Topology Pattern Mining: A Visual Approach for Detecting and Retrieving Design Patterns of Spatial Topology in a Case Library

Chieh-Jen Lin

Tainan University of Technology, 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, this paper proposes a visual interface named "Topology Pattern Mining (TPM)" to present and rank search results in response to users' queries. The purpose of TPM is to extend the capacity of OCS for representing the implicit knowledge of spatial topology. TPM can retrieve and classify design patterns of spatial topology, which thereby helps designers as well as participants to select appropriate proposals.

Keywords: case-based design, knowledge definition, spatial topology, data mining.
DOI: 10.3722/cadaps.2012.199-205

1 INTRODUCTION

Alexander proposed "a Pattern Language" for capturing perennial solutions of recurring problems within physical construction contexts [1]. His work inspired the idea of "design pattern" for formal documenting good practices of a solution to a design problem in a special expertise such as software engineering. In the field of architectural design, important precedents in textbooks or practices are important sources for acquiring design patterns. One reason for the development of an architectural design case library is the collection of the design knowledge and experience obtained from case studies[7]. Design patterns should be one of important knowledge within a design case library; however, design patterns are usually implicit and most of case libraries lack a feasible mechanism for automatically generalizing design patterns.

Currently, an architectural deign case library is usually constructed by applying 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 for the library. However, this indexing mechanism using 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[9]. 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[8]. 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 users' 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 designers as well as participants to select appropriate cases from the library.

2 THE APPROACH FOR MINING SPATIAL TOPOLOGY PATTERN

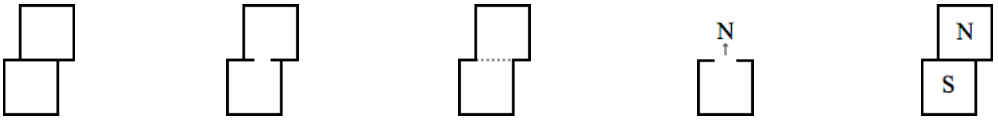
The nature of architectural design is to appropriately create and arrange topological and geometric relations between vacant spaces and physical components of a building. Eastman proposed the building product models [3], which categorized a building's information into three types – geometric, semantic, and topological information – and initiated research on building information model (BIM). Presently, BIM is already applied in commercial CAAD tools, and is widely adopted by the AEC application field. However, most implementations of BIM are more aligned to document generation and final construction application [5]. Consequently, those implementations all lack the topological information of indoor or outdoor spaces, which is required for 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 instead of the 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 is stored in the case itself. However, how to index, retrieve, and represent this design knowledge still remains a problem for the development of a case library. To solve this problem, a case library needs: (1) methods for encoding spatial topology of floor plans; (2) an algorithm for weighting the similarities of spatial topologies; (3) an algorithm for clustering similar spatial topologies, and; (4) a 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 sys-tem (GIS), but it is usually ignored in CAAD research. Basic topological relationships within GIS include adjacency, overlapping, disjointness, inclusion, etc. [11]. 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 modeled three manipulatable and two detectable spatial topologies for encoding spatial topology[9]. The three manipulatable spatial topologies are: (1) adjacency; (2) connection; and (3) combination. The two detectable spatial topologies are: (4) opening and (5) spatial orientation (Fig. 1).



(a) Adjacency (w = 0.25) (b) Connection (w = 0.5) (c) Combination (w = 1.0) (d) Opening (e) Orientation

Fig. 1: Five spatial topologies in STR.

STR modeled the adjacent topology of rectangular spaces as a Boolean type relation, which means that the two spaces must be adjacent when assigned an adjacent topology between them, but spaces could be adjacent even without any assigned topology (Fig. 1(a), 1(b), 1(c)). The other two manipulatable topologies are extended from adjacent topology. The connective topology, which is built by an access connecting two adjacent spaces, can represent the interior circulation (Fig. 1(b)). The combining topology, which is built by removing the boundary between two adjacent spaces, can represent a compound space; such as a living room joined with a dining room (Fig. 1(c)).

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 responses to climate or context of the site (Fig. 1-5). However, to avoid complicated inputs in encoding spatial topology, only the first three topologies are manipulatable in STR. The opening topology can only be assigned and detected, and STR automatically detects orientation of opening and adjacent spaces, in a manner similar to early systems, such as FABEL [2].

