A TRUST BASED RECOMMENDATION SYSTEM USING SELF ORGANIZED MAP FOR SERVICE SELECTION

SAMIA BOUZEFRAINE
samia.bouzefrane@cnam.fr
CEDRIC LAB – CNAM- PARIS

NEW ARCHITECTURE MODELS & PARADIGMS

(Industrial) IoT devices

Outsourcing

Generation of services

Virtualization of services

Multi-Access Edge Computing
TRUST OF SERVICES

A trusted service means what?

- A trusted service provider?
- A secure service?
- A good QoS to access the service?
- A good reputation of the service as given by the consumers?
TRUST RELATIONSHIP

• Trust relationship can be built based on many indicators:
  – QoS
  – Security
  – Reliability
  – Reputation (given by the requestor or by others)
  – etc.

• The trust can be built efficiently if many criteria can be collected to evaluate the trustworthy of the service.

• The issue is how to trust a service requested by a user for which he does not have any trust indicator?
OBJECTIVE OF THIS WORK

• Assumption:
  – The reputation is our indicator of trust.

• Goal:
  – Prediction of the trust value of a service that is requested by a user for which he does not have any previous evaluation.
TERMINOLOGY USED

• **Evaluation**: is a numerical data value which is a user's appreciation for a service
  – Based on attributes of the service/service provider: QoS, security, **reputation**, etc.

• **Prediction of the evaluation**: computes the most likely evaluation that a user would have assigned to a given service.

• **Recommendation**: uses prediction to evaluate a list of services and then to propose to the user the most appropriate service.
RECOMMENDATION SYSTEMS

• **Content-based filtering**: recommends new services based on the services already consumed by the user
  – Problem with a new user (the system does not know the preferences of the user)
  – The system will recommend only the services that are similar to the user profile.

• **Collaborative filtering**: can predict a specific service to a user based on the evaluations done by users.
  – This approach assumes that if several users have similar preferences for a group of services, they may have the same preferences for another group of services (if users agree for the evaluation of services in the past, they are likely to agree in the future).
COLLABORATIVE FILTERING

• Collaborative filtering is a part of ML approaches.
• Based on the experiences of the users

• Neighborhood Based (memory) Method
  – The reputation of a service $s$ is predicted thanks to the feedbacks given by similar (neighbors) users to user $u$ / The reputation is given by user $u$ to similar services as service $s$
  – Similarity ratings may be resource-intensive (to store the feedbacks of the users)

• Model-Based Recommendation Method
  – Off-line models (bayesian classification, K-means, PCA, etc.) for prediction.
  – Sensitive to missing data.
Aim: to predict classifications for new services by two ways:

- **User-based approach**: evaluates the interest of user $u$ for a service $s$ using the ratings given by neighbours of $u$ (who have similar ratings patterns as user $u$)

- **Service-based approach**: evaluates the interest of user $u$ for a service $s$ using the ratings of $u$ for services that are similar to $s$

$$k = \text{argmax} \ f(u, s)$$

$f$: regression/classification
NEIGHBORHOOD BASED/2

Aim: to predict classifications for new services by two ways:

- **User-based approach**: evaluates the interest of user $u$ for a service $s$ using the ratings given by neighbours of $u$ (who have similar ratings patterns as user $u$)

- **Service-based approach**: evaluates the interest of user $u$ for a service $s$ using the ratings of $u$ for services that are similar to $s$

  - The model is trained with a dataset
  - The model is then used to predict ratings of users for new services

Model: Support Vector Machine, Bayesian Clustering, etc.
**PREDICTING EVALUATIONS IN CF**

Predictive value

The mean of evaluations given by user u

Users who evaluated service s

Evaluation of service s by user v

Similarity between user u and user v (cosine similarity/Pearson correlation coeff.)

\[
\hat{r}_{us} = \bar{r}_u + \frac{\sum_{v \in U_s} Sim(u, v) \times (r_{us} - \bar{r}_v)}{\sum_{v \in U_s} Sim(u, v)}
\]

Predictive value
OUR APPROACH

• Based on SOM (Self Organizing Map)
• Recommendation system to predict the trust/reputation of a service
  – How to predict a trusted service which is not already evaluated/known by a user?
SOM: SELF-ORGANIZING MAP

