C. M. Herr, N. Gu, S. Roudavski, M. A. Schnabel, *Circuit Bending, Breaking and Mending: Proceedings of the 16th International Conference on Computer-Aided Architectural Design Research in Asia,* 535–543.

TOPOLOGY PATTERN MINING

A visual approach for representing and retrieving design patterns of spatial topology in a case library

CHIEH-JEN LIN
Tainan University of Technology, Yongkang, Tainan City, Taiwan, t60011@mail.tut.edu.tw

Abstract. This paper aims to apply the clustering analysis algorithm to analyze and classify the pattern of spatial topology of floor plans within a case library named “Open Case Study (OCS)”. Based on the results of classifications, this paper proposes a visual interface named “Topology Pattern Mining (TPM)” to present and rank searching results in response to user’s queries. The purpose of TPM is to extend the capacity of OCS for representing implicit knowledge of spatial topology. TPM can retrieve and classify design patterns of spatial topology, and thereby helps users to learn the design knowledge within relevant cases.

Keywords. Case-based design; case library; knowledge representation; spatial topology; data mining.

1. Introduction

Architectural design cases are important sources for design thinking while cases often lack of visual retrieval mechanism for indexing implicit knowledge. One purpose for developing an architectural design case library is to accumulate the design knowledge and experience obtained from case studies (Kolodner and Leake, 1996). However, the means for abstracting, encoding, and indexing the case information determine the way for how to retrieve, represent, and reuse the knowledge in the design cases.

Currently, an architectural deign case library is usually constructed by rational database technology, which requires selecting the most specific common features of a case and then converting those features into metadata of design information. The selected explicit features serve as the indexing mechanism and knowledge representation of the design cases in the library. However, the indexing mechanism using the explicit features often overlooks the acquisition, re-indexing, and generalization of the implicit knowledge among different cases.

This paper is a follow-up study of a previous project named “Spatial Topology Retrieval (STR).” STR established a visual and interactive tool for encoding and retrieving spatial topology of floor plans within a case library (Lin and Chiu, 2010). The case library is another previous project named “Open Case Study (OCS)”, which is an online house design case library with ontology-based authoring tools of metadata for case’s features (Lin and Chiu, 2009). This paper introduces clustering analysis algorithm to analyze and classify the pattern of spatial topology of floor plans within the case library. Based on the results of classifications, this paper proposes a visual interface named “Topology Pattern Mining (TPM)” to present and rank searching results in response to user’s queries. The purpose of TPM is to extend the capacity of OCS’s knowledge representation, which can not only store spatial topological information, but also retrieve and classify design patterns, which thereby helps users learn the design knowledge within relevant cases.

2. The approach for mining spatial topology pattern

The nature of architectural design is to appropriately create and arrange topological and geometric relations between vacant spaces and physical components of a building. Eastman (1999) proposed the building product models, which categorized a building’s information into three types – geometric, semantic, and topological information – and initiated the researches of building information model (BIM). Nowadays, BIM is already applied in commercial CAAD tools, and is widely adopted by AEC application field. However, most implementations of BIM are more aligned to document generation and final construction application (Howell and Batcheler, 2005). Consequently, those implementations all lack the topological information of the indoor or outdoor spaces, which is required for the early architectural design and performance analysis.

In the early design phase, an architect usually first considers the “topological relationship” among vacant spaces of a building program, rather than detailed geometric properties of physical components. For example, an architect may consider the adjacency among relevant rooms, the circulation formed by the connections of accesses and corridors among rooms, and directions of views formed by the opening, etc. From the view of case-based design, the topological knowledge for how to organize relevant spaces of a building project has been stored in the case itself. However, how to index, retrieve and represent those design knowledge still remains as a problem for developing a case library. For solving this problem, a case library needs (1) methods for encoding spatial topology of floor plans, (2) algorithm for weighting similarity of spatial topology, (3) algorithm for clustering similar spatial topology, and (4) means for representing the results of classification.

