INFORMATION MINING TO ENHANCE SHARED UNDERSTANDING IN COLLABORATIVE ARCHITECTURAL DESIGN

JU-HUNG LAN  
Department of Architecture, Kao Yuan Institute of Technology, Kaohsiung, Taiwan.  
jhlan@cc.kyit.edu.tw

AND

MAO-LIN CHIU  
Department of Architecture, National Cheng Kung University, Tainan, Taiwan.  
mc2p@mail.ncku.edu.tw

Abstract. This research focuses on how to enhance shared understanding in collaborative design from information point of view. The data mining techniques are applied to discover the design semantic patterns with information classification and association capabilities. A system prototype with an information visualization tool is developed to demonstrate the capability of enhancing shared understanding in collaborative design.

1. Introduction

Researches in collaborative design have been focused on exploring the capabilities of information technologies from synchronous design or asynchronous design points of view (Maher et al., 1993; Kalay, 1998; Mills, 1999; Morozumi et al., 2001; Jabi, 2003). However, through the use of collaborative design systems, designers have difficulty in dealing with huge amount of design information available in the systems. How to overcome this kind of information overload problem and benefit from it has become an important issue in this field (Chiu and Lan, 2003).

In this research, we are exploring the shared understanding issue from an asynchronous design point of view. We are focusing on how to enhance shared understandings with collaborative design information in a distributed design environment. The research assumption is that design collaborators can communicate only with collaborative design information. They cannot communicate with each other from a synchronous design point of view. A data mining approach to discover valuable patterns within collaborative design information is proposed to achieve the research intention.
2. The Difficulties of Shared Understanding

What is shared understanding? Instead of human-oriented viewpoints, the term of shared understanding in this research is defined as sharing realizations of the relationships and rationale among collaborative design communication information from information-oriented viewpoints. In other words, shared understanding is defined from human-to-information point of view, not from human-to-human point of view in this study. Therefore, how to make information with self-organizing and design-understanding capabilities is the key to enhancing shared understanding in a collaborative design environment.

2.1. THE ILL-STRUCTURED COMMUNICATION INFORMATION

Most communication data generated in a collaborative design system has ill-structured essence. For instance, typical communication data within an electronic whiteboard or an email system are shown in Figure 1. These data, sometimes with files attached of various formats, usually are sequential records in accordance with timing of information posted for a specific design situation. However, it is difficult for collaborators to find desired data addressing specific design issue and associate related design communication information while the amount of information is getting huge. Nor it is easier for collaborators to share understanding with these kinds of ill-structured design communication information.

![Figure 1. Typical design communication information.](image)

2.2. THE MISUNDERSTANDINGS OF COMMUNICATION INFORMATION

In an asynchronous design environment, different domains work together at different locations. They are relied on the communication tool to exchange design ideas,
argue each position, and compromise with each other. However, since the limitation of asynchronous design, design collaborators have to interpret the communication information by themselves and make design decisions according to their “best” understanding while time is urgent. Naturally, sometimes they might make some mistakes since different collaborators have different specialties and cognitions. There might be some misinterpretations or misunderstandings in the communication of information. Also, some important information might be ignored since there is no mechanism to make designers aware of which information should be taken into account.

3. Mining Design Semantic Patterns

To achieve the research intention, this paper proposes a data mining approach to discover design semantic patterns hidden within collaborative design communication information. Data mining or knowledge discovery in database (KDD) is to find previously unknown and potentially useful knowledge from vast amount of data. These are the processes of discovering meaningful patterns through database by using multi-disciplinary techniques (Adriaans and Zantinge, 1996; Mena, 1999).

3.1. COLLECTING DESIGN COMMUNICATION INFORMATION

For the application of data mining techniques, the research first collected design communication information by conducting collaborative design studio projects. The students are separated as different design teams. Each design team is divided into three sub-domains, including architecture design, construction design, and environmental design. The students are advised to generate design information from multiple design perspectives. All design teams are guided to set a team-oriented

Figure 2. The whiteboard system for recording communication information.
collaborative design process while conducting their design projects. During the collaborative design process, the team member can bring up specific design issues for group communication at a certain design stage. An electronic whiteboard system with backend database design is used to record the information generated by the team members, shown as Figure 2. The notion of collaborative design communication information collected in this research is shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Site Analysis</td>
<td>Climate</td>
<td>The climate is hot and humid. Issue of thermal performance must be taken care.</td>
<td>The average temperature is 24.1°C. Summer has more rain. The relative humidity is 78.1%.</td>
<td>The structure should consider to be exposed to the sun and rain.</td>
<td>The roof style, natural lighting and sun shading should be taken care of.</td>
</tr>
<tr>
<td>Elevation Design</td>
<td>Opening Design</td>
<td>To avoid energy consuming, design a covered porch at the east side.</td>
<td>Design sunshade, use natural ventilation instead of air conditioners.</td>
<td>Use insulation materials with low thermal capacity.</td>
<td>Design a porch. Use steel structure.</td>
</tr>
</tbody>
</table>