✓ Unsupervised learning (neural networks)
✓ based on vector quantization
✓ Reduction of dimensions
✓ Preserve the topological structure of the entry space (neighborhood between classes)
✓ Allows data analysis including non-linear relations
✓ Resiliency for missing data
TRUST EVALUATION USING SOM

We consider a set of similar users (to user u) evaluating a set of similar services (to service v)
Similar users will behave like user u when ranking similar services
PROPOSED MODEL

Users

1: Invoke service

2: Provide service

Services

Legend:

- Process
- Data

SOM Based Trust Recommendation System (SOM-BTR)

- Trust indicators dataset
  - SOM-Services Module
  - SOM-Users Module

- Services Model
- Users Model

- Trust indicators prediction Module
- Untrusted users detection Module (unreliable ratings)

- Recommendation Module

1: Invoke service

2: Provide service

3: Feedbacks

4: Ask for recommendation

5: Get recommendation
The mean of set $N$ is equal to 14.72

While:
- 60% of the values have a mean equal to 1.7
- 80% of the values have a mean equal to 3.1

The descriptive average is based on the classification with K-means (k=3)
We apply K-means on the set $N$ for $K$ clusters ordered like the following:

$$|C^1_N| \geq |C^2_N| \geq \ldots \geq |C^K_N|$$

We fix the parameter $\alpha$ between 0 and 1:

$$\frac{|C^1_N| + |C^2_N| + \ldots + |C^j_N|}{|N|} \geq \alpha$$

$$j = \arg \min_{1 \leq m \leq k} \sum_{m=1}^{j} \frac{|C^m_N|}{|N|} \geq \alpha, 0 < \alpha \leq 1$$

$K=3$, $\alpha=0.6$

$DA=1.7$

$$DAv^\alpha_k(N) = \frac{\sum_{i=1}^{\lfloor \xi \rfloor} n_{i,}}{|\xi|}, n_{i,} \in \xi$$

$$\xi = \bigcup_{i=1}^{j} C^i_N$$
PHASE 1: DESIGNING THE MODEL

Offline Process

Generation of the profiles of the services

Generation of the profiles of users

SOM-Users Module

SOM-Services Module

Services Model

Users Model

Trust indicators prediction Module

Untrusted users detection Module (unreliable ratings)

Recommendation Module

Trust indicators dataset

SOM model for services

SOM model for services
PHASE 2 : RECOMMENDATION OF TRUST

On-line Process

1. Compute similarity using BMN: Best Matching Neuron (based on the distance)

\[ S_v = \{ j \in S \mid BMN(j) = BMN(v) \} \]

2. Predict the reputation/trust of service s using Descriptive Average (using K-means)

\[ r'_{uv} = DA_v^\alpha(R_{uv}) = \frac{\sum_{r_{xy} \in \xi} r_{xy}}{|\xi|} \]

\[
Avec \quad \xi = \bigcup_{k=1}^{j} C_{R_{uv}}^k \quad et \quad j = \text{argmin} \sum_{1 \leq m \leq K} \sum_{m=1}^{j} \frac{|C_{R_{uv}}^m|}{|R_{uv}|} \geq \alpha , \quad 0 < \alpha \leq 1.
\]
PHASE 3: UPDATE OF SOM MODELS

Algorithm 1

High-level description of the update algorithm

Input: $\Delta T$ and $\Delta R$
Output: Updated model

Initialization: $t = 0$ time initialisation, $f = 0$ number of feedbacks, $M$ numbers of neurons for services SOM-model

1. while $t < \Delta T$ and $f < \Delta R$ do
2.   if $n$ new feedbacks are reported then
3.     $f = f + n$
4.   end if
5.   increment($t$)
6. end while
   // execute phase 1:
7. for all $s \in S$ do
8.   perform expressions (7) to (10)
9. end for
10. SOM update for services
11. for $1 < i < M$ do
12.   perform expressions (12) and (13)
13. end for
14. SOM update for users
15. return $SOM_U$ and $SOM_S$