Although all five topologies may be useful in presenting important information for spatial allocations knowledge in architectural design, only the first three (adjacency, connecting and combination) are more important for identifying and weighting the relevance of two different indoor spaces.

2.2 Weighting Similarity of Spatial Topology

For weighting the similarity of spatial topology of design cases, TPM assigned different weights to three kinds of topologies, which are: (1) adjacency is 0.25; (2) connecting is 0.5; and (3) combination is 1 (Fig. 1(a), 1(b), 1(c)). This weighting method can present the relevance of two spaces and keep the scores of similarities among cases to value less than 1. Each spatial topology can be regarded as single dimension of a multidimensional space. A weighting of the spatial topology can be regarded as a coordinate along the topological dimension. Next, distance metrics can be applied to measure the distance between two topological nodes in the multidimensional space.

Cosine similarity is a kind of distance metrics applied in data mining that can measure the distance between two vectors by finding the cosine of the angle between them [12]. For example, if there are two cases Ci and Cj , where $i \neq j$, and all cases have n types of spaces in the library, then the topological weightings of floor plans of cases Ci and Cj can be converted into vectors Ti and Tj using formula Eqn. (2):

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

The cosine similarity algorithm can be used to measure the similarity score of two cases Ci and Cj by application of the following formula Eqn. (3):

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

The higher the score of Similarity(Ci, Cj) between Ci and Cj, the more similar they are in spatial topologies Based on the similarity scores of cases, we can further apply clustering analysis algorithm to classify patterns of house plans in the OCS case library.

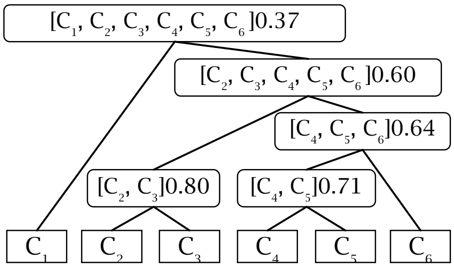
2.3 Clustering Spatial Topology Patterns

Clustering analysis algorithm is a kind of data mining algorithm, 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, if there are six cases in the library and their similarity scores matrix is as shown in Fig. 4(a), agglomerative hierarchical clustering method can then be used to build a binary tree by linking similar cases based on the maximum score of similarities. Therefore, if there are no given thresholds, there will be four levels of five different clusters after agglomeration. In other words, there were five patterns found in this demonstration (Fig. 4(b)). Based on the clusters, the system cannot only efficiently rank similar cases among cases provided by users but can also 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

(a) Similarity scores matrix of 6 cases.



(b) Binary tree by similarity scores.

Fig. 4: A demonstration of the agglomerative hierarchical clustering methods.

3 REPRESENTATION OF SPATIAL TOPOLOGY PATTERNS

“Design pattern” is a formal documentation 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, scholars have pointed out that most of the early case-based systems have failed to require, re-index, and generalize design knowledge within the case libraries[4], 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

The display of a multidimensional vector representing the features of a specific data point on a two-dimensional screen is usually a problem encountered when trying to visualize ranking and clustering results in data-mining domain. However, the problem for TPM to display the ranking results of spatial topologies on a two-dimensional 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 an other dimension of similarity. By integrating basic geometric features with topological similarities of separate floor plans, TPM can rank relevant floor plans for retrieved cases and display results of ranking on a two-dimensional screen at the same time that it avoids confusion. Figure 3 demonstrates the concepts of TPM’s interface for visualizing retrieved patterns. TPM arranges relevant floor plans along the y-axis by topological similarity scores and along x-axis by differences of total areas. The distances between y-axis among cases are based on the reciprocals of similarity scores. The distances between x-axis among cases are based the different percentages of total areas. Therefore, if the retrieved case is C6 of the demonstrations in Figure 2, then all six cases may be displayed in TPM in the manner shown in Figure 5.