3.2. INFORMATION MINING

3.2.1. A 6W Approach

In general, most design communication information is not standardized and without keyword features. Even these can be categorized into some structured formats, but are still difficult for abstracting relations among communication information. However, when data were organized as information for communication purposes, a specific design situation related to participants, timing, places, tasks, or rationales occurred. Therefore, a 6W (who, what, when, where, why, how) approach is adopted to associate related collaborative communication information. The communication information then can be organized into sub-datasets by the 6W-based classification, i.e. who are involved, or when, why, what, where, and how information are occurred.

3.2.2. Information Classification Patterns

During the mining processes, the text mining techniques based on information retrieval (IR) theories are used to discover the frequently used design terminologies within textual communication information. A vector space model is adopted to represent the collected design communication information by the discovered design
terminologies based on previous study (Lan, 2003). These design terminologies then are used to analyze the design semantic patterns with information classification capabilities.

By clustering analysis, these design terminologies can be classified according to design process (when), design issue (what), design domain (who, where) or design decision (how, why), shown in Table 2. Besides, there is an underlying hierarchical structure in Table 2, which is composed of design process, design issue, design domain and design decision. This hierarchical structure actually represents a semi-semantic relationship among design terminologies. Therefore, the notion of design semantic patterns with information classification capability can be shown as in Figure 3.

TABLE 2. The clustering analysis of design terminologies

![Figure 3. The notion of design semantic patterns with classification capability.](image-url)
3.2.3. Information Association Patterns

Besides information classification patterns, the research adopts basket analysis of E-commerce theory to discover the information association patterns. The idea is to build item sets of design terminologies from a design-issue-oriented point of view, as shown in Figure 4. These item sets are just like a series of transaction records of E-commerce. The basket analysis then can be applied to find the association rules among design terminologies. Typical association rule with confidence ratio (C) and support (S) ratio is showed as: \{Terminology 1\} \rightarrow \{Terminology 2\}, C=0.xx, S=0.yy (Agrawal et al., 1993). Figure 5 shows the results of association rule analysis by database techniques. Each number in \{Terminology Set 1\} or \{Terminology Set 2\} represents specific design terminology. While the \{Terminology Set 1\} at left-hand side appears, the association rule shows the \{Terminology Set 2\} at right-hand side will also appear with specific associative ratios. These association rules provide a computational model for calculating the design semantic relationships among design terminologies.

![Figure 4. Terminology Sets for Association Rule Analysis.](image1)

![Figure 5. The Results of Association Rule Analysis.](image2)

4. System Prototype with Shared Understanding

The design semantic patterns discovered in this research not only provide an effective mechanism for self-organizing design communication information, but also a design-understanding support for designers to realize the design rationales among design communication information, which correspond with the requirements of shared understanding that this research defined. To illustrate the capability of enhancing shared understanding in collaborative design, a system prototype with an information visualization tool, named as DECADE (a Distributed Environment for Collaborative Architectural Design), is developed in this research based on JAVA and web technologies.
4.1. DESIGN INFORMATION MANAGEMENT

Based on the 6W approach and the design semantic patterns discovered, a design information management tool with information classification and association capabilities is developed in DECADE. This tool provides a 6W-based information classification and association functions. The interface of design information management tool of DECADE is shown as Figure 6. The collaborative design information in DECADE is categorized according to design projects (what), project managers (who), team members (who), drawings (what), design discussions (who, when, what, why, how, where), meeting records (who, when, what, why, how, where) and message boards (Who, When, Why, What). The design participant can access different categories of design information by clicking a specific hyperlink. Once specific design information is retrieved, the underlying searching mechanism will assist users in traversing related information according to the network-like design semantic patterns this research discovered.

