Lessons from the Implementation of an Adaptive Parts Acquisition ePortal

Rafael Alonso Jeffrey A. Bloom Hua Li
Sarnoff Corporation
201 Washington Road
Princeton, NJ 08543
609-734-2172
{ralonso, jbloom, hli}@sarnoff.com

ABSTRACT

In recent work we have developed a novel approach to the design and implementation of an online portal (**ePortal**) to help application engineers find replacements for electronic parts that have become obsolete (and hence will no longer be produced). Our approach makes use of machine learning techniques to improve the performance of a database search function. However, the purpose of this note is not to describe in detail the application nor our technical solution - that has been done elsewhere (see [1,2]). Rather, it is our intention to present some of the lessons learned from our project. Below, we provide a brief introduction to the technical approach, concentrate on several of the most salient lessons, and conclude with a description of the current state of the project.

Categories and Subject Descriptors

H.2.8 [Database Applications]: Data Mining, H.3.3 [Information Search and Retrieval]: Search Process, Clustering, Information Filtering, I.2.6 [Learning]: Knowledge Acquisition.

General Terms

Algorithms, Management, Measurement, Performance, Design, Experimentation.

Keywords

Lessons Leaned, Normalization, Adaptive Search, *k*-Nearest Neighbor classification, User Profiling, Query by Example.

1. INTRODUCTION

This presentation discusses the lessons learned in a parts acquisition ePortal study [1, 2]. A central problem in parts acquisition occurs when electronic parts become obsolete. This means that a component in a long-lived electronic system is no longer available through the original source. An *application engineer* is tasked with resolving the problem by finding a replacement part that is "as similar as possible" to the original one. The current approach involves a laborious manual search through several electronic portals and data books. The search is difficult because potential replacements may differ from the original and from each other by one or more parameters. Worse still, the cumbersome nature of this process may cause the engineers to miss appropriate solutions amid the many thousands of parts listed in industry catalogs.

Our approach[1] to the problem centers around the use of an adaptive nearest neighbor search algorithm. We define a distance measure between the obsolete part (the *target*) and each of the other parts in a large data set, where this distance measure is guided by a dynamically-adaptive user profile. Potential replacement parts are sorted by distance and presented to the user. Feedback, obtained when the user selects an acceptable replacement part, is used to update the user profile.

A block diagram of our application algorithm is shown in Figure 1. Block 1 obtains a parametric description of the target. Block 2 provides a high-level filtering of the entire catalog of data (all heterogeneous parts databases taken together) to remove parts that are clearly unacceptable (e.g., one cannot replace a memory chip with a voltage regulator). Blocks 3-7 represent calculation and presentation of the distance measure, feedback, and profile adaptation.

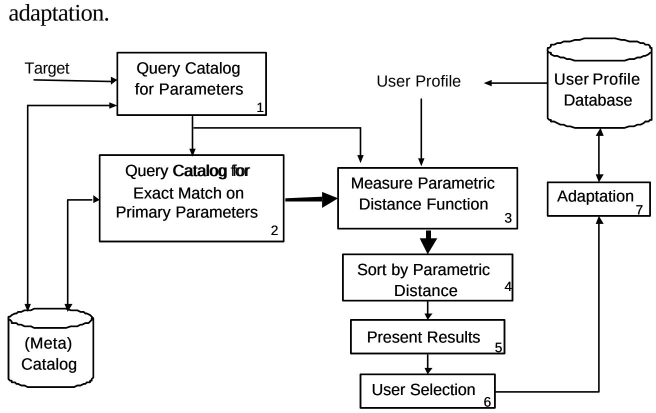

Figure 1. Adaptive Search Algorithm

CIKM'03, November 3-8, 2003, New Orleans, Louisiana, USA.

[1] Clearly, there are also database access and integration issues involved in this application. We ignore those issues in the present note since they are not relevant to the "lessons learned" described herein.

We have shown [1, 2] that our approach works well in a set of experiments employing both real-world parts databases and realistic user queries. We have developed and tested our algorithms in collaboration with a community of actual users, who have validated our research directions. Finally, the central benefit of our approach is that it reduces the time required by an engineer to locate a replacement part and also increases the likelihood that a replacement part can be found and an expensive redesign avoided.

