MULTI-ASPECT APPROACH TO THE SIMILARITY EVALUATION OF BIM-BASED STANDARD DWELLING UNITS

Tianfeng He[1], Jianping Zhang[2], Jiarui Lin[3], and Yungui Li[4]

[1]Ph.D. candidate, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: homdyan@126.com

[2]Professor, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: zhangjp@tsinghua.edu.cn

[3]Postdoctoral Research Associate, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: lin611@tsinghua.edu.cn (Corresponding author)

[4]Researcher, China State Construction Engineering Corporation Limited, Beijing, China. Email: liyungui@china.com

ABSTRACT

By standardize various dwelling units of residential buildings, standard dwelling unit (SDU) is a way to reuse knowledge for residential design, especially in the floor-plan design of multi-floor buildings. Recently, building information model (BIM) has emerged to capture not only the geometric data but also attributes and topological data of SDUs. However, conventional attribute-based search methods cannot utilize the full potential of BIM's rich information. Therefore, this paper proposes a multi-aspect approach to the similarity evaluation of SDUs to support the similarity-based search. This approach introduces six similarities from aspects of the attribute, topology, and shape of an SDU. Furthermore, a combined similarity is proposed to combine different similarities in order to reflect designers' search intentions. Finally, a database with 170 SDUs and a prototype design tool are developed and tested in two projects. This application shows that the approach is capable of evaluating the similarity of SDUs in a more comprehensive way and increasing the speed and flexibility of the search for SDU alternatives, thereby improving the design efficiency and quality.

Keywords: Standard Dwelling Unit; Similarity Evaluation; Building Information Model

INTRODUCTION

Residential buildings are known for their high similarity compared to other types of buildings. Standard designs can be summarized from successful projects and reused in other projects to improve efficiency while reducing both cost and risk. A standard design for a dwelling unit is also known as a standard dwelling unit (SDU). Generally, SDUs act as basic components of the residential design. By replacing the original SDUs in a floor-plan with similar alternatives, the final design solution will be gradually improved. Most SDUs are derived from as-built projects and include rich information. Traditionally, this valuable information is stored in drawings and documents, making it difficult to reuse the information (Li et al. 2004). In recent years, with the increased popularity of the Building Information Model (BIM) in the residential design process, SDUs tend to be modeled and stored as BIMs, for example, as Industry Foundation Classes (IFC) files. Compared to documents and drawings, the BIM integrates various data into a single repository, rendering information extraction and utilization much more convenient (Eastman et al. 2011)

Despite the change in the storage of SDUs, the method of searching SDUs remains the same. Attribute-based methods continue to be prevalent in practice (Langenhan et al. 2013). However, it is difficult or even impossible to represent some important features of SDUs, such as topology and shape, as attributes. Furthermore, most attribute-based search methods provide only results that strictly match certain rules, making them too rigid for designers to find proper SDU alternatives. To address this problem, a similarity-based search method is chosen. The search results will be ordered by the similarity between SDUs. Replacing a similar SDU can prevent major design changes in floor-plan design. However, few related research studies address the similarity evaluation of SDUs (He et al. 2016).

This paper proposes a multi-aspect approach to the similarity evaluation of SDUs to support the similarity-based SDU search. This approach contains six similarities from aspects of the attribute, topology, and shape of an SDU. A combined similarity combining the six similarities in order to reflect designers' search intentions. Then, a prototype database has been established, where the combined similarity is applied. Performance analysis of the similarity evaluations is provided to prove their efficiency, and the combined similarity is used for design recommendation in floor-plan design to test its practicality.

RELATED WORK

SDU and BIM

SDUs and their related design methods have been used for decades by real-estate developers, design institutes, and governments. Traditionally, documents and drawings are used to store SDUs, which makes it difficult to reuse the information. BIM is able to store various types of information, and it is gradually replacing documents and drawings recently. IFC is the most commonly used neutral format in BIM and is an ideal information source for the similarity evaluation of SDUs (Wu and Hsieh 2007). Langenhan and Petzold (2010) used IFC as the information source to extract a building's fingerprint. An IFC-based SDU database was established by He et al. (2016) for residential design and information modeling.

Isikdag et al. (2013) distinguished among semantic, Euclidean, and topological information spacesinBIM.TakingtheSDUasanexample, theEuclideanspacecontainsgeometricpresentations of rooms and other elements of an SDU; the topological space contains the topologies of its accessibility and adjacency; the semantic space contains definitions of the elements of an SDU and their properties. Such distinction of spaces should be considered when evaluating the similarity of SDUs. However, the attribute-based search method is not capable of making full use of the information, especially that in Euclidean and topological spaces. Therefore, instead of attribute-based search method, the similarity-based search method is chosen, which can utilize the rich information of BIM in a more comprehensive way in residential design.

Related Similarity Theories

Attribute-Based Similarity

In general, attributes are classified into 4 types: nominal attributes, binary attributes, numeric attributes and ordinary attributes (Han et al. 2011). Each type has its own approach to calculating the degree of similarity. Generally, nominal and binary attributes use the ratio of matches to measure the degree of similarity; numeric attributes use the Minkowski distance or the cosine similarity; ordinal attributes are considered nominal attributes or numeric attributes. The methods for evaluating the similarities of the attributes of an SDU depend on their types and specific requirements.

Topological Similarity

The spatial topology of a building strongly affects its functions (Hillier 2007). Similarly, SDUs with similar topologies of rooms may have similar functions and serve similar purposes. The topology of an SDU, including accessibility and adjacency, can be represented by an undirected attributed graph using the Poincaré Duality. Therefore, the topological similarity of SDUs can be evaluated by the similarity of their corresponding graphs.

Evaluating the similarity of undirected attributed graphs is a well-described problem. Graph edit distance (GED) is the most frequently used method. However, because the definition of the cost function is the key issue for GED algorithms and is strongly application dependent (Gao et al. 2010), most GED algorithms cannot be easily used to evaluate the similarity of SDUs. In contrast, (sub)graph isomorphism-based graph distances do not have to define a cost function, rendering them less limited and more suitable to the situation. Moreover, the error-tolerant graph matching algorithms (e.g., GED) will run faster only when finding the maximum common subgraph is not guaranteed. However, to search SDU alternatives for floor-plan design, designers have a strong need to find the best match. Consequently, there are no obvious efficiency advantages for GED against (sub)graph isomorphism-based graph distances. Therefore, (sub)graph isomorphism is chosen as the solution for evaluating the topological similarity of SDUs.

To use subgraph isomorphism as a graph distance, Bunke and Shearer (1998) and Wallis et al. (2001) used the maximum common subgraph, and Fernández and Valiente (2001) combined the maximumcommonsubgraphandtheminimumcommonsupergraphtogetherasadistancemeasure.

However, neither the maximum common subgraph nor the minimum common supergraph considers the similarity of smaller substructures. Because substructures may imply similar functions and design styles between SDUs, they are crucial to evaluating the similarity of SDUs.

Langenhan and Petzold proposed a decision-tree based subgraph isomorphism algorithm. This method was originally proposed by Messmer and Bunke (1999) and was improved by applying a new strategy in the tree generation (Weber et al. 2012), which can significantly reduce the size of the tree. The algorithm does not provide a distance or similarity measure. However, the decision tree constructed from the space topology is capable of calculating the similarities of substructures of any size in an SDU. Therefore, this paper proposes a topological similarity evaluation of SDUs based on that algorithm.

The Shape Similarity

Measuring similarities between 2D geometric shapes is a shape matching problem. One common approach to solving this problem is to use invariants or descriptors, such as moment invariants (Boyce and Hossack 1983), the Fourier descriptor (Persoon and Fu 1977), and the wavelet descriptor (Chuang and Kuo 1996). However, transformations may significantly change the living quality of an SDU. For example, a rotation will cause an orientation change, whereas a scaling will cause a size change. Consequently, most of the invariants described above cannot be considered 'invariant' in evaluating the shape similarity of SDUs. Various morphological features are also used for shape recognition, including the Hough transform (Ballard 1981), deformable templates (Christensen et al. 1996), and matching convex and concave structures (Ueda and Suzuki 1993). However, in most cases, SDUs are polygonal and lack morphological features. Therefore, using the overlap area of shapes as an evaluation of the shape similarity of SDUs is more suitable, and a straightforward polygonal description or approximation is capable of encoding the shape of an SDU.

