Automatic Optimization for a Clustering Based Approach to Support IT Management

Can BOZDOGAN, Nur ZINCIR-HEYWOOD, Yasemin GOKCEN
Faculty of Computer Science

Network Information Management and Security (NIMS) Lab
http://projects.cs.dal.ca/projectx/
Introduction

• We propose a system for automatically optimizing a decision support system for IT management tasks in order to manage problems in IT environments.

• Fixing failures and problems is a time consuming job and it is costly.
Framework
Problem(Title) and Solution Sample From Experience Data Sets

Network Information Management and Security (NIMS) Lab
http://projects.cs.dal.ca/projectx/
GOAL

1) **CES + MOGA** *(Proposed)*

   Individual Representation:

<table>
<thead>
<tr>
<th>C</th>
<th>D</th>
<th>CC</th>
<th>EC</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>C : Content Value (0-1)</td>
<td>D : Dictionary Threshold Value (0-1)</td>
<td>CC : Core Clusters Threshold Value (0-1)</td>
<td>EC : Expand Clusters Threshold Value (0-1)</td>
<td>S : Sophisticate Clusters Threshold Value (0-1)</td>
</tr>
</tbody>
</table>

2) **CES + Manual** *(Previous)*

   Three main steps:
   1) Creating core clusters
   2) Expanding clusters
   3) Sophistication - Fine tuning of the cluster

   Five parameters:
   - Dictionary Threshold
   - Cluster-Cluster Similarity Threshold
   - Cluster-Document Similarity Threshold
   - Document-Document Threshold
   - Content Value

Network Information Management and Security (NIMS) Lab

[http://projects.cs.dal.ca/projectx/]
Evolutionary Component of MOGA

1. Initialize Population
2. Calculate Ranks and Fitness of Individuals
3. Generate Offspring by Crossover
4. Mutate new Offspring
5. Replace new Offspring with least fit two Individuals
6. Are All Epochs Done?
   - Yes: FINISHED
   - No: Go back to 1
Data Sets Employed

- **ParallelsDS**: Parallels Desktop for Mac, Parallels Desktop for Upgrading to Windows 7 Licensing, Activation & Registration, Parallels Sphera, Parallels Helm, Parallels Confix, Parallels SiteStudio, Parallels Plesk Products, and Parallels Virtuozzo Containers
- **GoDaddyDS**: Registering Domain Names, SSL Certificates, and Express Email Marketing.
Experiments and Results

Performance Metrics

\[
\text{Precision} = \frac{\text{\#\{retrieved documents\} \cap \text{\#\{relevant documents\}}}}{\text{\#\{retrieved documents\}}}
\]

\[
\text{Recall} = \frac{\text{\#\{retrieved documents\} \cap \text{\#\{relevant documents\}}}}{\text{\#\{relevant documents\}}}
\]

\[
\text{AvgStdDev}(\text{Cluster}_C) = \alpha \times \text{AvgStdDev}(\text{Word}_{CQ}) + \alpha \times \text{AvgStdDev}(\text{Word}_{CA})
\]

where

\[
\text{AvgStdDev}(\text{Word}_{CQ}) = \frac{\sum_{j=1}^{n} \left( \sum_{i=1}^{m} \left( \text{Words}_{i,j} - \text{Mean}(\text{Words}_{i,j}) \right)^2 \right)^{1/2}}{n^{1/2}}
\]

\[
\text{AvgStdDev}(\text{Word}_{CA}) = \frac{\sum_{j=1}^{n} \left( \sum_{i=1}^{m} \left( \text{Words}_{i,j} - \text{Mean}(\text{Words}_{i,j}) \right)^2 \right)^{1/2}}{n^{1/2}}
\]

\[
\text{Mean}(\text{Word}_T) = \frac{\sum_{j=1}^{n} \text{Word}_T}{n}
\]
Experiments and Results

Performance Metrics

\[
F_{\text{within}} = \frac{\sum_{i=1}^{m} \text{AvgStdDev(Cluster}_c) \right)}{m}
\]

\[
F_{\text{between}} = \frac{\text{EuclideanDistanceFrom}_x \_ \text{to}_y}{\sqrt{(\text{AvgStdDev}_x)^2 + (\text{AvgStdDev}_y)^2}}
\]
Experiments and Results

• Algorithms and Distance Measures Results

Average Precision, Recall and Standard Deviation calculated for each technique on all databases

Network Information Management and Security (NIMS) Lab
http://projects.cs.dal.ca/projectx/
Experiments and Results

- Algorithms and Distance Measures Results

F\textsubscript{between} and F\textsubscript{within} calculated for each technique on all databases

Network Information Management and Security (NIMS) Lab
http://projects.cs.dal.ca/projectx/
Conclusions

• Possible to automate

• Possible to adapt and optimize

• For different publicly available dataset
Questions?

gokcen@cs.dal.ca