validate the generated clusters using VB validity index. Based on this the optimal number of clusters for resolution time for all these tuples in the data set is set to 3. For most of the QoS parameters we consider 3 clusters, *viz.*, 'hi', med' and 'low'. The FCM algorithm generated the cluster centers of each of the QoS parameters. Subsequently, the data set of each QoS are used to produce clustering models, which are fuzzy clustering models. In these models, each cluster is described by a curve. The cluster center is represented by the peak of the curve, which has the degree of membership of 1. For example, Figure 2 shows the membership curve for resolution time of rickets in 'Software' tuple belonging to AMD dataset. Here one can estimate the range of values for each cluster (curve) for this data.

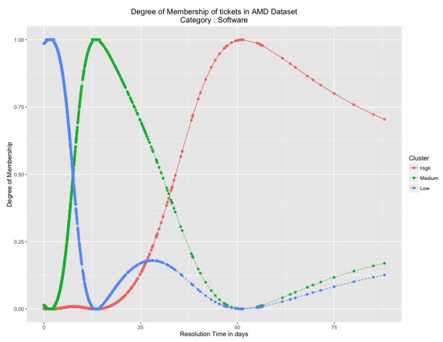

Figure 2: Fuzzy quality model for Resolution time for AMD dataset

4. PREDICTION OF QOS PARAMETERS

We use linear regression model for predicting QoS parameters modeled as Fuzzy sets. Dogman *et al.* proposed a QoS evaluation system for computer networks using a combination of fuzzy C-mean (FCM) and simple linear regression analysis by which they analyze for estimating QoS in a simulated network [1]. As in our case the fuzzy memberships are defined on discrete sets we choose to work with different linear regression models for these discrete fuzzy sets.

Now let us formulate the prediction problem for the QoS parameter Q. We have seen that each ticket T_i can be modeled using a feature vector $\mathbf{x}_i$. The responder variables can be expressed as membership values for the jth cluster as: $\mathbf{y^j} = [y_{1,j}(=\mu_{1,j}) \cdots y_{n,j}(=\mu_{n,j})]^T, 1 \leq j \leq \kappa$, where $\mu_{i,j}$ denotes membership of ticket T_i for the cluster $c_j, 1 \leq j \leq \kappa$. Further the TF*IDF matrix of tickets can be written as $\mathbf{X} = [\mathbf{1}\ \mathbf{x}_1^T\ \ldots\ \mathbf{x}_n^T]$. The regression equation for QoS parameter Q will look like $\mathbf{y^j} = X\mathbf{b^j}, 1 \leq j \leq \kappa$, where as the observed response is $\tilde{\mathbf{y}}^j$. In order to estimate $\mathbf{b}$, we formulate a non-negative least square (NNLS) problem as $\mathbf{b}^* = \text{argmin}_{\mathbf{b} \geq 0} \frac{1}{2} \left\| \left(\tilde{\mathbf{y}} - \mathbf{X}\mathbf{b} \right) \right\|^2$. Any update algorithm for solving NNLS can be used to solve this (convex programming) minimization problem. As the matrix X is sparse and skinny we use principal component analysis (PCA) to reduce its dimension before performing NNLS.

For a new ticket we need to predict its QoS parameter Q by computing its membership value *wrt* the each of the fuzzy clusters. Using NLP techniques discussed previously we can extract the feature vector of new ticket T_{in} as $\mathbf{x_{in}}^T = [1\ x_{in,1}\ \cdots\ x_{in,p}]$. For QoS Q we predict its membership value $y_{in,j} = \mu_{in,j} = \mathbf{x_{in}}\mathbf{b^j}$, where j $(1 \leq j \leq \kappa)$ denotes the jth fuzzy cluster. Need to ensure that $\sum_{i=1}^{\kappa} \mu_{in,i} = \mu_{in,1} + \cdots + \mu_{in,j} + \cdots + \mu_{in,\kappa} = 1$. Hence we normalize the membership values as $\bar{y}_{in,j} = \frac{\mu_{in,j}}{\sum_{i=1}^{\kappa} \mu_{in,i}}$, $1 \leq j \leq \kappa$, which is finally published. We assign the likely QoS of an incoming ticket as a particular cluster if the predicted membership value is the maximum for that cluster. The cluster curve can be used to determine the range of values for this predicted QoS For example, if the resolution time for a new ticket is predicted to be in 'med' cluster then we can say this ticket is likely to be resolved within 5 to 50 days by inspecting Figure 2.

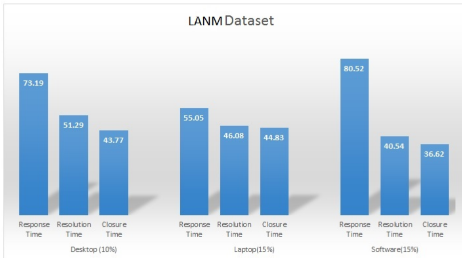

Figure 3: Matches for LANM tickets in different tuples

We validated our prediction model dividing the data sets into training and test data. For test data we predict the degree of membership of each cluster for a QoS and then compare it with the actual membership values. If the maximum predicted membership value and the actual maximum membership value belong to the same cluster we say there is a match. For LANM data we get maximum match of 51% for resolution time corresponding to 'Desktop' tuple, and 81% for response time corresponding to 'Software' tuple, these are depicted in Figure 3. For AMD data set we get a maximum match of 78% for 'administration' tuple. Travel data shows a low accuracy of maximum match as 40%. A better accuracy of matches can be obtained by refining dimension reduction technique which we leave for future work.

5. REFERENCES

[1] A. Dorgman, R. Saatchi, and S. AI-Khayatt. Quality of service evaluation using a combination of Fuzzy C-Means and Regression Model. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 6(1):62–69, 2012.

[2] M. H. Hasan, J. Jaafar, and M. F. Hassan. Development of web services fuzzy quality models using data clustering approach. In *Proceedings of the 1st DaEng-2013*, pages 631–640. Springer, 2014.