Calculating the maximum overlap area of two polygons is an NP-hard problem. Most of the currentresearchstudiesaremorespecific; forexample,theyconsideronlyconvexpolygons(DeBerg et al. 1998; Ahn et al. 2013) or the translation transformation (Mount et al. 1996; Har-Peled and Roy 2014). However, the shape of SDUs may contain concavities, and transformations other than translation (for example, reflection) may be applied. To solve this problem in a reasonable time, this paper proposes a simplified method according to a special pattern in the floor-plan design.

FRAMEWORK AND METHODOLOGY

SDU is a standard design for a dwelling unit. Generally, SDUs act as basic components of the residential design. Providing similar SDU alternatives in the floor-plan design of a residential building can improve efficiency while reducing cost and risk. However, according to the literature review, current approaches must be further improved to evaluate the similarity between SDUs. For attribute-based similarity, important attributes must be selected with proper evaluation methods.

For topological similarity, the decision-tree based algorithm proposed by Weber et al. (2012) must be improved in order to evaluate the similarities of the substructures in an SDU. For shape similarity, a novel and straightforward method is proposed according to a special pattern in the floor-plan design in this paper. Furthermore, when designers must consider different similarities at the same time, a combined similarity is also required. Moreover, to evaluate all similarities mentioned above, corresponding information-extraction methods and data preprocesses must be developed.

Therefore, a framework of the multi-aspect approach to the similarity evaluation of SDUs is proposed, as depicted in Fig. 1. First, the required information is extracted from the information models of SDUs, which are stored in IFC files. The extracted information is then reorganized into attributes, topologies, and shapes. Second, each model is preprocessed according to the regulations in the residential design. All types of similarities are then calculated and combined, providing a comprehensive index for the design recommendation. The following four sections with discuss the information extraction, the similarity evaluation, the application test, and the designer feedback respectively.

The aim of this research is not to build a complete standard design library, but to evaluate the similarities between SDUs. To fulfill that purpose, the SDUs only require the model to contain basic components (i.e. walls, doors, and windows) and spaces. Additional information like material, structure features are not required. The information-extraction process of an SDU requires all needed information to be properly and precisely modeled in the IFC file. Therefore, designers should follow certain modeling guidelines as listed below:

1. Set rooms in the model. Ensure every room is exported as IfcSpace with correct geometry.
2. There must be no overlap among rooms in the model.
3. There must be at least one room, one exterior door and one exterior window in the model.
4. Walls, doors, windows, rooms, and the relationships between them must be exported with the IFC file correctly.
5. Additional attributes that are not stored in IFC file must be input when the SDU is uploaded, including name, region, building height and the ratio of elevators or stairs per dwelling unit.

SIMILARITY DEFINITION AND INFORMATION EXTRACTION

Attribute-Based Similarities

The authors have surveyed several current Chinese dwelling unit databases and websites to identify the most frequently used attributes. These databases include the Zhibeizhen dwelling unit database(www.huxingku.com),theCSCECdwellingunitdatabase(Heetal.2014)andtheTsinghua

Database, which was made by the authors in a previous work (He et al. 2016). Websites include Kujiale (www.kujiale.com), Chinauhu (www.chinauhu.com) and Nbimer (www.nbimer.com). All attributes contained in the databases and websites are summarized and listed in TABLE 1.

However, with regard to calculating attribute-based similarities, only those attributes that are reusable, easy to normalize and not considered in topological or shape similarities are taken into account. For the attributes of name, code, category, developer, and project, even though they are important to database management, they are difficult to standardize and their relation to the similarity of SDUs is not clear. For the attributes of breadth and depth, they only act only a rough estimate of an SDU’s shape. A specialized measurement of shape similarity will be much better. For the attribute of price, not all SDUs have such an attribute. In addition, the price is influenced by too many irrelevant factors such as region, time, and housing policies, most of which do not have a clear relation to the similarity of SDUs. Consequently, only room count, room area, building height, location, elevator or stair count, and orientation are considered.

Room count and room area are represented by lists with nonnegative integers and real numbers, respectively. Both of them can be calculated by the cosine similarity. Building height is normalized asa nominalattributeconsisting of‘high’, ‘middle’ and‘low’. Thelocationis representedasclimate region and is normalized according to the Chinese ‘Code for Design of Civil Buildings (GB50352-2005)’, which contains ‘cold’, ‘warm’ and ‘hot’. Elevator or stair count is represented as the ratio of elevators or stairs per dwelling unit (such as 1 stair per 2 dwelling units), which is an index commonly used in China to reflect a building's residential density. Orientation is represented as 'N', 'S', 'E', 'W' and their combinations. All four nominal attributes are calculated together using the ratio of matches. In summary, there are 3 independent attribute-based similarities, which are the room count similarity (RCS), the room area similarity (RAS) and the nominal attribute similarity (NAS), as depicted in TABLE 2.

The Topological Similarities

Topology retrieval from IFC files

There are two important types of topologies of an SDU: accessibility and adjacency. The accessibility of rooms is determined by whether a resident can move from one room to another and vice versa. The adjacency of rooms is determined by whether two rooms are next to each other.

Clearly, two rooms can be accessible and adjacent to each other at the same time. According to these two topologies, the accessibility similarity (ACS) and the adjacency similarity (ADS) are proposed.

Topology information can be retrieved from IFC files, but additional processes are required. Generally, if two rooms are separated by one wall, then they might be adjacent to each other. Such relationships can be found through IfcSpace, IfcRelSpaceBoundary, and IfcWall, as depicted in Fig. 2a. If there is a door in the wall that acts as the boundary for two rooms, then these two rooms are accessible to each other. Such relationships can be found through IfcDoor, IfcRelVoidsElement, and IfcOpenningElement, as depicted in Fig. 2b. However, there are exceptions to those situations described above. If the wall is not separated properly, then the determination may lead to an incorrect topology. It should be checked if the nearest distances between the two rooms are less than the thickness of the wall. If they are, then the two rooms are adjacent or accessible to each other. In addition to walls, virtual room separators also provide rooms with boundaries. However, rooms separated by virtual room separators are accessible to each other because the rooms are not physically separated. If a boundary is a virtual room separator, the accessibility must be determined by whether the nearest distance between two virtual boundaries is zero, because the boundaries do not refer to any real elements.

Preprocesses

After the original topologies are extracted from IFC, the topologies must be further processed accordingto therequirements ofthe residentialdesignbeforeevaluatingthetopologicalsimilarities.

The preprocesses primarily include four cases: indoor corridors, multifunctional rooms, virtually separated rooms and the orientation of rooms (see Fig. 3).

Indoor corridors: An indoor corridor is a passageway providing access between rooms inside an SDU. Therefore, corridors should be considered related to accessibility, not as rooms with specific functions. In other words, they should be translated into edges instead of nodes in the graph representing the accessibility of the SDU. All rooms that are accessible via the same corridor should be accessible to each other in this case, as depicted in Fig. 3a.

Multifunctional rooms: A room may have multiple functions, particularly in small apartment designs. For example, one room can function either as a kitchen and a dining room simultaneously or as a bedroom and a living room simultaneously. However, multiple functions generate unspecific labels in the attributed graph representing the SDU, making the determination of subgraph isomorphism more complicated and difficult. Consequently, a multifunctional room is treated as separate single function rooms in this process. Each separate single function room is accessible and adjacent to one another and has all the accessibilities and adjacencies as the original multifunctional room, as depicted in Fig. 3b.

Virtuallyseparatedrooms: Designersdivideasinglespaceintoseveralroomsusingvirtualroom separators in an arbitrary way. Such behavior may cause uncertainty in the extracted topology. For example, if a virtual room separator is near a door, the accessibility topology will differ according to where the designer put the separator line. This uncertainty will cause errors in calculating subgraph isomorphism when evaluating the topological similarities. To eliminate such uncertainty, we first find which rooms belong to a single space but are divided by virtual room separators. Then, all related rooms are set to be accessible and adjacent to each other, and every room is set to share accessibility and adjacency with other rooms, as depicted in Fig. 3c.

The orientation of rooms: The orientation of a room is an important feature. Even with identical layouts and sizes, rooms with different orientations may strongly impact the living quality of their SDU. To consider this impact, predefined orientations (including north, south, east, and west) are added into the adjacency graph as nodes. If an exterior wall has a window facing outside, then the rooms that take this wall as a boundary may have the same orientation as depicted in Fig. 3d.