2.1. ENCODING SPATIAL TOPOLOGY

The data model, which represents and reasons topological relationships among spatial features, is an important issue in geographic information system (GIS), but it is usually ignored in CAAD research. Basic topological relationships within GIS include adjacency, overlapping, disjointness, inclusion, etc. (Rigaux, Scholl and Voisard, 2002). However, not all topologies are useful for architectural design, such as overlapping and inclusion. Other relations such as the accessibility, which is formed by accesses between interior spaces, are more important than basic topologies in GIS.

In the previous study, STR therefore modelled three manipulatible and two detectable spatial topology for encoding spatial topology. The three manipulatible spatial topologies are: (1) adjacency, (2) connection, and (3) combination. The two detectable spatial topologies are (4) opening, and (5) spatial orientation (Figure 1).

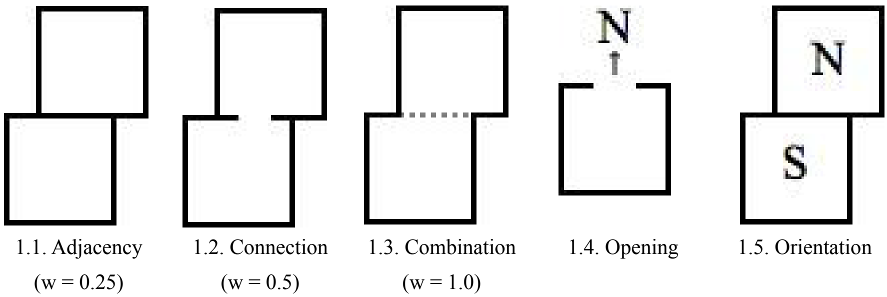


Figure 1. Five spatial topologies in STR.

STR modelled adjacent topology of rectangular spaces as a Boolean type relation, which means that two spaces must be adjacent when assigned an adjacent topology between them, but spaces could be adjacent even without any assigned topology (Figure 1.1). Other two manipulatible topologies are extended form adjacent topology. The connective topology, which is built by an access connecting two adjacent spaces, can represent the interior circulation (Figure 1.2). The combining topology, which is built by removing the boundary between two adjacent spaces, can represent a compound space such as a living room combining with a dinner room (Figure 1.3).

The opening topology, which is built by attaching an opening, can represent the relation of an interior space with outdoor spaces for the view and natural lighting (Fig. 1.4). The orientating topology, which is determined by relative positions of two adjacent spaces, can represent the functional requirements such as a response to the climate or context of the site (Fig. 1.5). However, to avoid complicated inputs in encoding spatial topology, only the first three topologies are manipulatible in STR. The opening topology could only be assigned and detected, and STR automatically detects orientating of opening and adjacent spaces like the early system such as FABEL (Coulon, 1995).

Five topologies may be all useful for present some important knowledge for spatial allocations knowledge in architectural design. However, only first three topologies, i.e. adjacency, connecting and combination, are more important for identifying and weighting relevance of two different indoor spaces, then further to weighting the similarity of spatial allocations among different cases.

2.2. WEIGHTING SIMILARITY OF SPATIAL TOPOLOGY

For weighting the similarity of spatial topology of design cases, TPM assigned different weight to three kinds of topologies, which are: (1) adjacency is 0.25, (2) connecting is 0.5 and (3) combination is 1 (Figure 1.1, 1.2, 1.3). This weighting method can present the relevance of two spaces and keep the scores of similarities among cases less than 1. Every one spatial topology can be regarded as one dimension of a multi-dimensional space. A weighting of the spatial topology can be regarded as a coordinate along the topological dimension. Then distance metrics can be applied to measure the distance between two topological nodes in the multi-dimensional space.