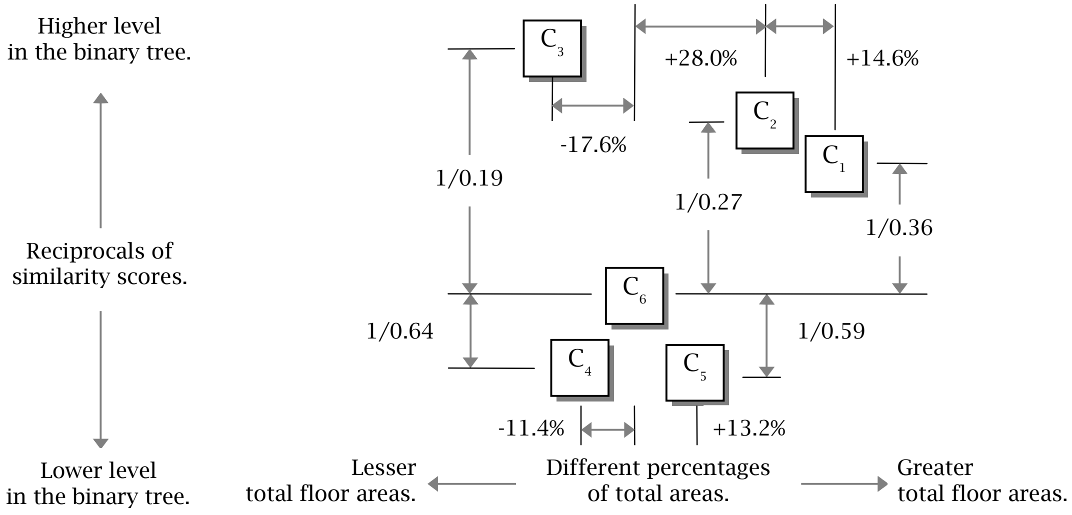


Fig. 5: The concept of TPM for visualizing patterns of retrieved cases.

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 be easily ranked according to partial similarities or differences between those plans. However, since the metric of every type of spatial topology is cumulated within 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 the partial representation of spatial topologies and for retrieving relevant floor plans according to the ranking of topological weighting among different types of spaces.

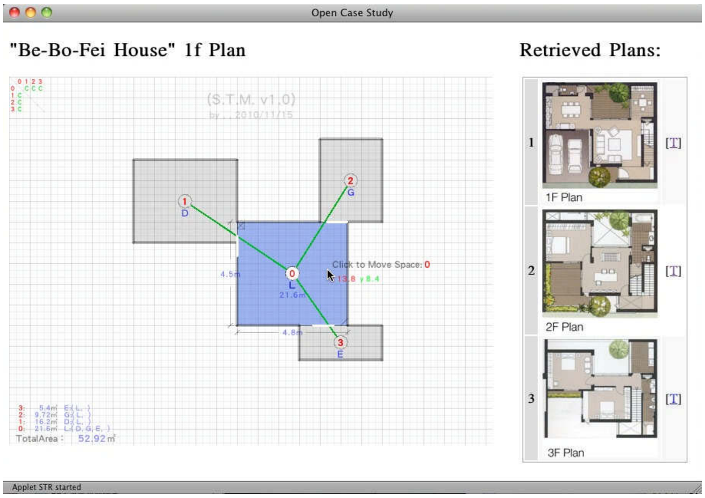


Fig. 6: The prototype of TPM's interface for retrieving cases by partial spatial topologies.

Fig. 6 demonstrates the prototype of TPM's interface for retrieving floor plans using partial spatial topologies. However, TPM can only map the topological weightings of all spaces in the query and retrieve the floor plans of cases in the library, but cannot immediately measure the similarity score of a new query with all the floor plans in the library.

3.3 The Implementation of TPM

OCS is an online case library of house design cases, which was developed in MySQL rational database, PHP script language for server-side data access, 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 using Processing, which is a simplified software IDE developed by MIT [10]. Processing is based on JAVA programming language but aims to teach fundamentals of computer programming in the fields of visual art and design education. Through the 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 DISCUSSIONS

4.1 Design Pattern vs. Design Situation

Good design solutions always relate with their situations. Alexander therefore documented design solutions with relevant building contexts for describing the situations of a design pattern. However, topological patterns, which are detected by applying data-mining algorithms to classify floor plans, cannot avoid losing their contexts with external situations. For associating external situations, topological patterns need more clues to build the relations with other features of the cases, such as site contexts, spatial functions, and building types of the buildings. Additional research is necessary in order to assist users in establishing these kinds of associations.

