Semi-Automated Assessment of Annotation Trustworthiness

Davide Ceolin, Archana Nottamkandath, Wan Fokkink
Museums...
Museums... have a problem.
Let’s recruit some help
Red!
Red!

Tulips!
Tasty!

Red!

Tulips!

Tasty!
Red!

Tulips!

Tasty!

???
Museums are meticulous.

- The crowd alone does not solve the problem.
- We need to select only the most trustworthy annotations.
- System requirements:
  - reliable
  - with minimum effort
  - efficient
User Reputation

User reputation should reflect the trustworthiness of their annotations.

Flower
Tulip
Ugly
Reputation estimation

Tulip

Flower

Red

Pink

Purple
Reputation estimation

We compute reputations as in subjective logic.
Reputation estimation

- Tulip
- Flower
- Red
- Pink
- Purple

Few annotations per author:
✓ minimum effort

![Reputation values](image)
Interpreting reputation
Interpreting reputation

User reputation: 0.6
Interpreting reputation

User reputation: 0.6

Should we accept all his annotations? Reject all?
Interpreting reputation

User reputation: 0.6

Should we accept all his annotations? Reject all?

No: accept only the best 60%.
How?
Estimating expertise can help.
We are almost there...

Rank the annotations and evaluate them.

Test set

Rose 0.9
Violet 0.8
...
...
...
Tomato 0.3

60% accepted
40% rejected

But we can improve the efficiency of the system
Semantic clustering

Clustering the training set semantically helps in improving the system efficiency.
Semantic clustering

Clustering the training set semantically helps in improving the system efficiency.
Semantic clustering

Clustering the training set semantically helps in improving the system efficiency
Semantic clustering

Clustering the training set semantically helps in improving the system efficiency.
Results

• We run some experiments with from 5 to 30 annotations per user. On two datasets.

• Accuracy: 68% - 84%

• Precision: 87% - 88%

• Recall: 80% - 96%

• Time saved by clustering: 44% - 52%
Discussion

• Results are satisfactory but they can further be improved
• Provenance can help
• Image analysis programs can help
• Reuse of evaluations can reduce the museum effort
Recap

- We propose a system to evaluate the trustworthiness of museum annotations
- Uses reputation, semantic similarity and clustering
- Reliable
- With minimum effort
- Efficient
Recap

- We propose a system to evaluate the trustworthiness of museum annotations
- Uses reputation, semantic similarity and clustering
- Reliable
- With minimum effort
- Efficient

Thanks!