Recommender Systems in Requirements Engineering

Bamshad Mobasher and Jane Cleland-Huang

Requirements engineering in large-scale industrial, government, and international projects can be a highly complex process involving thousands or even hundreds of thousands of potentially distributed stakeholders. The process can result in massive amounts of noisy and semistructured data that must be analyzed and distilled in order to extract useful requirements. As a result, many human-intensive tasks in requirements elicitation, analysis, and management processes can be augmented and supported through the use of recommender system and machine-learning techniques. In this article we describe several areas in which recommendation technologies have been applied to the requirements engineering domain, namely stakeholder identification, domain analysis, requirements elicitation, and decision support across several requirements analysis and prioritization tasks. We also highlight ongoing challenges and opportunities for applying recommender systems in the requirements engineering domain.

Requirements engineering is the practice of eliciting, analyzing, prioritizing, negotiating, and specifying the requirements for a software-intensive system (Robertson and Robertson 1999). These activities engage various stakeholders in the task of identifying and producing an agreed-upon set of requirements that clearly specify the functionality, behavior, and constraints of the proposed system. The importance of the requirements engineering task is illustrated by several studies, which have shown that requirements-related issues, such as poorly specified requirements (Leffingwell 1997) and incomplete and changing requirements and specifications, are the root cause of many failed projects. To address these problems, researchers and practitioners have developed processes for identifying relevant stakeholders; discovering their needs, wants, and desires for the system; prioritizing and negotiating requirements; and specifying them in understandable, measurable, and testable ways (Robertson and Robertson 1999). In this article we focus particularly on recent efforts to use machine learning and recommender systems technologies for automating requirements engineering processes and enabling stakeholder and designer decision support.

In general, the task of a recommender system in any domain is to identify items of interest to a given user of which that user may otherwise be unaware. The recommendation problem is
Stakeholder Identification and Analysis

Stakeholder analysis is a critical component of every requirements engineering process. It involves identifying a set of stakeholders capable of providing a relatively complete and accurate description of the product under development. Although it is fairly straightforward to identify an initial group of stakeholders, it is far more difficult to make sure that all important stakeholder groups are represented by collaborative, responsible, authorized, committed, and knowledgeable stakeholders (Boehm and Turner 2004). There are two primary recommendation techniques that have been used effectively in this area. The first utilizes social networking and crowd sourcing to identify and recommend stakeholders for a project. The second is a hybrid recommender system that recommends stakeholders for particular discussion threads in online forums and that recommends discussion threads to individual stakeholders.

Discovering Project Stakeholders

Lim and others developed a recommender tool named Stakesource, which harnesses the power of crowd sourcing to discover project stakeholders (Lim, Quercia, and Finkelstein 2010a and 2010b). First an initial set of stakeholders is identified and asked to refer other stakeholders whom they consider important for the project. These stakeholders in turn are asked to refer additional people. The process continues until few additional stakeholders are discovered. At that time, a social network graph is constructed in which nodes represent candidate stakeholders and arcs represent referrals. Social network analysis techniques are then used to compute salience measures and to rank stakeholders according to their potential importance to the project.

Lim, Quercia, and Finkelstein also evaluated several different salience metrics; however, four of them stand out as most useful for ranking the potential importance of stakeholders. Betweenness centrality measures a stakeholder \( S \)'s ability to serve as a broker between other groups of stakeholders by summing the shortest paths between pairs of stakeholders that pass through \( S \) (Brandes 2001). Load centrality measures the influence stakeholder \( S \) has on the transfer of information between each pair of stakeholders. It is computed by passing one unit of information between each pair of stakeholders along the shortest path. When \( n \) paths are equal to the shortest length, the unit is divided by \( n \) and passed equally along each of the paths. Load centrality is then computed for each stakeholder \( S \) by summing the units, or fractions of units, that pass through it during stakeholder communication. These two metrics are used to identify core

typically formulated as a prediction task in which a predictive model is built according to prior observations and then used in conjunction with the dynamic profile of a new user to predict the level of interest by that user on a target item. Recommender systems generally fall into the three broad categories of content-based systems, which make recommendations based on content similarity between user profiles and the semantic descriptions of items (Pazzani and Billsus 2007); collaborative-filtering systems, which make recommendations by examining implicitly or explicitly derived preferences of a user on items and identifying other users with similar interests (Schafer et al. 2007); and knowledge-based systems, which make recommendations based on heuristic rules that depend on prior knowledge of the users or the domain (Felfernig et al. 2006). The output of a recommender system is usually a set of items that were previously unknown to the user and that scored the highest predicted interest values.