![Figure 6. The Design Information Management Tool of DECADE.](image)

4.2. INFORMATION VISUALIZATION

Traditional information management systems use keywords to search available information in database. The result of search is often listed with detailed, separated and much undesired information. However, this kind of searching mechanism has difficulties in dealing with vast amount of information (Chiang et al., 2001). It is
hard for designers to find related design information regarding a specific design situation. On the other hand, although the information management tool of DECADE provides information classification and association capabilities, the design participants still find it difficult to figure out the relationships among collaborative design information. This is because traditional graphical user interface has difficulty in effectively depicting information relationships with multi-W features at the same time. For instance, design communication messages in a discussion board usually are regarded with 3W (Who, When, What) or even more complicated 6W features. It is difficult for designers to understand a long list of 6W-related information at the same time by traditional user interface design. On the contrary, based on our previous study, a visual interface of linking design information shows great potential for discovering associative and meaningful information (Chiu and Lan, 2003). These associative links enable users to connect information into a network of logical association.

In this research, we develop an information visualization tool for associating related design information based on the design semantic patterns discovered. This tool provides a design-understanding support for designers to realize the design rationales among collaborative design communication information by a visual interface, as depicted in Figure 7. This visualization tool could be accessed from project-oriented, process-oriented, issue-oriented or designer-oriented point of views. To reduce the information overload problem, there is a threshold setting to show three levels of related design information with different colours and associative ratios at the same time. The designer can retrieve related information by clicking a specific information node in the visual interface. This network-like visualization interface provides visual effects to assist design collaborators in filtering and retrieving desired information. It is also helpful for design collaborator to become

Figure 7. The Information Visualization Tool of DECADE.
aware of hidden related information and have a better understanding of collaborative design communication information.

4.3. SHARED UNDERSTANDING EVALUATION

To evaluate the shared understanding achievement in this research, here we assume an interior design situation regarding the noise problem issue. We are using a keyword-based search function developed in DECADE. The search result shows the design terms related with noise problem with different ratios. These design terms include landscape, air quality, noise, traffic, plants, etc., as shown in Figure 8. For this design situation, the terms such as noise and traffic related with noise problem are obvious. However, the terms such as landscape, air quality and plants related with noise problem are worth studying further. These hidden-related terms might provide useful design information for designer’s reference.

Figure 8. The search result regarding noise problem.

To find a reasonable explanation of the above situation, we are further examining the contents of design communication information in database. By looking into the contents of communication information, we find that noise problem addressed at various design stages, not only at interior design stage. For instance, during site analysis stage, normally a designer will explore the noise problem issue, and usually traffic is the key factor for consideration. Besides, many designers prefer to plant a row of trees to solve the noise problem arising from traffic by landscape design point of view. As to the air quality, some information contents tell us that indoor plants could improve the air quality. Figure 9 shows the notion of the overall analysis of noise problem mentioned earlier. The arrows indicate the associative levels among these design terms. As you can see, noise problem and noise are associative by only one level. It is the most obvious pattern that we can realize by common sense. The terms landscape and traffic are associative with noise problem by two levels. These are easily recognizable patterns for general architectural designers. As to the terms
landscape, plants and air quality, they are associative with noise problem by three or four levels. Even to professional designers, sometimes it is still difficult to understand the overall design rationales. With this kind of shared understanding evaluation, the research shows there is great potential to enhance shared understanding in collaborative architectural design based on the research findings.

Figure 9. Shared understanding with the noise problem.

5. Summary and Future Works

5.1. SUMMARY

In summary, the research focuses on how to enhance shared understanding in collaborative design from information point of view. The data mining techniques are applied to discover the design semantic patterns with information classification and association capabilities. A system prototype with an information visualization tool is developed to demonstrate the capability of enhancing shared understanding in collaborative design. The research findings raise the following issues for further studies.

5.2. FUTURE WORKS

Although the information visualization tool provides more user friendly and more effective mechanism for retrieving collaborative design information, it requires more sophisticated interface when the number of nodes or information is growing. For further studies, the links among information in regarding associative ratios can be highlighted by various colours, thickness, or representations. Also, there might be
a threshold setting designed for collaborators to choose how much information should appear on the screen at the same time.

The research currently evaluates the system performance of shared understanding from the information point of view. Although the evaluation result shows great potentials of shared understanding, it will be more persuasive to have quantitative proof from the human point of view. For further studies, the research will adopt the semantic differential (SD) method, introduced by Osgood (Osgood, 1976), to evaluate the shared understanding performance from human point of view.

Acknowledgements

This research is partially supported by National Science Council in Taiwan, NSC92-2626-E-244-001. The authors are grateful to this support.

References

Mena, J. 1999, Data Mining Your Website, Digital Press.
Osgood, C.E. 1976, Focus on meaning, Mouton.