Amathae: Information Discovery and Filtering using a Multiagent Evolving Ecosystem

Alexandros Mouskas
MIT Media Lab.

Motivation
- Information overload
- Suited to this domain
  - Automatical evolution rate control
  - Quick adaptation to an user’s new interest
  - Exploration VS Exploition
- Various application domains
  - WWW (using search engines)
  - information streams (news server)
  - site monitoring

Features
- Digest reporting
- Background operation
- Multiple agents
  - Information Filtering Agents (IFAs)
  - Information Discovery Agents (IDAs)

Question
The two most important features of Amathae?

Features
- Bottom-up approach for intelligent behavior
  - An user usually has some overlapping interests.
  - Very simple agents cooperate and compete with one another
  - Distributed adaptation
- Training agents using an genetic algorithm

Bootstrapping Amathae
- Generation of filtering agents
  - Bookmarks
  - Browsing history file
  - Pointing a specific page
  - Pretrained packages of agents
- Generation of discovery agents
  - Random assignment of WWW search engines
**Amalthaea’s Architecture**

- User Interface - Configuration
  - Monitoring Amalthaea’s Operation
    - Information Filtering Agents
      - Document Representation
        - Weighted keyword vector representation
        - Preprocessing
        - Weighting keywords by TFIDF measure
        - Adjusting weights with HTML tags
          
          \[
          idf_i = \log \left( \frac{N}{df_i} \right) 
          \]

          \[
          W_i = H \cdot T \cdot idf_i 
          \]

**User Interface - Site Monitoring**

**Information Filtering Agents**

- General Properties
  - An IFA represents a very specific interest of the user.
  - Similar agents are clustered together.
  - An user-creating agent is assumed to represent a long-term interest of the user. The agent is not destroyed easily.

- Operation
  - request documents whose types they are interested in.
  - filter documents using the distance between two keyword vectors a, b

\[
D_{ac} = \sqrt{\sum_{j=1}^{m} \left( \frac{a_j - b_j}{\sqrt{(a_j^2 - a_j^2)} \sqrt{(b_j^2 - b_j^2)}} \right)^2} 
\]
Information Discovery Agents

- **Genotype**
  - Search Engine / information source to use
  - Number of keywords
  - Query type ("AND", "OR", etc)
  - Minimum Hits, Maximum Hits

Interaction between IFAs and IDAs

- IFAs post their requests in the request list.
- Each IDA selects the request of a suitable IFA by checking its transaction history. The IDA erases the request in the request list. (It implies n:1 relationships)
- IDAs get documents.
- IDAs present corresponding IFAs retrieved documents.

The Digest and Relevance Feedback

- **Making the digest**
  - Each IFA submits one document to Amalthaea.
  - Amalthaea shows the user top-ranked documents only with high confidence level.
  - Relevance Feedback
    - Site rating
    - Picking keywords that best describe the given document
    - Monitoring the user’s behavior

User Interface - The Digest

Ecosystem

- **Credit Assignment**
  - IFAs receive credits based on the user’s feedback
  - IFAs give part of the credits to corresponding IDAs.
  - Its credit is Its fitness.
  - Evolution happens every given amount of generations.

Evolution (Learning)

- **Selection:** only the fittest agents allowed
- **Crossover:** 2-point crossover
  - The child inherits parts of the genotypes from parents
- **Mutation:** the genotype of an agent + a randomly selected, randomly weighted keyword from other agents or highly-rated documents
- **Cloning:** just copying good agents