Maalej and Thurimella (2009) outlined a preliminary research agenda for recommender systems within the requirements engineering domain. They proposed several ways in which recommender technologies could help stakeholders create, analyze, specify, and manage requirements. These included using a recommender system to identify relevant background information for a given requirement, to recommend suitable templates for a given task, to suggest opportunities for saving time and effort by reusing requirements, for identifying pertinent design rationales related to specific types of requirements, and for proposing traceability relationships between various requirements and design elements. Although many of these proposed methods are based on heuristic approaches, more recent work in the area has focused on recommender systems that go beyond heuristic rules or search mechanisms and focus on more traditional machine-learning techniques.

Recommender systems have been used in the requirements engineering domain to address three specific kinds of problems. The first is to identify potential stakeholders for a given project. The second is to discover user requirements or features for a system, and the third is to provide support for requirements-related decision making such as requirements prioritization. Although current work is still in the early experimental stage, various approaches have already been evaluated on large-scale simulations using data from open-source forums, reenactments of large-scale projects with thousands of stakeholders, and small-scale proof-of-concept software development projects. Results from these studies show significant promise for transference to real industrial projects in the future.
project stakeholders. Two other metrics are used to identify and recommend stakeholders who are significantly influenced by a project. These are PageRank and in-degree metrics, both of which measure the degree to which a stakeholder is authoritative (locally in the case of in-degree and globally in the case of PageRank). Salience measures are computed for each stakeholder, and recommendations are made for any stakeholders scoring above a certain salience threshold.

### Recommending Stakeholders in Forums

In distributed projects, it is often infeasible to bring stakeholders together on a regular basis, and as a result, numerous organizations are adopting the practice of using wikis and forums to gather an initial set of feature requests and then to extract a more formal set of requirements from those requests (Cleland-Huang et al. 2009). Although such tools allow more stakeholders to engage in the requirements elicitation process, they also introduce challenges related to information overload, noisy and redundant data, incomplete discussions, dissenting opinions, and frustrated stakeholders who often receive no feedback for their contributions. Fortunately, many of these problems can be addressed through using data-mining techniques to keep feature requests organized in cohesive and nonredundant threads (Cleland-Huang et al. 2009) and also through using recommender systems to keep stakeholders informed of interesting topics and to help project managers find appropriate stakeholders to engage in open discussions.

The organizer and promoter of collaborative ideas (OPCI) approach, developed by Castro-Herrera, Cleland-Huang, and Mobasher (2009) and Castro-Herrera, Duan, Cleland-Huang, and Mobasher (2009) is depicted in figure 1. While OPCI can be used to support any online forum, it was originally designed to support the requirements engineering process (Castro-Herrera et al. 2009). OPCI’s recommender system serves two purposes: (1) recommending specific topics to individual stakeholders in order fully to engage each stakeholder in the requirements elicitation process, and (2) recommending stakeholders for specific topics in order to ensure that each topic has sufficient coverage.

In OPCI, the initial feature requests are represented as weighted term vectors. Specifically, each feature request, $x$, is represented as a vector of terms $(x_1, \ldots, x_W)$ where $x_i$ is the normalized weight associated with term $t_i$ in $x$, and $W$ represents the total number of unique terms in the entire set of feature requests. This normalized weight is calculated using a standard information-retrieval technique known as tf-idf, where $tf$ represents the frequency of term $t$ in the feature request and $idf$ represents the inverse document frequency. The Cosine similarity between two feature request vectors $a = (a_1, \ldots, a_W)$ and $b = (b_1, \ldots, b_W)$ is then computed as:
Table 1. Users’ Profiles with Additional Known Data.