Cosine similarity is a kind of distance metrics applied in data mining, that can measure distance between two vectors by finding the cosine of the angle between them (Tan, Steinbach and Kumar, 2005). For example, there are two cases C_i and C_j, where $i \neq j$, and all cases has n types of spaces in the library. Therefore, the topological weightings of floor plans of case C_i and C_j can be converted into vector T_i and T_j as formula (1). Then cosine similarity algorithm can measure the similarity score of two cases C_i and C_j by formula (2):

$$T_i = [t_{i.1}, t_{i.2}, \dots t_{i.n}],\ T_j = [t_{j.1}, t_{j.2}, \dots t_{j.n}] \quad (1)$$

$$\text{Similarity}(C_i, C_j) = \frac{T_i \bullet T_j}{\|T_i\|\|T_j\|} = \frac{\sum_{k=1}^{n} t_{i.k} \times t_{j.k}}{\sqrt{\sum_{k=1}^{n} (t_{i.k})^2} \times \sqrt{\sum_{k=1}^{n} (t_{j.k})^2}} \quad (2)$$

The higher score of Similarity(C_i, C_j) means that C_i and C_j are more similar in their spatial topologies. Base on the similarity scores of cases, we can further apply clustering analysis algorithm for classifying patterns of houses' plans in the OCS case library.

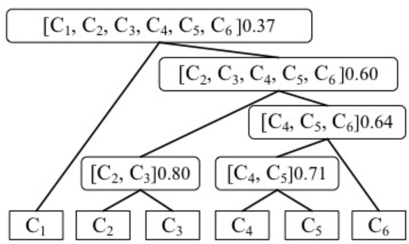
2.3. CLUSTERING SPATIAL TOPOLOGY PATTERNS

Clustering analysis algorithm is a kind of data mining algorithms, such as K-Means for partitional clustering and complete-linkage method (CLM) for hierarchical clustering, which is applied to automatically classify mass data. Since cosine similarity is applied to measure the similarities of spatial topologies of cases in the library, agglomerative hierarchical clustering method should be satisfied and easier for clustering the patterns by ranking the similarity scores among cases.

For example, there are 6 cases in the library and their similarity scores matrix is shown in Figure 2.1. Then, agglomerative hierarchical clustering method can build a binary tree by linking similar cases based on the max score of similarities. Therefore, if no any given threshold, there are 4 levels of 5 different clusters after agglomerated. In other words, there are 5 patterns found in this demonstration (Figure 2.2). Based on the clusters, the system cannot efficiently ranking similar cases of a given cases by user, nut also can discover design patterns.

	C_2	C_3	C_4	C_5	C_6
C_1	**0.37**	0.18	0.37	0.11	0.36
C_2		**0.80**	**0.60**	0.37	0.27
C_3			0.55	0.40	0.19
C_4				**0.71**	**0.64**
C_5					0.59

2.1. Similarity scores matrix of 6 cases.



2.2. Binary tree by similarity scores.

Figure 2. A demonstration of the agglomerative hierarchical clustering methods.

3. Representation of spatial topology patterns

"Design pattern" is a formal documenting of a solution to a design problem in a particular field of expertise. Therefore, "design pattern" seems to be a satisfied representation of design knowledge within a case library. However, Heylighen and Neuckermans (2001) pointed out that most of early case-based systems failed to require, re-index, and generalize design knowledge within the case libraries, and not to mention to represent design patterns of cases.

The purpose of TPM is to extend the capacity of OCS's knowledge representation, which can not only store spatial topological information, but also retrieve and classify design patterns of spatial topologies. For representing design patterns of spatial topologies, TPM proposes two approaches for visualizing similarities among floor plans of design cases and individual spaces in separate floor plans.

3.1. HYBRID REPRESENTATION OF FLOOR PLAN PATTERNS

How to display a multi-dimensional vector of data's features on a 2-D screen usually is a problem for visualizing ranking and clustering results in data-mining domain. However, the problem for TPM to display the ranking results of spatial topologies on 2-D screen should be only one metric of similarity scores.