2. Lessons Learned

In designing and implementing our technical approach, we were often surprised by unanticipated features of the problem. It is fair to say that several times we encountered difficulties where we thought all issues were straightforward, and conversely, we were happy to have smooth sailing in places where we anticipated difficulties. In this section, we present a number of the lessons we learned in applying the adaptive search algorithm (Figure 1) to the obsolete part problem. We present them here because we believe these lessons have general implications to similar applied research.

2.1 Normalization of Features

Normalization of features is very important especially in the case where different features of an object or system are combined to produce a composite metric. In our study, we defined the distance between two parts as weighted linear combination of all the parametric differences (Block 3 in Figure 1). Here each parameter (e.g., current, voltage) is a feature. The central concern here is the equivalence of the parametric distances of different features. For example, is the distance of 1 mV on voltage the equivalent of 1 mA in current? To resolve that issue, we expected to sit down with a domain expert to work out the correct estimates in short order. In practice, it turned out to be rather difficult to concretely define the normalization functions, and we had to make our own reasonable assumptions much of the time.

2.2 Evaluation Criteria

Finding the appropriate evaluation criteria for an adaptive algorithm is not always straightforward. To evaluate the performance of the profile adaptation algorithm (Block 7 in Figure 1) in our study, we asked the system to learn a preset base profile based on the feedback (i.e., the top-ranked part) provided using the base profile. We initially selected as evaluation criteria the difference between the base profile and the learned profile (see top panel of Figure 2). We observed that the learned profile did get closer to base profile but they rarely converged within a reasonable number of feedback cycles. Even when the learned profile did not converge with the base profile, the predictions by the learned profile matches the selections by the base profile very well. Thus it was not required for two profiles to be identical to produce identical predictions. Although obvious in retrospect, this had not been evident to us in the initial design. Consequently, we changed our evaluation criterion to be the errors in prediction made by the learned profile compared with the base profile (bottom panel of Figure 2). This criterion indicated our algorithm performs very well.

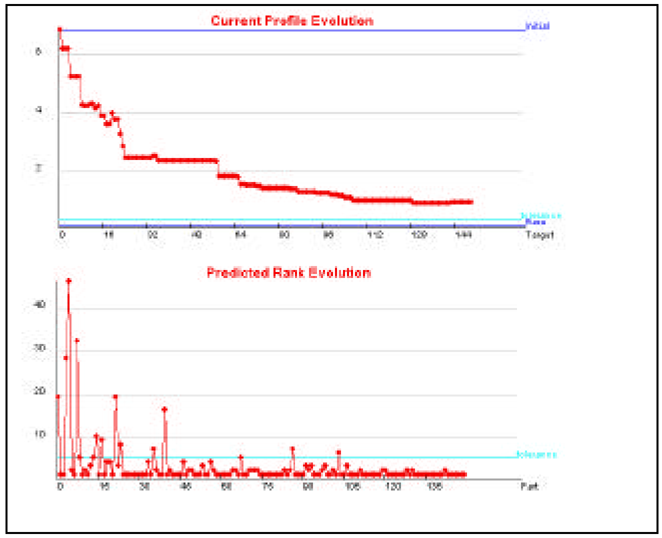

Figure 2. Evaluation Criteria: difference vs. prediction

2.3 Analyzing Unexpected Results

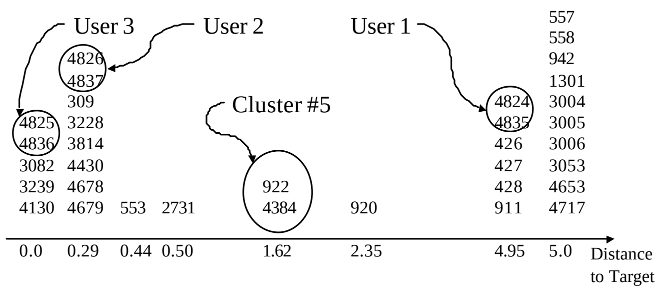

Figure 3. Parts Similar to 4823 Using a Default Profile

Simulation or experimental results can often provide insights into the problem at hand. We expect the results to confirm our predictions, however we often see unexpected results. When this happens we need to carefully analyze the results to gain new insights into the problem. In our study, we found that many parts ranked the same because they have the same exact distance from the target part (Block 5 of Figure 1). Thus there is no ordering among them. We then decided to modify our initial approach to include the concept of a *cluster* to refer to the set of parts with same ranks. Both the rank and the size of the cluster are interesting parameters. We give one example below.