Another prerequisite is to ensure that the wall is separated properly. If the nearest distance between the window and the room is less than the thickness of the wall, then the room is facing the same direction.

Note that because SDUs are assumed to be single-floor units, the topology is represented as a planar graph. However, if an SDU has undergone any of the preprocesses listed above, it may lose its planarity (e.g., the indoor corridors or multifunctional rooms). This is the reason why a faster subgraph isomorphism algorithm (Hopcroft and Wong 1974) is not applied here.

The Shape Similarity

As mentioned above, a shape similarity (SS) is proposed based on the overlap area of two SDUs. However, most of the similarity evaluations are either restricted or inefficient. To solve this problem in a reasonable time, a simplified method is proposed in this paper according to a special pattern in the floor-plan design. The public spaces of an apartment building, including corridors and the vertical transportation system, are the core of a floor-plan. All dwelling units are connected to this core by exterior doors. Retaining the layout of the core and changing its connected SDUs is a common approach for customizing the floor-plan design. Therefore, if an SDU is replaced by another SDU, the locations of their exterior doors are very close to each other. According to this pattern, the maximum overlap area of two SDUs is calculated by ensuring that their exterior doors are coincident. Here the location of the exterior door is taken as a point. When the exterior doors are coincident, the SDUs can be mirrored across the axis of the wall in which the exterior door is located or across the line perpendicular to that wall at the exterior door’s position. Therefore, four possible cases exist for two SDUs. The maximum overlap area of these cases can be used as an evaluation of shape similarity.

The shape of an SDU is calculated by the shapes of its rooms. The shape of a room can be extracted from the geometric representation of an IfcSpace. However, because of the existence of interior walls in an SDU, there are gaps between room shapes. Those gaps must be filled before calculating the shape of an SDU. First, the gaps are filled using the polygon offset algorithm proposed by Chen and McMains (2005). Next, the shape is calculated by merging all rooms' shapes using the Boolean union algorithm proposed by Vatti (1992).

SIMILARITY EVALUATION

Attribute-Based Similarities

As mentioned above, room count, room area, building height, climate region, the ratio of elevators or stairs per dwelling unit, and orientation are considered in the following attribute-based similarities: RCS, RAS, and NAS. All three similarities require the transformation of an SDU into a vector. For RCS or RAS, the dimension of the vector is equal to the count of the room types. For NAS, the dimension of the vector is equal to the number of attributes. Given two SDUs, A and B, the following are the definitions and formulations used for calculating RCS/RAS:

- C is a set of room types from both A and B; n is the size of C.
- $L :\rightarrow C$ is a function that assigns room types to rooms. For example, room type L(i) is assigned to the No.i room in C.
- $\mathbf{x} = (x_1, x_2, ..., x_n)$ is a vector transformed from A, where x_i is the count/area of rooms in A whose type is equal to L(i). Vector $\mathbf{y}$ is transformed from B in the same manner.
- $\text{sim}(A, B) = \frac{\mathbf{x}\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}$ is the measure of RCS/RAS.

Below are the definitions and formulations used for calculating NAS:

• $\mathbf{x} = (x_i, x_2, x_3, x_4)$ is a vector transformed from A; x_1 is the building's height; x_2 is the climate region; x_3 is the ratio of elevators or stairs per dwelling unit; x_4 is the orientation.

Vector **y** is transformed from B in the same manner.

• $\text{match}(i) = \begin{cases} 1, & x_i = y_i \\ 0, & x_i \neq y_i \end{cases}$ is a function used for matching elements in **x** and **y**.

• $\text{sim}(A, B) = \frac{\sum_{i=1}^{4} \text{match}(i)}{4}$ is the measure of NAS.

Clearly, RCS, RAS, and NAS are all symmetric (sim(A, B) = sim(B, A)) and range between 0 and 1 ($0 \leq \text{sim} \leq 1$).

Topological Similarities

According to the topologies of accessibility and adjacency of an SDU, ACS, and ADS are proposed respectively. Both similarities are evaluated by (sub)graph isomorphism-based methods. To better evaluate the substructures inside SDUs, a decision-tree method proposed by (Weber et al. 2012) is adopted and extended. That method computes all permutations of the adjacency matrices under certain constraints (a weight function and a well-founded total order) and then transforms them into row-column vectors, which are then used to generate the decision tree. Weber's method is capable of considering directed graphs, whereas this research considers only undirected graphs. Consequently, the matrices representing adjacency and accessibility are symmetric. Therefore, the row-column vectors can be represented by half of their original length because of the symmetry.

Consider SDUs A and B with m and n rooms, respectively. Function sim(A,B) calculates the similarity between A and B. Function sim(A,B,k) calculates the similarity between A and B considering only k × k order submatrices. Function match(A,B,k) calculates the matches of graph isomorphism in k × k order submatrices by traversing and matching trees to the depth of k. Thus, we have:

$$\mathrm{sim}(A, B, k) = \frac{2 \times \mathrm{match}(A, B, k)}{\mathrm{match}(A, A, k) + \mathrm{match}(B, B, k)} \quad (1)$$

$$\mathrm{sim}(A, B) = \frac{\sum_{k=1}^{\min(m,n)} \mathrm{sim}(A, B, k)}{\max(m, n)} = \frac{\sum_{k=1}^{\min(m,n)} \frac{2 \times \mathrm{match}(A, B, k)}{\mathrm{match}(A, A, k) + \mathrm{match}(B, B, k)}}{\max(m, n)} \quad (2)$$

According to the definitions, match(A,A,k) is equal to the count of the k-depth nodes of the tree representing SDU A. Clearly, both ADS and ACS are symmetric and range between 0 and 1. Fig. 4 displays an example of the ADS between two SDUs with four rooms with functions tagged as P, Q, R, and S. Each SDU is represented by a layout, an undirected attributed graph, a matrix and a tree. In this example, the ADS is $\mathrm{sim}(A, B) = \frac{1}{4}\left(\frac{2}{3} + \frac{3}{5} + \frac{1}{5}\right) = \frac{11}{30} \approx 36.67\%$ according to Eq. 1 and Eq. 2.

The Shape Similarity

As mentioned above, SS is evaluated by calculating the overlap area of the shapes of two SDUs when the shapes of the two SDUs coincide at the exterior doors. When the exterior doors are coincident, the SDUs can be mirrored across the axis of the wall in which the exterior door is located or across the line perpendicular to that wall at the exterior door's position. Four possible cases exist for two SDUs, as depicted in Fig. 5.

The maximum overlap area of these cases is used as a measure of their shape similarity. Here, the Jaccard coefficient is used to evaluate the shape similarity. Consider SDUs A and B with shapes S_A and S_B, respectively. Thus:

$$\mathrm{sim}(A, B) = \frac{\mathrm{Area}(S_A \cap S_B)}{\mathrm{Area}(S_A \cup S_B)} \quad (3)$$

where $S_A \cap S_B$ represents the Boolean intersection result of S_A and S_B (the overlap area) and $S_A \cup S_B$ representstheBooleanunionresultof S_A and S_B. BothBooleanoperationscanbeperformed using the algorithms proposed by Chen and McMains (Chen and McMains 2005). The SSs are also symmetric and range between 0 and 1.

The Combined Similarity

When designers try to find proper SDU alternatives, they need to consider more than one similarities at the same time. For that purpose, we provide a combined similarity that comprises a linear combination of the six proposed similarities to reflect designers' search intentions. Let weight vector $\mathbf{w} = (w_{RCS}, w_{RAS}, w_{NAS}, w_{ACS}, w_{ADS}, w_{SS})^T$ represent 6 weights for each similarities, with $0 \le w_{RCS}, w_{RAS}, w_{NAS}, w_{ACS}, w_{ADS}, w_{SS} \le 1$ and $w_{RCS} + w_{RAS} + w_{NAS} + w_{ACS} + w_{ADS} + w_{SS} = 1$. Let vector $\mathbf{s} = (s_{RCS}, s_{RAS}, s_{NAS}, s_{ACS}, s_{ADS}, s_{SS})^T$ represent 6 similarities between 2 SDUs named A and B. Thus, the combined similarity sim_{COM} can be presented as follows:

$$\mathrm{sim}_{COM}(A, B) = \mathbf{w}^T\mathbf{s} \quad (4)$$

A designer can define his/her search intention by adjusting the weights of the combination. To further help users make their combination faster, this paper proposes a semi-automated approach based on a genetic algorithm (GA) to accelerate the process, as depicted in Fig. 6.