4.2 Spatial Topology vs. Spatial Semantics

The emerging generative grammar of computer science in 1970 inspired Alexander, who tried to hold a design pattern as a word, then organized 253 design patterns into a design language. Since Alexander's work was described by natural language, the semantic relations among components of a design patterns, their contexts, and different patterns are very clear and easy to understand. However, since spatial relations rather than natural language describe topological patterns, users may not easily be able to understand the significances of those patterns. Every topological relation between two spaces is actually a representation of some design concepts about their correlations. For example, the kitchen must serve the dining room, and the partition should prevent the exchange of pollution between the two spaces. Attaching more semantic portrayals to spatial topologies should allow designers to quickly acquire design knowledge. However, further investigations need to be done on how to assist users in organizing spatial topologies with their semantics.

4.3 Design Solution with Design Pattern

The selection of different design proposals is the major task of most design meetings. Since the contexts of design meetings may change each time, the selection of design proposals is usually a torturous process for both the architects and their clients. Although some design proposals may evolve and become the final project, most proposals usually died in the process, with nothing left behind [6]. A tool like TRM, which can reveal past design patterns, should prove to be very helpful for architects and their clients in selecting appropriate proposals and avoiding tedious and inefficient communications. However, as mentioned above, the design situations and semantic explanations are still the keys for successful representations of design knowledge in the patterns. Further investigation is necessary for improving TRM to assist in communications between architects and their clients.

5 CONCLUSIONS

Alexander proposed the idea of "pattern language" for describing and organizing good design practices of building 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, due to the lack of generalizing means for design patterns, the only representation of design pattern in most 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, which automatically derives the patterns of spatial topologies. By applying data-mining algorithms to weight similarity among floor plans of cases, TPM can cluster similar floor plans in order to reveal their 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 the searching and mapping functions of database technology, TPM can map the 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 calculations. When a new permutation of spatial topologies appears, it requires time to re-measure and regenerate similarity scores among all cases. Therefore, an efficient implementation for weighting and clustering complex building types is a challenge for this approach. Finally, further investigation needs to be conducted on 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.

5.1 Acknowledgements

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

REFERENCES

[1] Alexander, C.; Ishikawa, S.; Silverstein, M.: A Pattern Language: Towns, Buildings, Construction, Oxford University Press, New York, 1977.

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

[3] Eastman, C.-M.: Building Product Models: Computer Environments Supporting Design and Construction, CRC Press, Boca Raton, FLA, 1999.

[4] Heylighen, A.; Neuckermans H.: A case base of Case-Based Design tools for architecture, Computer-Aided Design, 33(14), 2001, 1111-1122. doi:10.1016/S0010-4485(01)00055-0

[5] Howell, I.; Batcheler B.: Building Information Modeling Two Years Later – Huge Potential, Some Success and Several Limitations, 2009, http://www.laiserin.com/features/bim/newforma_bim.pdf

[6] Ingels, B.: Yes Is More: An Archicomic on Architectural Evolution. Köln: Evergreen, 2010.

[7] Kolodner, J.-L.; Leake D.B.: A Tutorial Introduction to Case-Based Reasoning, In D. B. Leake (Ed.), Case-based reasoning: experiences, lessons & future directions. AAAI Press, Menlo Park, CA., 1996, 31-65.

[8] Lin, C.-J.; Chiu, M.-L.: Open Case Study. CAADRIA 2009, 2009, 393-399.

[9] Lin, C.-J.; Chiu, M.-L.: Spatial topology retrieval: a visual approach for representing and retrieving spatial topology in a case library, CAADRIA 2010, 2010, 147-154.

[10] Reas, C.; Fry B.: Processing: a programming handbook for visual designers and artists, MIT Press, Cambridge, Mass., 2007.

[11] Rigaux, P.; Scholl, M.-O.; Voisard, A.: Spatial Databases: With Application to GIS, Morgan Kaufmann, San Francisco, 2002.

[12] Tan, P.-N.; Steinbach M.; Kumar V.: Introduction to data mining. Pearson Addison Wesley, Boston, 2005.