A Fuzzy Logic Approach for Dynamic User Interests Profiling

Authors: Abd El Heq Silem, Hajer Taktak, Faouzi Moussa

Presented by: Abd El Heq Silem
Affiliation: Faculty of Sciences of Tunis, LIPAH LR11ES14, University of Tunis El Manar
Email: hakou.silem@gmail.com
My name is Silem Abd El Heq, I am a PhD student at the University of Tunis El Manar.

I got my master degree in 2017 in the field of distributed architecture (computer science). I have a published paper as a second author in an indexed journal titled: “NorJADE: An Open Source JADE-Based Framework for Programming Normative Multi-Agent Systems”.

Currently, I am working in the field of context-awareness and personalization systems to get my PhD.

Finally, this is my first published paper in this field.
Plan

- Introduction
- Related work
- Fuzzy logic
- Proposed approach
- Conclusion
Introduction

- The exponential growth of information available in the web make the search for specific information very difficult and time consuming.
- The personalization systems has been invented to solve this problem.
- These systems provide relevant information to the user based on his needs.
- The user needs are collected and stored in a database called user profile.
- User profile is a virtual representation of the user that holds variety of user information such us: preferences, interests, enviroments,
In literature, the researches create the user profile based on
- The text retrieved from the visited Web pages (first group)
- The text retrieved from the visited Web page and the user behavior in this web page (second group)

Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1

<table>
<thead>
<tr>
<th>Fuzzy</th>
<th>Logic</th>
<th>Is</th>
<th>a</th>
<th>form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Of</td>
<td>Many</td>
<td>Valued</td>
<td>Logic</td>
<td>Which</td>
</tr>
<tr>
<td>The</td>
<td>Truth</td>
<td>values</td>
<td>Of</td>
<td>variables</td>
</tr>
<tr>
<td>may</td>
<td>be</td>
<td>any</td>
<td>real</td>
<td>number</td>
</tr>
<tr>
<td>between</td>
<td>0</td>
<td>and</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Fuzzy Logic Is a form
Of Many Valued Logic Which
The Truth values Of variables
may be any real number
between 0 and 1

Step 2: Remove stop words

Fuzzy Logic form

Value Logic
Truth values variables
real number between

Step 3: Stem all words

Fuzzy Logic form

Value Logic
Truth valu variables
real number
Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1.

- After the generation of term vector, the approaches will assign a weight to each word based on the occurrence and the position in the document (website).
- The resulted weights will be used to update the user profile.
- As a disadvantage, this group of approaches does not differentiate between interesting and uninteresting websites which generate many wrong interests when the user visits uninteresting websites.
Perform all the steps of the first group, then uses the user behavior to calculate a unit score and add it to the weight of each word.

The user behavior (factors) such as: bookmarking, copying, scrolling speed, time spent.

- **Discrete variable**
  - Add a unit score X to the calculated weight
- **Continuous variables**
  - Calculate the weight of the captured values with fixed ranges (if $a < \text{scrolling speed} < b$ then $X$)
Related works

Term Vector = { (Fuzzi, 1+X₁), (Logic, 2+ X₂), (Form, 1+X₃), (Valu, X₄), (Truth, X₅), (Variabl, 1+ X₆), (real, 1+ X₇), (number, 1+ X₈) }

User Behavior

Term Vector = { (Fuzzi, 1+X₁), (Logic, 2+ X₂), (Form, 1+X₃), (Valu, 2+ X₄), (Truth, X₅), (Variabl, 1+ X₆), (real, 1+ X₇), (number, 1+ X₈) + Y }

Disadvantage:

The second group of approaches solve the problem of the previous one. But the evaluation of user behavior (use fixed ranges) is not accurate to all users.
Fuzzy logic

- Fuzzy logic belongs to many-valued logic.
- The fuzzy logic represents the situation with a continuous value from 0 to 1 unlike the binary logic that represent the situation with 0 or 1.
- The fuzzy logic give the computer the ability to represent the human unclear idea.
- For example: when describing a room brightness using binary logic, we can only represent two situation where the room is dark or bright (0 or 1).
- With fuzzy logic we can represent the degree of light and say little bright (0.6), little dark (0.4), very dark (0), very bright (1).
In this paper, we propose an approach that predict the user interest in the content of a Web page based on three factors are: scrolling speed, Time spent, and number of visits.

The proposed approach is based on the fuzzy logic.

This approach has three main steps are: fuzzification, inference and defuzzification (see figure below).
Proposed approach

- Fuzzification phase:
  - The fuzzification phase is the first phase of the approach. It transforms a crisp value to a fuzzy set.
  - The fuzzy sets are sets that contain factor values (crisp values) with the same interpretation.
  - This process adapts the user ranges according to his habits as follows:
    - First, the approach collects information about the user (the values of the factor “scrolling speed” in this case). Then, sort all the collected data in ascending order, finally calculate the initial ranges.
    - When a new data is captured, the approach sorts this new data with the old ones in ascending order and recalculates the ranges.
  - The figure in the next diapositive explains the whole process of fuzzification.
Proposed approach
The Inference Engine is the core of the Fuzzy Logic system responsible for the calculation of one fuzzy output from a set of fuzzy inputs (fuzzy inputs calculated in the fuzzification phase).

The inference engine is composed of a set of rules as follows:

IF input1 is A AND input2 is B AND input3 is C THEN output is D

Some of the rules used in our Fuzzy logic system:

- IF SC is High AND TS is High AND NV is High THEN User is Interesting
- IF SC is High AND TS is Low AND NV is Medium THEN User is Likely Interesting
- IF SC is Medium AND TS is Medium AND NV is Medium THEN User is Likely Interesting
- IF SC is Low AND TS is Low AND NV is Low THEN User is Uninteresting

Where SC → Scrolling speed, TS → Time Spent, NV → number of visits.
Proposed approach

Inference phase

- The output of this phase (fuzzy output) is the aggregation of the results of all rules.
- The figure below represents a fuzzy output.
Proposed approach

Defuzzification phase

- The defuzzification process is the inverse of the fuzzification process.
- In this phase, the approach transforms the fuzzy output (the output of the inference phase) into a crisp output that can be used by other applications.
- The defuzzification is performed based on a decision-making algorithm called Center Of Gravity (COG), this function returns the center of the fuzzy output area (see the figure).
In this paper, we use the ontology-based model to represent and store the user profile. This profile contains the user interesting topics with the interest degree. It has two main classes:

- **User Interests**: contain user interest websites. This class has five attributes: URL of the website, scrolling speed, time spent, number of visits, interest degree calculated by our approach.

- **Topic**: represents the topic of the website. This class has only one attribute “Label” that represents the name of the topic (e.g., machine learning, sport).

The user profile is created manually using protégé (for the first time).
Proposed approach

User profile
Proposed approach

User profile

- The approach updates the user profile every time a new website arrives (the user browses a website) as follows:
  - If the website already exists in the user profile:
    - Update all attributes (scrolling speed, time spent, number of visits, and interest degree) with the average between the new and the old values (number of visits are rounded).
  - Else (the website does not exist in the user profile):
    - Add the new website with all attributes to the user profile.
Conclusion

- The user profile contains information about the user that helps the customization systems to provide data or service to the user’s needs.
- We propose a new approach to automatically construct and update the user profile using a Fuzzy Logic system.
- This approach uses three factors are scrolling speed, time spent, and number of visits.
- The approach solves the problems presented in the literature (factor weight misinterpretation and the generation of false interests).
Thank you for your attention