To determine the combinational weights, one or more designers are invited to evaluate the similarities of SDUs manually. Each designer divides a certain number of SDUs into several groups and evaluates the degree of similarity among these groups. The similarity measures are then averaged and normalized between 0 and 1. Given N SDUs X_1, X_2, ..., X_N, the similarity matrix is a matrix defined as follows:

$$\mathbf{S} = \begin{pmatrix} \mathrm{sim}(X_1, X_1) & \cdots & \mathrm{sim}(X_1, X_N) \\ \vdots & \ddots & \vdots \\ \mathrm{sim}(X_N, X_1) & \cdots & \mathrm{sim}(X_N, X_N) \end{pmatrix} \quad (5)$$

Then, thesimilaritymatrixconstructedfromthedesigners’evaluationcanbepresentedas$\mathbf{S}_{target}$.

Because all 6 similarities are symmetric and range between 0 and 1, the combined similarity is also symmetric and ranges between 0 and 1. Let the similarity matrices of all 6 proposed similarities be $\mathbf{S}_{RCS}$, $\mathbf{S}_{RAS}$, $\mathbf{S}_{NAS}$, $\mathbf{S}_{ACS}$, $\mathbf{S}_{ADS}$ *and* $\mathbf{S}_{SS}$. The determination of the weights in the linear combination can be described as the following optimization problem:

$$\min cost(\mathbf{x}) = \mathbf{S}_{target} - w_{RCS}\mathbf{S}_{RCS} - w_{RAS}\mathbf{S}_{RAS} - w_{NAS}\mathbf{S}_{NAS} - w_{ACS}\mathbf{S}_{ACS} - w_{ADS}\mathbf{S}_{ADS} - w_{SS}\mathbf{S}_{SS}$$

$$s.t.\ 0 \leq w_{RCS}, w_{RAS}, w_{NAS}, w_{ACS}, w_{ADS}, w_{SS} \leq 1$$

$$w_{RCS} + w_{RAS} + w_{NAS} + w_{ACS} + w_{ADS} + w_{SS} = 1 \quad (6)$$

where k**S**k is the matrix 2-norm. The weights have defined the specific searching intentions made from the experience and opinions of the invited designers. However, the relationship between designers' intentions and proposed similarities is unknown. To find an acceptable weight combination in a reasonable time, we used a GA as a heuristic way to solve the problem. In the GA, each chromosome has 60 genes, and each gene is represented by a bit. A set of every 10 genes together represents a weight value that ranges from 0 to $2^{10} - 1$. Next, each weight value is normalized to meet the constraints before calculating the cost. Such coding makes the crossover operator and the mutation operator simple and straightforward. The crossover operator randomly selects genes from their parents, as depicted in Fig. 7a. In addition, the mutation operator randomly changes one gene to a random bit value (0 or 1), as depicted in Fig. 7b. The cost function is used as the unfitness score. Roulette selection is used as the selection operator. Since the combination is linear, the GA is simple and stable enough to solve the problem. The process of applying this GA and the results are described and discussed in the next section.

APPLICATION AND ANALYSIS

The SDU Database

The authors collected 170 SDUs by collaborating with the Architectural Design and Research

Institute of Tsinghua University (THAD) and the China State Construction Technical Center (CSCECTC). The SDUs are from 13 cities in China, covering 3 different climate regions according to the Chinese '*Code for Design of Civil Buildings* (GB50352-2005)'. The areas of the SDUs range from 30 m^2 to 200 m^2. All SDUs were modeled using Autodesk Revit 2014 or 2016 and were exported as IFC 2x3 files. A prototype SDU database and a design tool were developed based on our previous research (see Fig. 8). The calculation of the proposed similarities and the corresponding information-extraction methods were all realized.

The prototype database and design tool and all SDUs are applied and tested in two residential projects in China. The first project is the Changyang Tiandi project located in Beijing with 12 apartment buildings and a total floor area of 70,100 m^2. The second project is the Hupan New City project located in Hefei with 39 apartment buildings and a total floor area of 446,500 m^2.

Application of the Combined Similarity

In this research, a combined similarity using a linear combination is proposed to help designers define their search intentions. Following the proposed GA-based approach, 7 designers from THAD were invited to evaluate the similarities of SDUs manually. Each invited designer was told to divide all 170 SDUs into not less than 5 groups and to evaluate the degree of similarity among these groups. The similarity measures were then averaged and normalized between 0 and 1. Next, the GA described above was implemented to determine a satisfactory weight combination. The initial population is set to 100 and is kept in each generation. The algorithm will stop if it runs for more than 10000 generations or if the best solution has not been improved for more than 500 generations.

Among all 170 SDUs, the first 100 SDUs are chosen as the training set, whereas the last 70 SDUs are chosen as the test set. This GA was run 1000 times, and all runs were found to converge near the best run. The best run and its results are listed in TABLE 3 and TABLE 4:

The distribution of the weights provides designers' different perspectives on various aspects of the SDUs. NAS has the highest weight, which indicates that climate region, building height, residential density and orientation remain the focus of designers. ACS has the second-highest weight, which indicates that designers can recognize substructures of the space topology of SDUs and consider them while evaluating their similarities. RAS has the third-highest weight, which indicates the influence of room sizes on the living quality of SDUs. The last aspect is SS, whose weight is less than 4%, which may indicate that the shape of an SDU is not among designers' greatest concerns. The weights of RCS and ADS are set to zero. This may be caused by the overlap of aspects between RCS and RAS and between ACS and ADS, and the weights are restricted to be positive during the evolving process. It shows that RAS and ACS reflect the designers' search intention better than RCS and ADS.

Note that the weights are used to approximate the matrix $\mathbf{S}_{target}$. However, $\mathbf{S}_{target}$ is strongly influenced by the experience and opinions of the invited designers. In fact, the matrices generated by the designers are quite different from each other. consequently, the combined similarity is a result of compromise. If a better user experience is required in a specific case, then the best weight combination can be generated according to the similarity evaluations of a certain designer using the same GA-based method.

Performance Analysis

The attribute-based similarities are calculated either by the cosine similarity or by the ratio of matches. Both methods have a time complexity of O(N). For RCS and RAS, N is the number of room types; for NAS, N is the number of attributes. The topological similarities, including ACS and ADS, are calculated by a method based on Weber’s subgraph isomorphism algorithm. This method takes O(N!) (Langenhan et al. 2013) to generate the decision tree in the worst circumstances and takes O(N^2) to complete matching and calculate the similarity. Here, N is the number of rooms.

The shape similarity is calculated by Vatti’s algorithm, with a time complexity of O(NlogN) (Vatti 1992). N is the number of points of the polygon representing the shape of an SDU. The shape similarity also must offset the rooms’ shapes before calculating the shape of the SDU, which is realized by implementing Chen’s algorithm, whose time complexity is also O(NlogN) (Chen and McMains 2005).

All SDUs were imported into the database, and all similarities between every pair of the 170 SDUs were evaluated. All calculations were performed on a computer equipped with an i7-6700 CPU (3.4 GHz, 4 cores, and 8 threads) and DDR4 8GB RAM. The time consumptions are listed in TABLE 5.

According to the chart above, it costs only 3547.1 ms to calculate all similarities of the 170 SDUs without repetition (time consumption I, 14365 pairs in total). The time is primarily consumed by ACS, ADS, and SS, which cost 50.64%, 20.02%, and 24.69% of the total time consumption (time consumption I), respectively. If the time used to generate a decision tree for calculating ACS and ADS is included, then the entire process takes no more than 5 minutes.

When an SDU is imported, all related similarities are calculated and stored automatically on the server. As the SDUs are being imported, the new imported SDU must be paired with all imported SDUs to calculate the similarities, resulting in a linear increase in time consumption with the increasing size of the database. In addition, the characteristics of a specific SDU will influence the amount of time consumed (see Fig. 9).