<table>
<thead>
<tr>
<th>Users</th>
<th>Known Data</th>
<th>Threads</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K₁</td>
<td>K₂</td>
</tr>
<tr>
<td>U₁</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>U₂</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>U₃</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U₄</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ sim(a, b) = \frac{\sum_{i=1}^{W} a_i \cdot b_i}{\sqrt{\sum_{i=1}^{W} a_i^2 \cdot \sum_{i=1}^{W} b_i^2}}. \]

Intuitively this formula returns high similarity scores between two feature requests that share a significant number of relatively rare terms.

OPCI uses a consensus-based clustering approach to merge similar discussion threads and to extract topics that are dispersed across multiple threads into new discussion themes (Castro-Herrera et al. 2009; Duan, Cleland-Huang, and Mobasher 2008). Stakeholders who have contributed ideas to the initial threads or the reconstructed threads are assumed to be interested in the topics of those threads.

An initial stakeholder profile is created for each stakeholder according to the discussion threads that the stakeholder participates in, and also according to known information such as the role of the stakeholder in the project or the stakeholder’s subject matter expertise. These profiles form a profiles matrix \( M := (m_{ij})_{U \times K} \), where \( U \) is the number of stakeholders, \( T \) is the number of discussion threads, and \( M[i,j] \) is the weight representing the stakeholder’s interest in the discussion thread (derived as part of the clustering process that generates the initial discussion forums). The profiles are used to feed a collaborative recommender system (Schafer et al. 2007) based on the standard \( k \)-nearest-neighbor (\( k \)NN) strategy (Castro-Herrera et al. 2009). It computes similarities among stakeholder profiles in order to create a “neighborhood” for a targeted stakeholder. The resulting neighborhood is then used to predict the interest that the targeted stakeholder will have in any particular topic. These recommendations serve to foster the cross-pollination of ideas, prevent the problem of missed or stagnant discussions, and promote topics that address key areas of the system that have so far not been explored.

Two variations of the standard \( k \)NN algorithm were explored (Castro-Herrera, Cleland-Huang, and Mobasher 2009). In the first variation, user profiles were modeled using a binary matrix in which a 1 represented membership in a thread, and a 0 represented nonmembership. This representation was shown experimentally to be a good fit for the forum environment since users do not normally specify degrees of interest in forums or posts. The second variation involved extending the profiles with features corresponding to semantic knowledge about the domain or prior knowledge about the users. Examples of additional known data available for the requirements elicitation domain include: roles that the stakeholders play in the project, interests in key features of the system, and interests in cross-cutting concerns such as security or usability. Formally, the profiles matrix \( M := (m_{ij})_{U \times (K+T)} \) now includes \( K \) columns that capture a user’s interest in a known metadata element. Table 1 shows a schematic example of the enhanced matrix.

The binary recommender was shown to perform relatively well, and in all four of the datasets demonstrated the ability to recommend back approximately 25 percent to 50 percent of the known interests in the top 10 recommendations (Castro-Herrera, Cleland-Huang, and Mobasher 2009). The experiments also clearly demonstrated the usefulness of incorporating prior knowledge into the users’ profiles in order to enrich knowledge of the users’ interests in early stages of the forum. Recommendations that incorporated known data were able to recommend back approximately 70 percent to 90 percent of known interests in the top 10 recommendations.

Figure 2 shows hit ratio values for four different approaches including standard \( k \)NN, binary representation, binary with known data, and the baseline case in which items are recommended in random order. Hit ratio represents the probability that a recommended item among the top \( n \) items is selected by the user. The railway data set represents more than 1600 feature requests created by extracting requirements from the public specifications of two large-scale railway systems, the Canadian Rail Operating Rules and the Standard Code of Operating rules published by the Association of American Railroads. In this data set, known data was derived from the users’ individual roles in the project.

Recommender Systems in Requirements Elicitation

Recommender systems have also been used to recommend new ideas, features, and requirements during the initial domain analysis and elicitation process. Most projects include both a domain analysis phase, in which project stakeholders analyze related software systems to identify, organize, and represent features common to systems within
a domain (Arango and Prieto-Diaz 1989), and a focused requirements elicitation phase, in which the actual project requirements are elicited. Recommender systems have been used to support each of these individual tasks.

