E-COLLABORATION ON SMARTPARTICIPATION

STUDENT NAME: ZHANG YI
STUDENT NUMBER: 02-202-489
COURSE NAME: Bachelor of informatics
DEPARTMENT: Department of informatics
COURSE CODE: IN.5010

SUPERVISOR: TERAN Luis

DATE OF SUBMISSION: 15 Oct. 2013
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>4</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>4</td>
</tr>
<tr>
<td>METHODS</td>
<td>4</td>
</tr>
<tr>
<td>RESULTS</td>
<td>4</td>
</tr>
<tr>
<td>PROBLEM STATEMENT</td>
<td>4</td>
</tr>
<tr>
<td>OVERVIEW</td>
<td>4</td>
</tr>
<tr>
<td>RESEARCH QUESTION/HYPOTHESIS</td>
<td>5</td>
</tr>
<tr>
<td>OBJECTIVES AND AIMS</td>
<td>5</td>
</tr>
<tr>
<td>OVERALL OBJECTIVE</td>
<td>5</td>
</tr>
<tr>
<td>SPECIFIC AIMS</td>
<td>5</td>
</tr>
<tr>
<td>BACKGROUND AND SIGNIFICANCE</td>
<td>6</td>
</tr>
<tr>
<td>Why use dynamic fuzzy clustering?</td>
<td>6</td>
</tr>
<tr>
<td>Why use sentiment analysis?</td>
<td>6</td>
</tr>
<tr>
<td>Why content awareness recommendation?</td>
<td>6</td>
</tr>
<tr>
<td>Why use Drupal as Framework?</td>
<td>7</td>
</tr>
<tr>
<td>RESEARCH DESIGN AND METHODS</td>
<td>7</td>
</tr>
<tr>
<td>TIME SCHEDULE</td>
<td>7</td>
</tr>
<tr>
<td>STRENGTHS AND WEAKNESSES OF THE STUDY</td>
<td>7</td>
</tr>
<tr>
<td>STRENGTHS</td>
<td>7</td>
</tr>
<tr>
<td>WEAKNESS</td>
<td>7</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>8</td>
</tr>
<tr>
<td>1. SmartParticipation Platform</td>
<td>8</td>
</tr>
<tr>
<td>2. eCollaboration Extension for SmartParticipation Platform</td>
<td>8</td>
</tr>
<tr>
<td>3. Fuzzy Logic Algorithm</td>
<td>8</td>
</tr>
<tr>
<td>3.1 Fuzzy Logic, Fuzzy Sets, and Fuzzy Clustering</td>
<td>8</td>
</tr>
<tr>
<td>3.2 Fuzzy C-Means Algorithm</td>
<td>8</td>
</tr>
<tr>
<td>3.3 Comparison with sharp logic algorithm (K-means Algorithm)</td>
<td>8</td>
</tr>
<tr>
<td>4. Introduction of Sentiment Analysis</td>
<td>8</td>
</tr>
<tr>
<td>4.1 Document-level sentiment analysis;</td>
<td>8</td>
</tr>
<tr>
<td>4.2 Sentence-level sentiment analysis;</td>
<td>8</td>
</tr>
<tr>
<td>4.3 Aspect-based sentiment analysis;</td>
<td>8</td>
</tr>
<tr>
<td>4.4 Comparative sentiment analysis;</td>
<td>8</td>
</tr>
<tr>
<td>4.5 Sentiment lexicon acquisition;</td>
<td>8</td>
</tr>
<tr>
<td>5. Implementation of eCollaboration Extension for SmartParticipation</td>
<td>8</td>
</tr>
<tr>
<td>5.1 Architecture Overview</td>
<td>8</td>
</tr>
</tbody>
</table>
ABSTRACT

Background

Nowadays more and more people have been used to connecting with each other, in order to sharing their ideas, opinions, suggestions or other information in any kind of media form via internet. Each person in eCollaboration is facing with a very critical problem – How can a user with profile obtain the most favorable and/or most needed information dynamically?

Methods

The project is based on
1. a user profile generation by dynamic fuzzy clustering
   1.1 User Profile Generation
   1.2 Modified fuzzy c-means algorithm
   1.3 Clustering plan with Cron-job process, if necessary
2. A recommendation Engine with following features:
   2.1 Sentiment analysis, of which the architecture will be decided based on the user profile
   2.2 Content awareness recommendation
3. The project will be implemented with Drupal 7 framework.

Results

Incl. a literatural thesis and an implementation of the eCollaboration Extension of Smartpartition with PHP/MySQL (Drupal framework)

PROBLEM STATEMENT

Overview

All the necessary theories for eCollaboration have been done or even been applied in some practical field. For example, economist are used to applying FCM for trying to find the best valued leads (potential customers) in CRM.

However, this project are not only related to numbers, like ages, assets, or invests. A further challenge is encountered against the implementing: Sentiment analysis, in other word, opinion mining.
Briefly, the users in eCollaboration are playing as roles with two faces. They are not only the subject to affect the social network, but also are affected by the changed platform, which will advice them with the context-awareness recommendation, based on the fuzzy clustering categories, which are evolved with all efforts of the users.