In Fig. 9, the RCS and RAS are influenced by the number of room types in an SDU; the ACS and ADS are influenced by the number of rooms in an SDU; SS is influenced by the complexity of the shape of an SDU; NAS is almost uninfluenced by any characteristics of an SDU. Although time consumption without tree generation will increase linearly as the size of the database increases, the time consumption for generating the decision tree will not, as the latter time consumption is only related to the characteristics of a certain SDU. Therefore, according to an approximately linear regression, if 10^3 SDUs are imported into the database, it will take approximately 30 minutes to generate the decision tree and approximately 7 minutes to calculate all the similarities. This result is promising because it proves the efficiency and feasibility of the proposed multi-aspect approach for the similarity evaluation of SDUs.

Design Recommendation in the Floor-plan Design

The public spaces of an apartment building, including the vertical transportation system and the corridors, are the core of a floor-plan. All dwelling units are connected to this core by exterior doors. Retaining the layout of the core and changing its connected SDUs is a common approach in the floor-plan design process. In this research, a prototype design tool (plugin) based on Revit 2016 is developed to provide design recommendations according to the combined similarity and preview the replacement of the SDUs. An example from the floor-plan design of the project is depicted in Fig. 10a.

Fig. 10b shows the steps of SDU search and replacement. First, the user designs a floor-plan using Revit and then sets the rooms to several SDUs and public spaces. Second, the user selects an SDU to search for similar SDUs in the database. The design tool will provide SDU alternatives sorted according to their similarities in descending order. Third, the user can select any recommended SDU to preview the replacement. Finally, the user can load the selected SDU into the Revit platform and then replace the former one. Because the preview and replacement are based on the exterior doors, the user must adjust the layout manually after the replacement process.

An example of the comparison between attribute-based search and similarity-based search is shown in Fig. 10b and Fig. 10c. For attribute-based search, the designer needs to look for proper SDU alternatives in the search result. For similarity-based search, the designer will look for proper SDU alternatives in the order of similarity. Even though the search areas of the two methods are the same, the efficiencies are different. In the given example, one can find a proper alternative very soon using NAS, SS, or the combined similarity. This example is a common case during the application test. Therefore, in most cases, the approach proposed in this paper is more efficient than the attribute-based search.

The design tool acts as a means of searching for SDU alternatives using the combined similarity. Using the tool will not change the existing process of employing SDUs in the floor-plan design. The main purpose is to help designer find better SDU alternatives in less time. The feedback and discussion can be found in the next section.

DESIGNER FEEDBACK AND DISCUSSION

The prototype database and design tool and all SDUs are applied and tested in the two above-mentioned residential projects in China, Changyang Tiandi in Beijing and Hupan New City project in Hefei. Several designers from the projects participated in the modeling of the SDUs and the design-recommendation process. In general, the designers approved the purpose and effect of the proposed multi-aspect approach to a similarity evaluation of SDUs. The approved advantages of this approach are summarized as below:

1. The establishment of the database is easy and fast.

2. Six independent quantitative similarities of SDUs (RCS, RAS, NAS, ADS, ACS, and SS) cover important aspects of the attribute, topology, and geometry of SDUs. Compared to the attribute-based search method, the approach exploits the rich information of BIM to help find proper SDU alternatives in a more comprehensive manner.

3. The combined similarity provides a flexible way for designers to use all six similarities together. The GA-based approach is capable of building up a combination quicklyaccording to the manual evaluation of one or more designers. The result can be explained and adjusted for actual usage.

4. Searching for similar alternatives through the SDU database using the design tool is more efficient than searching those databases using attribute-based methods, which reduces the design efforts.

5. During SDU searching, the design tool can detect similarities that are not noticed by the designers, which sometimes brought inspiration to the designers.

6. This method only requires the company to design and store their SDUs in IFC file format, and follow several modeling guidelines proposed in the framework section. Since BIM and IFC are becoming more and more popular in the residential design, it will not be difficult for companies to adopt the approach proposed.

However, there are still some limitations of this approach according to the application:

1. Although this research has proposed six different similarities covering semantic, geometric, and topological aspects of SDUs, there still are more attributes could be taken into account. For example, an SDU also contains other topologies, such as the visibility topology con-

structed from isovists, that can be used to measure the similarity between SDUs as well. By taking more interesting attributes into account, a new similarity may lead to better results and better reflect the views of the designers.

2. The combined similarity could be more flexible. This research used a linear combination of similarities as a way to help designers consider more than one similarities during the floor-plan design. A non-linear combination may provide better performance. Furthermore, continuous learning and self-improvement for the combination of multi-aspect similarities based on user habits could be provided, thereby satisfying the requirements of different designers or users.

3. Importing a large (more than 15-room) SDU into the database or replacing a large SDU in the floor-plan design is slow. The calculation efficiency for large SDUs could be improved by improving the algorithms or applying other algorithms.

4. In the floor-plan design, designers need to manually adjust the design after the replacement of SDUs, because most alternative SDUs do not precisely suit the situation. If an automated adjustment process follows the replacement of SDUs according to certain requirements and specifications, it would greatly improve the approach’s efficiency and effect.

5. Information extraction of an SDU requires that all information needed is properly and precisely modeled in the IFC file. For exceptions, new methods could be used, such as voxelization (Daum et al. 2014) or a straight medial axis transform (Taneja et al. 2011).

6. The SDU model needs to be reanalyzed if certain requirements change. For example, if the SDU name normalization rules change, the adjacent graph and accessible graph of every SDU need to be adjusted due to the change. Automation in the database maintenance could help solve this problem.

CONCLUSION AND FUTURE WORK

This paper proposes a multi-aspect approach to the similarity evaluation of SDUs to support the similarity-based SDU search. From aspects of the attribute, topology, and shape of an SDU, this approach contains the RCS, RAS, NAS, ADS, ACS, and SS with corresponding information-extraction and similarity evaluation methods. Furthermore, a combined similarity is adopted to combine different similarities to reflect designers' search intentions. Finally, a database with 170 SDUs and a prototype design tool are developed and tested in two residential projects. According to the application and analysis, the proposed approach is capable of evaluating the similarity of SDUs in a more comprehensive way and making searching for SDU alternatives faster and more flexible, thus improving the design efficiency and quality. The main contributions of this paper are summarized as below:

1. Six independent similarities (RCS, RAS, NAS, ADS, ACS, and SS) have been proposed fromtheaspectsoftheattribute, topology, andgeometryofanSDU.Comparedtothecurrent practice, this approach can evaluate the similarities of SDUs in a more comprehensive way, and take advantage of the rich information of BIM.
2. The similarity evaluation process is efficient. In our test, it takes approximately only 3.5 seconds to calculate the similarities of 170 SDUs. If the tree generation is included, it takes fewer than 5 minutes. According to an approximately linear regression, this method is capable of processing a database with at least 10^3 SDUs in a reasonable time. Compared to manual evaluation by designers in current practice, the proposed approach can considerably reduce designers' efforts to find proper SDU alternatives.

3. The combined similarity can help define designers' search intentions. Currently, designers need to find proper attributes by themselves before searching for SDU alternatives. The proposed GA-based approach, in contrast, helps designers customize their search criteria semi-automatically and heuristically. This also helps designers find better SDU alternatives in less time and reduce design efforts.

4. Instead of providing results that strictly match certain rules, the proposed approach provides SDU alternatives that are ordered by the similarity combinations customized by users.

Replacing a similar SDU can prevent major design changes in the floor-plan design. In addition, the algorithm sometimes detects similarities that are not noticed by the designers. This 'unpredictability' in the search results can bring inspiration to the designers, thereby improving the design quality.

Because it is the initial demonstration of the approach, the proposed approach can be further improved in the following ways:

1. More attributes and more similarities will be added to further utilize the rich information in BIM. For example, calculating the similarity of modular component and furniture in SDUs will assist the industrialized residential floor-plan design.

2. Use nonlinear combination and other types of data-driven classification systems to help build a better-combined similarity which will reflect the designers' search intention in a more effective way.

3. For supporting larger SDU databases and more complex SDU models, the efficiency of similarity evaluation will be further improved by improving the algorithms or applying other algorithms.

4. Currently the SDU alternatives need adjustment to fit the floor-plan after the replacement of original SDUs in most cases. However, if the difference between the original SDU and alternative SDU is small enough, this work could be done by the computer automatically.

Therefore, algorithms for automating the adjustment process will be developed to further support the floor-plan design.

5. For making the similarity evaluation more robust and error-tolerant, additional algorithms should be developed to deal with defective SDU models.