Recommending Requirements
Kumar, Ajmeri, and Ghaisas (2010), proposed a conversation-style recommender for use in agile development in which the system recommends a set of basic user stories (that is, brief descriptions of functionality) and tasks in response to an initial project description. The approach utilizes different ontologies that represent the environmental context, agile requirements, general requirements, and the problem domain. The authors provide an initial validation of their approach for an agile project and illustratively show that it generates useful recommendations. In related work, Romero-Mari-ona, Ziv, and Richardson (2008) developed an ad hoc knowledge-based system that takes initial goals of the system and recommends relevant models and requirements in the form of misuse cases and abuse scenarios, antimodels, and security-related goal models.

Recommending Features
The identification and timely reuse of domain assets in early stages of the requirements process can potentially reduce development costs, shorten time to market, improve quality of the delivered system, and increase product competitiveness. Existing domain analysis techniques primarily focus on either manual or semiautomated analyses of a small number of requirements specifications or product brochures in order to identify and extract reusable features. As a result, domain analysis tends to be quite labor intensive, dependent upon the expertise of available analysts or subject matter experts, and constrained by the availability of requirements specifications from previous related projects.

Dumitru et al. (2011) developed a novel recommender system that addresses these limitations by mining raw feature descriptions from online project repositories such as SoftPedia, clustering the descriptions into features, generating a feature model for a given product category, and then using this feature model to recommend features and combinations of features that might be relevant for a new product under development. The approach uses association rule mining (Agrawal, Imielinski, and Swami 1993) to identify affinities among product features and the kNN machine-learning strategy to identify similar products and make predictions about the existence of features not previously identified (figure 3).

More formally, once feature descriptions have been mined and clustered, a product-by-feature matrix is generated, $M := (m_{ij})_{P \times F}$, where $P$ represents the number of products and $F$ the number of identified features. To generate a recommendation for a new item, a new profile is created by parsing an initial product description and then using cosine similarity to match the terms in that description with features in the matrix. In many
cases, the initial product profiles may be rather sparse and unable to provide sufficient information to generate high-quality recommendations using more traditional techniques such as content-based or collaborative filtering.

Given a set of product profiles $P$, a set of features $F = \{F_1, F_2, \ldots, F_k\}$, and a feature set $f_s \subseteq F$, let $P_{f_s} \subseteq P$ be the set of products that have all the features in $f_s$. The support of the feature set $f_s$ is defined as $\sigma(f_s) = |P_{f_s}| / |P|$. Feature sets that satisfy a predefined support threshold are referred to as frequent feature sets. An association rule $r$ is expressed in the form $X \Rightarrow Y(\sigma_r, \alpha_r)$, where $X$ and $Y$ are feature sets, $\sigma_r$ is the support of the feature set $X \cup Y$, and $\alpha_r$ is the confidence for the rule $r$ given by $\sigma_r(X \cup Y) / \sigma_r(Y)$.

After generating the association rules, new features can be recommended based on an initial product profile by finding all the matching rules. In order to reduce the search time, the frequent feature sets are stored in a directed acyclic graph, called a frequent itemset graph, or FIG (Mobasher et al. 2001; Sandvig, Mobasher, and Burke 2007). The graph is organized into levels from 0 to $l$, where $l$ is the maximum size among all discovered frequent feature sets. Each node at depth $d$ in the graph corresponds to a feature set $I$ of size $d$ and is linked to feature sets of size $d + 1$ that contain $I$ at the next level. Given a product profile with feature set $f_s$, a depth-first search is performed on the graph to level $|f_s|$ in the graph. If a match is found, then the children of the matching node are used to generate candidate recommendations; and each child of this node corresponds to a frequent itemset $f_s \cup \{r\}$. The feature $r$ will be added to the recommendation list if the confidence of the rule exceeds a prespecified minimum threshold. The same procedure will be repeated for all the subsets of the feature set $f$.

The association-based recommender tends to generate a small, yet relatively accurate set of features. These recommended features are automati-
cally added to the initial product profile. Then a standard \(k\)NN learning strategy is used to compute a feature-based similarity measure between the new product and each existing product. The top \(k\) most similar products are considered neighbors of the new product, and this information is used to recommend features that might be relevant to the current analysis. In general, prediction scores will be computed for each candidate feature, and the features with highest predictions will be recommended.