Therefore, TPM takes a hybrid approach, which applies basic geometric features of floor plans, i.e. the differences of total areas of floor plans, as another dimension of similarity. By integrating basic geometric features with topological similarities of separate floor plans, TPM can rank relevant floor plans for retrieved cases, display results of ranking on 2-D screen but avoids confusion. Figure 3 demonstrates the concepts of STM's interface for visualizing retrieved patterns. TPM arranges relevant floor plans along Y axis by topological similarity scores and along X axis by differences of total areas. The distances of Y axis among cases are based the reciprocals of similarity scores. The distances of X axis among cases are based the different percentages of total areas. Therefore, if the retrieved case is C_6 of the demonstrations in Figure 2, then all the 6 cases may be displayed in TPM as Figure 3.

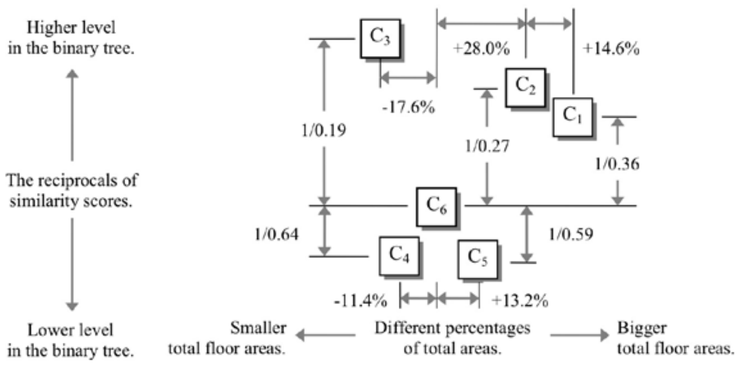


Figure 3. The concept of TPM for visualizing patterns of retrieved cases.

3.2. PARTIAL REPRESENTATION OF SPATIAL TOPOLOGY PATTERNS IN A FLOOR PLAN

The similarity scores of spatial topologies in TPM are measured by all topologies in floor plans of all cases. Therefore, the similarity scores of two floor plans cannot easily ranking partial similarities or differences between those plans. However, since the metric of every type of spatial topology is cumulated in the system, the ranking of topological weighting among different types of spaces can effectively reduce permutations of different types of spaces. Therefore, it is possible to implement an interface for partial representing of spatial topologies and for retrieving relevant floor plans by the ranking of topological weighting among different types of spaces.
Figure 4 demonstrates the prototype of TPM's interface for retrieving floor plans by partial spatial topologies. However, TPM can only map the topological weightings of all spaces in the query then retrieves the floor plans of cases in the library, and cannot immediately measure similarity score of a new query with all floor plan in the library.

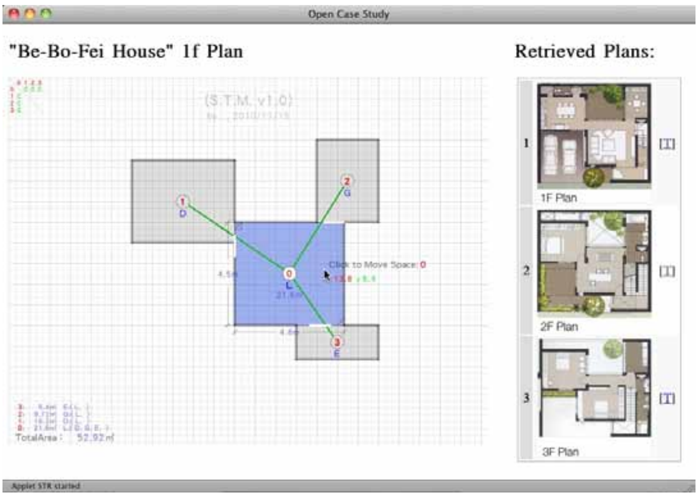


Figure 4. The prototype of STM's interface for retrieving cases by partial spatial topologies.