The 34 parts closest to the target part 4823 are shown in Figure 3. The plot shows the 34 parts arranged vertically in 8 groups. All parts in each group are the same distance from the target and that distance is indicated below the x-axis. The 5 parts in the first group (shown on the left) are a distance of 0.0 from the target. The 8 parts in the second group have a distance of 0.29 to the target. We refer to each grouping as a cluster and note the number of parts in each cluster. Note that since all of the parts in any one cluster are the same distance from the target, the order in which they are listed in Figure 3 is arbitrary.

In order to assess the "ranking" of a part, we need to consider three values: the rank of the cluster in which it lies, the size of the cluster in which it lies, and the total number of parts in all clusters closer to the target. For example, the distance between part 4384 and the target, 4823, is 1.62. There are two parts that both have this distance and they are considered a cluster of size two. This is the fifth cluster and there are 15 parts closer to the target.

2.4 Algorithm Stopping Criteria

For our work, we needed to devise a number of stopping criteria (Block 7 of Figure 1) to deal with specific input situations. After some experimentation we settled on a "one-shot" strategy to solve the problem of thrashing. Thrashing occurs when the same candidate parts for profile adaptation are being processed repeatedly without improving the profile. The "one-shot" variant of our algorithm allows each candidate part only one chance of being used to update the profile. When all candidates are exhausted, the algorithm stops.

2.5 Using Specific Scripted Experiments to Collect User Data

User data is very difficult to collect. We originally expected to have a substantial amount of user-generated data to test our algorithms. In practice, it proved costly and time-consuming to generate even relatively small amounts of real-world trace data. We had to develop testing approaches that were parsimonious in the use of this test data. In addition, we needed to make sure that any user data gathered would be useful. We found that providing the users with specific scripts helped to achieve these goals. Figure 4 shows a sample script used to gather user data.

3. CONCLUSIONS AND CURRENT STATUS

Although we believe our Acquisition Parts ePortal project was successful, we had several unexpected results along the way. In this short note we have briefly outlined several of these "lessons learned." We hope they prove valuable to the reader.

Our project is now in a second phase, where we expect to integrate our prototype software into a deployable implementation.

4. ACKNOWLEDGMENTS

The authors would like to thank Application Engineers Kevin Kovar, Steve Leith, and Tom Smith from Rockwell Collins for providing invaluable insight into the obsolescence problem and resolution process and for providing us with the user data needed to evaluate our algorithms. Kevin also provided the domain expertise required for this project. The parts data set used for our experiments was a snapshot from CAPS Expert™, provided courtesy of IHS Corporation. The authors would further like to thank Don O'Brien and Leo Plonsky from the DLA and Curtis Holcomb from ATI Corp. for their guidance and feedback in this project. This work was completed under the sponsorship of the Defense Sustainment Consortium (DSC) as part of Defense Logistics Agency Contract No.: N00140-00-BAA3339.

Your job at GI JANE-JOE INC is to find solutions to obsolete parts for military missiles and munitions. You have a Bill of Material (BOM) of 20 parts that are all classified as not procurable. There still may be parts in distribution (short term solution) but the E-Portal Team is looking for the long term solution to the problem. It is up to you to find this solution in order to keep the factory building and aircraft flying.

General Info.

Your operating environment is considered to be severe. Your product is located in uninhabited spaces usually the wings of aircraft. This product will be exposed to extreme temperatures, large G forces, and high shock. You need to consider extreme pressure, vibration, temperature and moisture intrusion. Your product has been in the field for a long period of time, uses a 5V power supply and there is no money to redesign any circuit cards. Although a part redesign as a drop in replacement may be considered.

Good luck…

Figure 4. An example experiment script.

5. REFERENCES

[1] Alonso, R., Bloom, J., and Li, H. SmartSearch for Obsolete Parts Acquisition. *Technical Report*, Sarnoff Corporation, Princeton, NJ, 2002

[2] R. Alonso, J. A. Bloom, H. Li, and C. Basu, "An Adaptive Nearest Neighbor Search for a Parts Acquisition ePortal", Proceedings of the ninth ACM International Conference on Knowledge Discovery and Data Mining, SIGKDD'03, Washington, August 2003., Missouri, 1999.