ACKNOWLEDGEMENTS

The authors are grateful for the support provided by the National High-Tech R&D Program of China (No.2013AA041307), the National Natural Science Foundation of China (No.51278274), the China Postdoctoral Science Foundation Grant (No. 2016M601038) and the Young Elite Scientists Sponsorship Program by CAST (No.YESS20160122).

REFERENCES

Ahn, H.-K., Cheng, S.-W., and Reinbacher, I. (2013). “Maximum overlap of convex polytopes under translation.” *Computational Geometry*, 46(5), 552–565.

Ballard, D. H. (1981). “Generalizing the Hough transform to detect arbitrary shapes.” *Pattern recognition*, 13(2), 111–122.

Boyce, J. F. and Hossack, W. (1983). “Moment invariants for pattern recognition.” *Pattern Recognition Letters*, 1(5-6), 451–456.

Bunke, H. and Shearer, K. (1998). “A graph distance metric based on the maximal common subgraph.” *Pattern recognition letters*, 19(3), 255–259.

Chen, X. and McMains, S. (2005). “Polygon offsetting by computing winding numbers.” *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, 565–575.

Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1996). “Deformable templates using large deformation kinematics.” *IEEE transactions on image processing*, 5(10), 1435–1447.

Chuang, G.-H. and Kuo, C.-C. (1996). “Wavelet descriptor of planar curves: Theory and applications.” *IEEE Transactions on Image Processing*, 5(1), 56–70.

Daum, S., Borrmann, A., Langenhan, C., and Petzold, F. (2014). “Automated generation of building fingerprints using a spatio-semantic query language for building information models." *eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2014*, 87.

De Berg, M., Cheong, O., Devillers, O., Van Kreveld, M., and Teillaud, M. (1998). "Computing the maximum overlap of two convex polygons under translations." *Theory of computing systems*, 31(5), 613–628.

Eastman, C.M., Eastman, C., Teicholz, P., Sacks, R., andListon, K.(2011).*BIMhandbook: Aguide to building information modeling for owners, managers, designers, engineers and contractors*. John Wiley & Sons.

Fernández, M.-L. and Valiente, G. (2001). "A graph distance metric combining maximum common subgraph and minimum common supergraph." *Pattern Recognition Letters*, 22(6), 753–758.

Gao, X., Xiao, B., Tao, D., and Li, X. (2010). "A survey of graph edit distance." *Pattern Analysis and applications*, 13(1), 113–129.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Har-Peled, S. and Roy, S. (2014). "Approximating the maximum overlap of polygons under translation." *European Symposium on Algorithms*, Springer, 542–553.

He, T., Lin, J., Hu, Z., Zhang, J., Jia, W., and Su, J. (2014). "BIM-Based Plan Modeling System at Preliminary Stage for Residential Real Estate Projects." *Computing in Civil and Building Engineering*, Orlando, USA, 488–495.

He, T., Zhang, J., Lin, J., Jia, W., and Hu, Z. (2016). "A Topological and Hierarchical Information Integration Approach for Standard-unit-based Residential Planning." *Proceedings of the 16th International Conference on Computing in Civil and Building Engineering*, 381–388.

Hillier, B. (2007). *Space is the machine: a configurational theory of architecture*. Space Syntax.

Hopcroft, J. E. and Wong, J.-K. (1974). "Linear time algorithm for isomorphism of planar graphs (preliminary report)." *Proceedings of the sixth annual ACM symposium on Theory of computing*, ACM, 172–184.

Isikdag, U., Zlatanova, S., and Underwood, J. (2013). “A BIM-Oriented Model for supporting indoor navigation requirements.” *Computers, Environment and Urban Systems*, 41, 112–123.

Langenhan, C. and Petzold, F. (2010). “The fingerprint of architecture-sketch-based design methods for researching building layouts through the semantic fingerprinting of floor plans.” *International electronic scientific-educational journal: Architecture and Modern Information Technologies*, 4, 13.

Langenhan, C., Weber, M., Liwicki, M., Petzold, F., and Dengel, A. (2013). “Graph-based retrieval of building information models for supporting the early design stages.” *Advanced Engineering Informatics*, 27(4), 413–426.

Li, H., Cao, Y., and Lu, M. (2004). “Semiautomated detection of design errors in 2d drawings using 3d reconstruction.” *Computer-Aided Civil and Infrastructure Engineering*, 19(4), 288–294.

Messmer, B.T.andBunke, H.(1999).“Adecisiontreeapproachtographandsubgraphisomorphism detection.” *Pattern recognition*, 32(12), 1979–1998.

Mount, D. M., Silverman, R., and Wu, A. Y. (1996). “On the area of overlap of translated polygons.” *Computer Vision and Image Understanding*, 64(1), 53–61.

Persoon, E. and Fu, K.-S. (1977). “Shape discrimination using Fourier descriptors.” *IEEE Transactions on systems, man, and cybernetics*, 7(3), 170–179.

Taneja, S., Akinci, B., Garrett, J., Soibelman, L., and East, B. (2011). “Transforming ifc-based building layout information into a geometric topology network for indoor navigation assistance.” *Computing in Civil Engineering (2011)*, 315–322.

Ueda, N. and Suzuki, S. (1993). “Learning visual models from shape contours using multiscale convex/concave structure matching.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4), 337–352.

Vatti, B. R. (1992). “A generic solution to polygon clipping.” *Communications of the ACM*, 35(7), 56–63.

Wallis, W. D., Shoubridge, P., Kraetz, M., and Ray, D. (2001). “Graph distances using graph union.” *Pattern Recognition Letters*, 22(6), 701–704.

Weber, M., Liwicki, M., and Dengel, A. (2012). “Faster subgraph isomorphism detection by well-founded total order indexing.” *Pattern Recognition Letters*, 33(15), 2011–2019.

Wu, I.-C. and Hsieh, S.-H. (2007). “Transformation from ifc data model to gml data model: methodology and tool development.” *Journal of the Chinese Institute of Engineers*, 30(6), 1085–1090.

List of Tables

1 Survey of attributes used in current dwelling unit databases and websites 28

2 Attributes for similarity calculation of SDUs 29

3 Weights of the best run of the proposed GA 30

4 Convergence details of the best run of the proposed GA 31

5 Time consumptions for calculating similarities (Unit: ms) 32

TABLE 1. Survey of attributes used in current dwelling unit databases and websites

Source	Nbimer website	Chinauhu website	Kujiale website	Zhibeizhen database	CSCEC database	Tsinghua database
Details						
Type	website	website	website	database	database	database
Publicly accessible	◦	◦	◦	×	×	×
Storage	□.dwg	□.dwg	□.png /3d model	□.dwg	□.rvt	□.ifc
Attributes						
Name	◦	◦	◦	◦	◦	◦
Area	◦	◦	◦	◦	◦	◦
Bedroom count	◦	◦	◦	◦	◦	◦
Living room count	◦	◦	◦	◦	◦	◦
Bathroom count	◦	◦	◦	◦	◦	◦
Breadth	×	◦	◦	◦	◦	◦
Depth	×	◦	◦	◦	◦	◦
Building height	×	◦	◦	◦	◦	◦
Elevator/stair count	×	◦	◦	◦	◦	◦
Location	×	◦	◦	◦	◦	◦
Code	×	◦	◦	◦	◦	◦
Developer	×	×	◦	◦	◦	◦
Project	×	×	◦	◦	◦	◦
Category	×	×	◦	◦	◦	◦
Orientation	×	×	×	×	×	◦
Price	×	×	×	◦	◦	×

TABLE 2. Attributes for similarity calculation of SDUs

Attributes	Type	Range (Nominal attributes) / Unit (Numeric attributes)	Calculation
Room count	Numeric	none	Cosine similarity

Room area	Numeric	m^2	Cosine similarity
Building's height⁑	Nominal	High (>10 floors), Middle (4-9 floors), Low (1-3 floors)	Ratio of matches
Climate region⁑	Nominal	Cold (Region II), Warm (Region III), Hot (Region IV)	Ratio of matches
Ratio of elevators or stairs per dwelling unit⁑⁑	Nominal	N elevators/stairs per M dwelling units	Ratio of matches
Orientation	Nominal	N, S, E, W, NE, NW, SE, SW	Ratio of matches

⁑ According to classification in '*Chinese Code for Design of Civil Buildings* (GB50352-2005)'.

⁑⁑ An index commonly used in China to reflect a building's residential density.