$\Delta R$: window to update the users feedbacks
$\Delta T$: time window to update the SOM models
IDENTIFICATION OF UNTRUSTED USERS

1. Invocation density: computes the density of calls done by a specific user \( u \) compared to others (\( u \) may have malicious behavior)

2. Coefficient of user Aberration: is the deviation of the feedbacks of user \( u \) compared with the feedbacks of other similar users

3. Credibility coefficient of the user \( u \).
INVOCATION DENSITY OF THE USER

Number of times user $u$ invokes the services of group $S$

$$D_{u,s} = \frac{n_{us} + 1}{DAv_k^\alpha(N_s) + 1}$$

$\approx 1$ means normal behavior

Distribution of the number of times that all the users invoke services of group $S$
COEFFICIENT OF USER ABERRATION

\[ A_{br u,s} = \sum_{i=1}^{k} |C_i| \times D_{c i,s} \]
**USER CREDIBILITY FACTOR**

\[ CrC_{u,s} = Abr_{u,s} \times e^{D_{u,s}} \]

 Invocation Density

Aberration coefficient

\( \gamma \) is a threshold

If \( CrC_u \leq \gamma \)

user u is trusted
EXPERIMENTAL EVALUATION

WS-DREAM repository
✓ 339 users (U) distributed over 30 countries
✓ 5825 Web services (S) distributed over 73 countries
✓ Matrix UxS has 1 974 675 evaluations
✓ Values of Response Times vary between 0 and 20 seconds.

Density of Matrix DM to simulate sparsity
✓ MD = 30 % (70% missing data)
✓ MD = 10 % (90% missing data)
✓ MD = 05 % (95% missing data)

Untrusted users
✓ 2.9% users, 5.9% users
✓ 8.8% users, …, 17.7% users
ROBUSTNESS AGAINST UNTRUSTED USERS

RMSE: root-mean square error

Smaller is RMSE, greater is the prediction accuracy

The model is stable despite the increasing number of untrusted users.
COMPARATIVE ANALYSIS OF SOM-BTR

1. UPCC : similarity between users – PCC (Pearson Correlation Coefficient)
2. IPCC : similarity between services – based on cosine
3. WSRec : combines UPCC (User-based PCC) and IPCC (Item-based PCC)
4. GNMF : geographical neighbourhood - PCC - matrix factorization
5. TAP : IPCC - K-means for users
6. GURAP : geographical neighbourhood - reputation of users
7. SOM-BTR : similarity of services and users based on SOM model
COMPARATIVE ANALYSIS OF SOM-BTR

MD = (5%, 10%, 20% & 30%)
Untrusted users = 10 (over 339 users)

<table>
<thead>
<tr>
<th>Methods</th>
<th>MD=5%</th>
<th>MD=10%</th>
<th>MD=20%</th>
<th>MD=30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPCC</td>
<td>2.305</td>
<td>2.183</td>
<td>2.149</td>
<td>2.079</td>
</tr>
<tr>
<td>UPCC</td>
<td>2.200</td>
<td>2.106</td>
<td>2.149</td>
<td>2.096</td>
</tr>
<tr>
<td>GNMF</td>
<td>2.275</td>
<td>2.078</td>
<td>1.703</td>
<td>1.655</td>
</tr>
<tr>
<td>WSRec</td>
<td>2.052</td>
<td>2.036</td>
<td>2.029</td>
<td>1.967</td>
</tr>
<tr>
<td>TAP</td>
<td>1.664</td>
<td>1.631</td>
<td>1.652</td>
<td>1.685</td>
</tr>
<tr>
<td>GURAP</td>
<td>1.494</td>
<td>1.433</td>
<td>1.332</td>
<td>1.284</td>
</tr>
<tr>
<td>SOM-BTR</td>
<td>1.396</td>
<td>1.352</td>
<td>1.333</td>
<td>1.320</td>
</tr>
</tbody>
</table>
CONCLUSION

• SOM-based approach for trust recommendation (first work that investigates SOM in this context)
• The selection is based on one trust indicator => needs extension to several trust indicators
  – Hyrid solution
• Adapt this approach to practical contexts (related to 5G or 6G for example).