The first key point in implementation should be building the data set and the process of dynamic fuzzy clustering.

Then, sentiment analysis and context-awareness recommender engine.

**Research Question/Hypothesis**

Prerequisites: It is assumed, there is NO SUPERVISOR/ADMIN, which can change the user profile from back-end. The eCollaboration is based on a fully autonomous system.

Each user profile is a member of reference set. (not fuzzy)

Each category or class of user is a member of fuzzy clustering.

The dynamic fuzzy clustering with modified fuzzy C-means algorithm in following conditions:

- An initial default fuzzy clustering
- Each Add/Delete/Modification of the user profile will affect or even restructure the clustering
- The modification of user profile will be and will only be caused by the operation of user.

Thus, the all concerned research questions shall be like following:

1. How many dimensions must be included in the user profile?
2. How the weights shall be participated into the dimensions?

By sentiment analysis, what kind of level should be applied and for which dimensions (articles, comments, ratings, and etc.)

**OBJECTIVES AND AIMS**

**Overall Objective**

A fully autonomous eCollaboration-based CMS or even a SNS (Social network system)

**Specific Aims**

- User Profile generation, which contains:
  i. User Profile data structure
  ii. A quantified regular-set of User Profile generation/modification, which could cover (be mapping to) all the user interactive operations.

- Dynamic fuzzy clustering, which contains the following controllable variables
i. Dimensions of classification
ii. Weights for each dimension

- Sentiment analysis for user Profile modification / re-generation
  Obviously, the unclassified or undefined “Tag”, which should be recognized and combined to a certain clear point of view by everyone, cannot be missed, when a well-defined sentiment analysis were implemented. The primitive method could be a Tag-system, like RDF by manual works.

- Context-awareness recommender engine
  This is not a main research topic in this project, but a key point, which will influence the platform over time. If the user cannot obtain their wished, interested or even critical information, then the platform cannot reflect what the user really think and need. Therefore a suitable context-awareness recommender will be applied.

- Easy-to-use and visual-friendly UI

**BACKGROUND AND SIGNIFICANCE**

- Why use dynamic fuzzy clustering?
  People don’t think in 0 and 1, as a computer. A man that likes a red hat, maybe could accept black berries. Thus, fuzzy data set should be applied for a better simulation.
  Every minute people obtain the information and at the same time they also create the information. During the informative interaction their minds / opinions (user profile) will be changed all the time. Thus, the classification for all the user profiles should not be a constant set, but dynamic.

- Why use sentiment analysis?
  Sentiment analysis, in another word, opinion mining is the key to modify the user profile based on the relevant specification in one or more proper dimensions.

- Why content awareness recommendation?
  The eCollaboration platform should have the functionality, in which the system can offer the probable user or recommended information to each user after login according to current user classification.
Why use Drupal as Framework?

Since Drupal, a free and open-source content management framework written in PHP and distributed under the GNU General Public License, which conforms the OOP and MVC standards.

RESEARCH DESIGN AND METHODS

Time schedule

STRENGTHS AND WEAKNESSES OF THE STUDY

Strengths:
+ Dynamic fuzzy clustering based on FCM algorithm, which is well known and developed.
+ Easy-to-use implementing Framework: Drupal

Weakness:
- Unknown needed dimensions for user profile, due to lack of requirement analysis.
- Unsure needed sentiment analysis level.
- Unknown definition of context glossary, for sentiment analysis and recommendation.
- Unknown complexity and performance of the dynamic fuzzy clustering, when a massive amount of parallel user profile modification is encountered.
# TABLE OF CONTENTS

1. SmartParticipation Platform

2. eCollaboration Extension for SmartParticipation Platform

3. Fuzzy Logic algorithm.
   3.1 Fuzzy Logic, Fuzzy Sets, and Fuzzy Clustering
   3.2 Fuzzy C-Means Algorithm
   3.3 Comparison with sharp logic algorithm (K-means Algorithm)

4. Introduction of Sentiment analysis
   4.1 Document-level sentiment analysis;
   4.2 Sentence-level sentiment analysis;
   4.3 Aspect-based sentiment analysis;
   4.4 Comparative sentiment analysis;
   4.5 Sentiment lexicon acquisition.

5. Implementation of eCollaboration Extension for SmartParticipation
   5.1 Architecture Overview
   5.2 Implementing data structure of user profile
   5.3 Implementing architecture of a generic sentiment analysis system
   5.4 Implementing Dynamic Fuzzy Clustering
   5.5 Applying some Context-awareness recommender engine and Top-N recommender engine

6. Test with sample user profile

7. Conclusion and Outlook in future

(The order and the contents of headings could be modified during the studying process.)
Referencing

- Ronen, Feldman. "Techniques and Applications for Sentiment Analysis"
  DOI:10.1145/2436256.2436274
- Linas, Baltrunas. “Context-Aware Collaborative Filtering Recommender Systems”
- Drupal Community Documentation by https://drupal.org/documentation