TABLE 3. Weights of the best run of the proposed GA

RCS	RAS	NAS	ACS	ADS	SS
0	0.114249037	0.534659820	0.311296534	0	0.039794608

TABLE 4. Convergence details of the best run of the proposed GA

Iteration	Cost (Unfitness)	Average Deviation (Training Set)	Variance (Training Set)	Average Deviation (Test Set)	Variance (Test Set)
1046	184.2608294	0.109716777	0.018612205	0.108075706	0.018796000

TABLE 5. Time consumptions for calculating similarities (Unit: ms)

Similarity	Time consumption I[*]	Time consumption II[**]	Time consumption for each SDU pair				
			Avg.	Med.	Min.	Max.	SD.
RCS	110.5	589.6	0.0204	0.0171	0.0723	11.7170	0.0078
RAS	51.7	536.3	0.0186	0.0138	0.1214	12.8321	0.0069
NAS	0.5	3.1	0.0001	0.0000	0.0000	0.0135	0.0002
ACS	555.9	12082.0	0.4181	0.2627	0.4702	8.0569	0.0123
ADS	1377.4	4865.3	0.1683	0.1067	0.5522	89.2192	0.0135
SS	1451.6	6013.2	0.2081	0.1799	0.8356	137.5379	0.0189
Combined	3547.1	24089.6	0.8336	0.5973	1.6405	253.3339	0.0791

[*] Time consumption for calculating similarities without repetition (14365 pairs). Parallel computing is used.

[**] Time consumption for calculating all possible combinations (28900 pairs). For SDUs A and B, sim(A,B) and sim(B,A) are both calculated. Parallel computing is not used.

List of Figures

1 The framework of the multi-aspect approach for the similarity evaluation of SDUs . 34

2 Topology representations in IFC 35

2(a) Adjacency 35

2(b) Accessibility 35

3 Preprocesses for calculating topological similarities 36

3(a) Indoor corridors 36

3(b) Multifunctional rooms 36

3(c) Virtually separated rooms. 36

3(d) The orientation of rooms 36

4 A topological similarity example 37

5 Shape similarity based on the overlap area of 2 SDUs 38

6 The GA-based approach for customizing a combined similarity 39

7 Operators of the GA 40

7(a) Mutation operator 40

7(b) Crossover operator 40

8 The interface of the design tool based on the similarity of SDUs 41

9 Time consumption for calculating similarities as the size of the SDU database increases 42

9(a) RCS & RAS 42

9(b) NAS 42

9(c) ADS & ACS 42

9(d) SS 42

10 An example of the design recommendation in the floor-plan design 43

10(a) SDU search and replacement 43
10(b) Attribute-based search 43
10(c) Similarity-based search 43

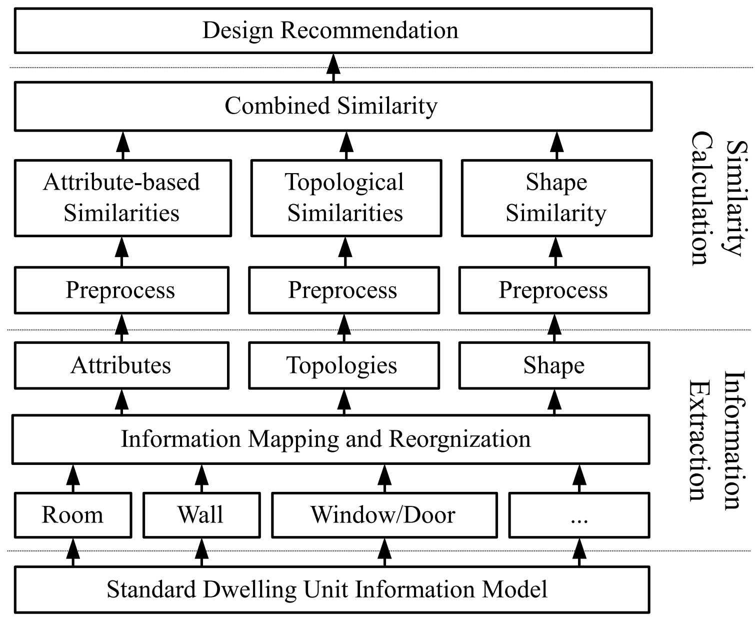
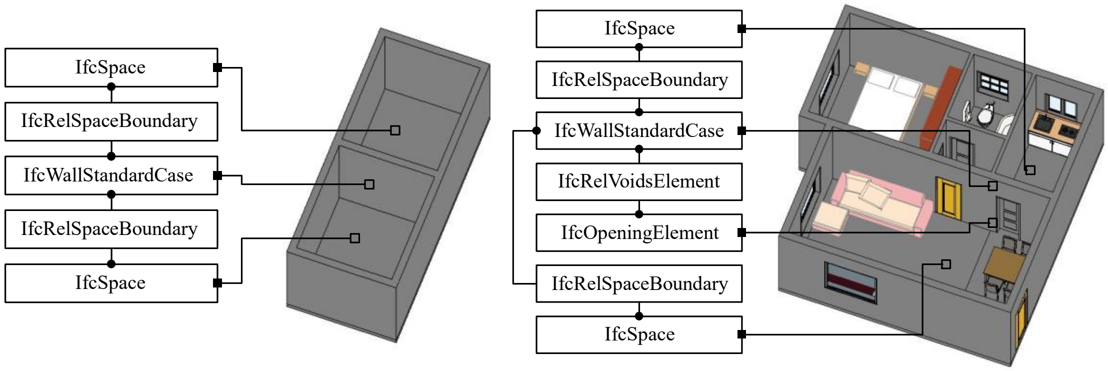
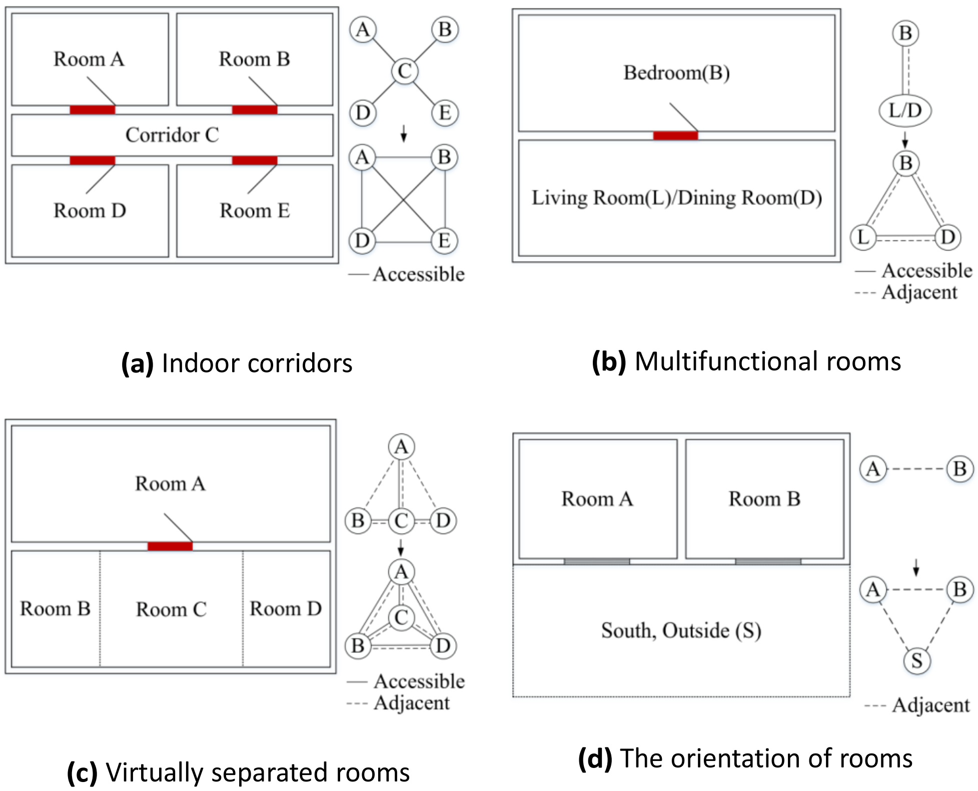


Fig. 1. The framework of the multi-aspect approach for the similarity evaluation of SDUs