Figure 4 illustrates a feature recommendation scenario for an antivirus product. An initial product description is mapped to four features in the recommender's knowledge base related to spam detection, disk scans, virus definitions, and virus databases, which serve as seeds for generating feature recommendations. User feedback is fed back into the system and used to drive future recommendations.

**Decision Support**

The final emerging area of recommender systems provides support for decision making in the requirements engineering process. More specifically, it is designed to address the four areas of decision making identified by Regnell et al. (2001), namely evaluating the quality of requirements, selecting requirements for the next release, classifying requirements by type, and making estimation decisions. Felfernig et al. (2010) have developed a recommender system named IntelliReq, which provides support for individual and group decisions. IntelliReq is designed to capture individual opinions related to requirements qualities, estimations, and priorities and then to use utility functions and a recommender system to help resolve conflicts, recommend requirements relevant to the current task, highlight possible requirement dependencies, and finally to recommend ways in which to maximize agreement between stakeholders. Parts of these scenarios have already been integrated in the IntelliReq decision support environment utilizing collaborative filtering, content-based filtering, knowledge-based recommendation, and group-based recommendation algorithms. In its current form IntelliReq utilizes group recommendation algorithms to support requirements negotiation.

**Conclusions**

The ongoing shift in the requirements engineering domain toward large-scale, distributed, collaborative, and tool-supported environments offers numerous opportunities for supporting human-intensive tasks through the use of recommendation technologies. In this article we have described three active areas of recommender system research related to identifying stakeholders, generating requirements and features, and providing support for several decision-making tasks. The work reported in these areas shows significant promise for the effective use of recommender systems in the requirements engineering domain.

However, making recommendations as a natural part of a complex requirements engineering task is nontrivial and introduces numerous human-computer interaction factors that need to be taken into consideration. For example, we need to ensure that recommendations are timely and contextualized and, furthermore, that they actually help the requirements engineer perform the task more effectively.

In many respects it is harder to build a successful recommender system in the requirements engineering domain than it is in a more traditional e-commerce context. Whereas a user purchasing books online may be pleased with a recommendation for a single interesting book, a domain analyst...
may not find a recommender system useful or trustworthy unless the system can successfully recommend all the basic and commonly occurring features in a targeted product. In certain cases, the role of the recommender system therefore transitions from the goal of generating serendipitous ideas and suggestions of potentially interesting items to the goal of making the complete and accurate set of recommendations needed to support a complex requirements engineering task.

Despite these difficulties, the various applications of recommender systems described in this article have highlighted the potential usefulness in the requirements engineering domain. Future work is expected to develop advanced techniques, algorithms, and processes for utilizing recommender systems in the requirements engineering domain and to explore the usefulness of the various approaches within the context of real software projects.

Acknowledgments

The work described in this article was partially funded by grant III:0916852 from the National Science Foundation.

References


Bamshad Mobasher is a professor of computer science and the director of the Center for Web Intelligence at the School of Computing at DePaul University in Chicago. He has published extensively in the areas of web data mining, web personalization, recommender systems, and predictive user modeling. He has also served in leadership positions for numerous related conferences and workshops, including as general chair, program chair and steering committee member of the ACM International Conference on Recommender Systems. Mobasher has served as an associate editor for ACM Transactions on the Web and on the editorial boards of several other prominent computing journals, including User Modeling and User-Adapted Interaction and the Journal of Web Semantics.

Jane Cleland-Huang is an associate professor of software engineering and director of the Systems and Requirements Engineering Center at the School of Computing of DePaul University in Chicago. She also currently serves as North American director of the Center of Excellence for Software Traceability (coest.org). She publishes extensively in the areas of requirements engineering and software traceability. She is currently associate editor for IEEE Transactions on Software Engineering, IEEE Software magazine, and the Requirements Engineering Journal, and has also served in various organizational and program committee roles for conferences such as the International Conference on Software Engineering (ICSE) and the International Requirements Engineering Conference (RE).