3.3. THE IMPLEMENTATION OF TPM

OCS is an online case library of house design cases, which is developed in MySQL rational database, PHP script language for server-side data accessing, and JavaScript for client-side interface. TPM extends the abilities of STR from encoding and retrieving spatial topology of individual floor plans, to weighting and ranking similarities among different floor plans.

Just like STR, TPM is developed in Processing - a simplified software IDE developed by MIT (Reas and Fry, 2007). Processing is based on JAVA programming language, but aims to teach fundamentals of the computer programming in the visual art and design education. Via a visual, interactive, Internet-ready interface and the MySQL database connective capabilities of Processing, TPM thereby serves as a visual and interactive tool for retrieving and representing the design patterns of spatial topologies in floor plans of cases within the OCS library.

4. Conclusions

Alexander (1977) proposed the idea of "pattern language" for describing and organizing good design practices of architectural design, and inspired the concept of "design pattern" for many complex engineering tasks such as software engineering. Therefore, for representing design solutions and its problem, a case library can also be an architectural design pattern library. However, because lack of generalizing means for design patterns, the only representation of design pattern in most of case libraries is the case itself. As a result, no case library can provide a mechanism for acquiring then retrieving design patterns.

TPM demonstrates a data-mining approach for detecting the topological features of floor plans, then derive out the patterns of spatial topologies automatically. By applying data-mining algorithms to weight similarity among floor plans of cases, TPM can cluster similar floor plans in order to reveal the patterns of spatial topologies. TPM implements a visual retrieval interface for displaying relevant floor plans of cases based on topological similarity scores and basic geometric features. By applying searching and mapping functions of database technology, TPM can map partial topological layout as a query for retrieving similar floor plans from the library.

However, the data-mining approach is a method that requires a large quantity of calculation. When a new permutation of spatial topologies appears, it needs time to re-measured similarity scores among all cases. Therefore, efficient algorithm for weighting and clustering complex building types will become a challenge for this approach. On the other hand, how to integrate more topological features, such as separating, opening and orientation, and geometric features, such as shape, dimensions and area of individual spaces in the floor plans, into the weighting and clustering algorithm still needs more investigations.

Acknowledgements

This research is supported by the National Science Council of Taiwan under grant number NSC 99-2221-E-165-002.

References

Coulon, C.H.: 1995, Automatic Indexing, Retrieval and Reuse of Topologies in Architectural Layouts, *CAADFutures 1995*, Singapore, 557-586.

Eastman, C. M.: 1999, *Building Product Models: Computer Environments Supporting Design and Construction*, CRC Press, Boca Raton.

Heylighen, A. and Neuckermans, H.: 2001, A case base of Case-Based Design tools for architecture, *Computer-Aided Design*, **33**(14), 1111-1122.

Howell, I., and Batcheler, B.: 2005, "Building Information Modelling Two Years Later – Huge Potential, Some Success and Several Limitations Retrieved". Available from: LaiserinLetter <http://www.laiserin.com/features/bim/newforma_bim.pdf> (accessed 1 November 2010).

Kolodner, J.L. and Leake, D.B.: 1996, A tutorial introduction to case-based reasoning, in Leake, D.B. (ed.), *Case-Based Reasoning: Experiences, Lessons, and Future Directions*, MIT Press, Cambridge, 31-65.

Lin, C.J. and Chiu, M.L: 2010, Open case study, *CAADRIA09*, Yunlin, Taiwan, 393-399.

Lin, C.J. and Chiu, M.L: 2010, Spatial topology retrieval: a visual approach for representing and retrieving spatial topology in a case library, *CAADRIA10*, Hong Kong, 147-154.

Reas, C. and Fry, B.: 2007, *Processing: A Programming Handbook for Visual Designers and Artists*, MIT Press, Cambridge.

Rigaux, P, Scholl, M. and Voisard, A.: 2002, *Spatial Databases: With Application to GIS*, Morgan Kaufmann, San Francisco.

Tan, P.N., Steinbach, M. and Kumar, V.: 2005, *Introduction to Data Mining*, Addison-Wesley, Boston.