(a) Adjacency
(b) Accessibility

Fig. 2. Topology representations in IFC



(a) Indoor corridors

(b) Multifunctional rooms

(c) Virtually separated rooms

(d) The orientation of rooms

Fig. 3. Preprocesses for calculating topological similarities

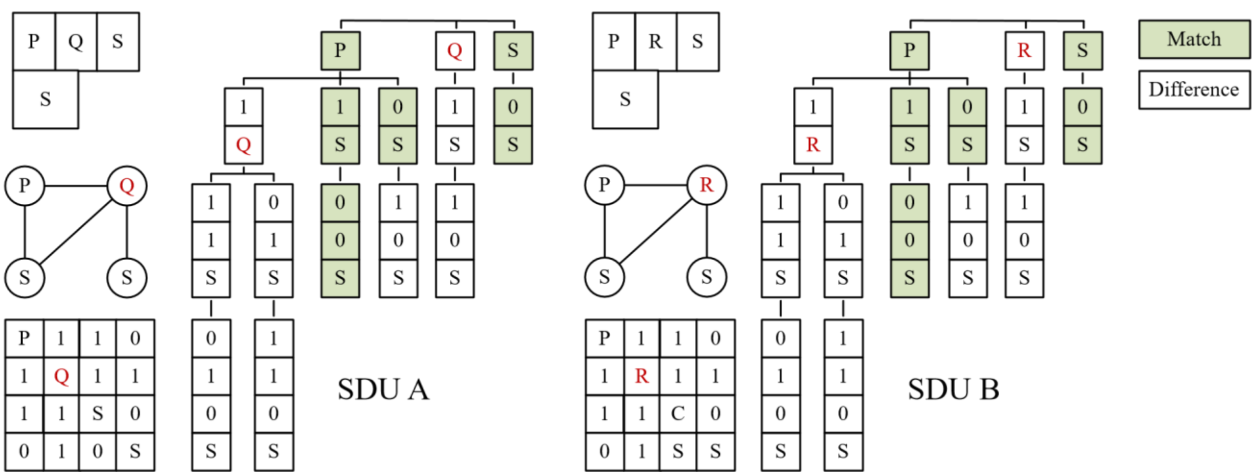


Fig. 4. A topological similarity example

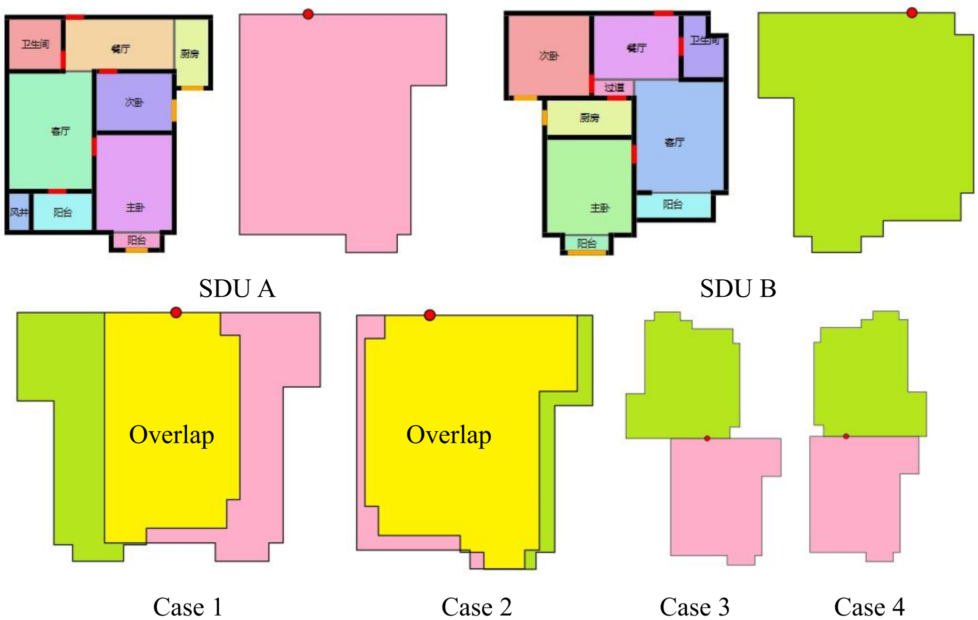


Fig. 5. Shape similarity based on the overlap area of 2 SDUs

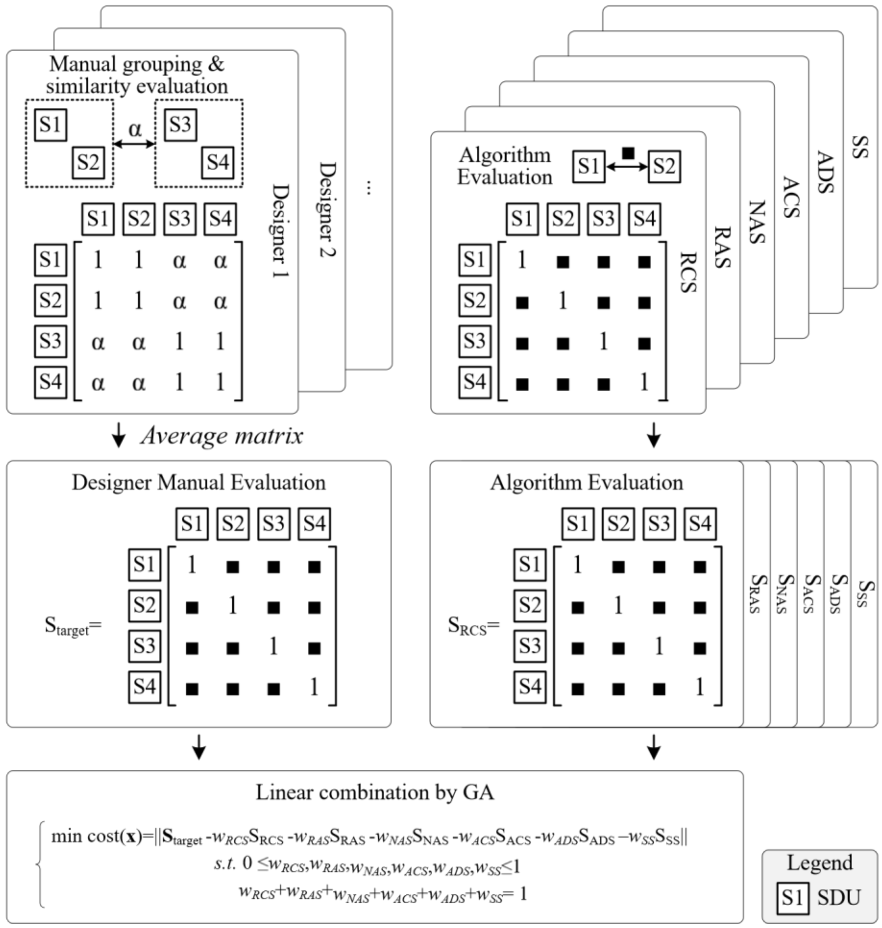
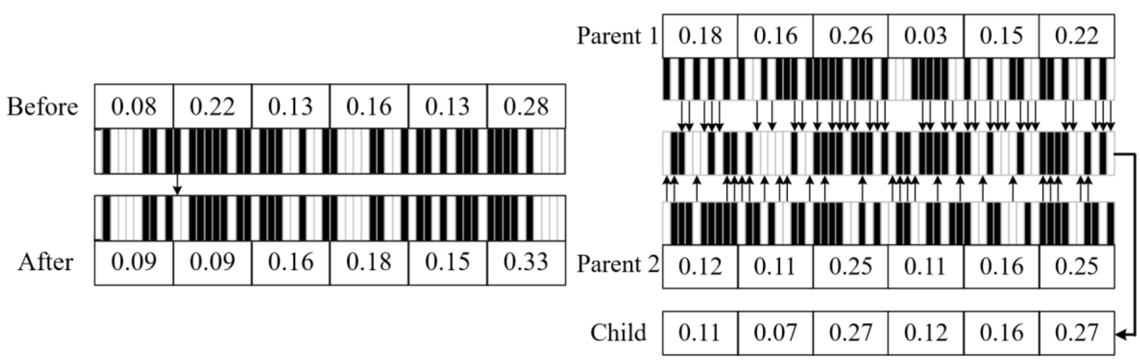


Fig. 6. The GA-based approach for customizing a combined similarity



(a) Mutation operator

(b) Crossover operator

Fig. 7. Operators of the GA

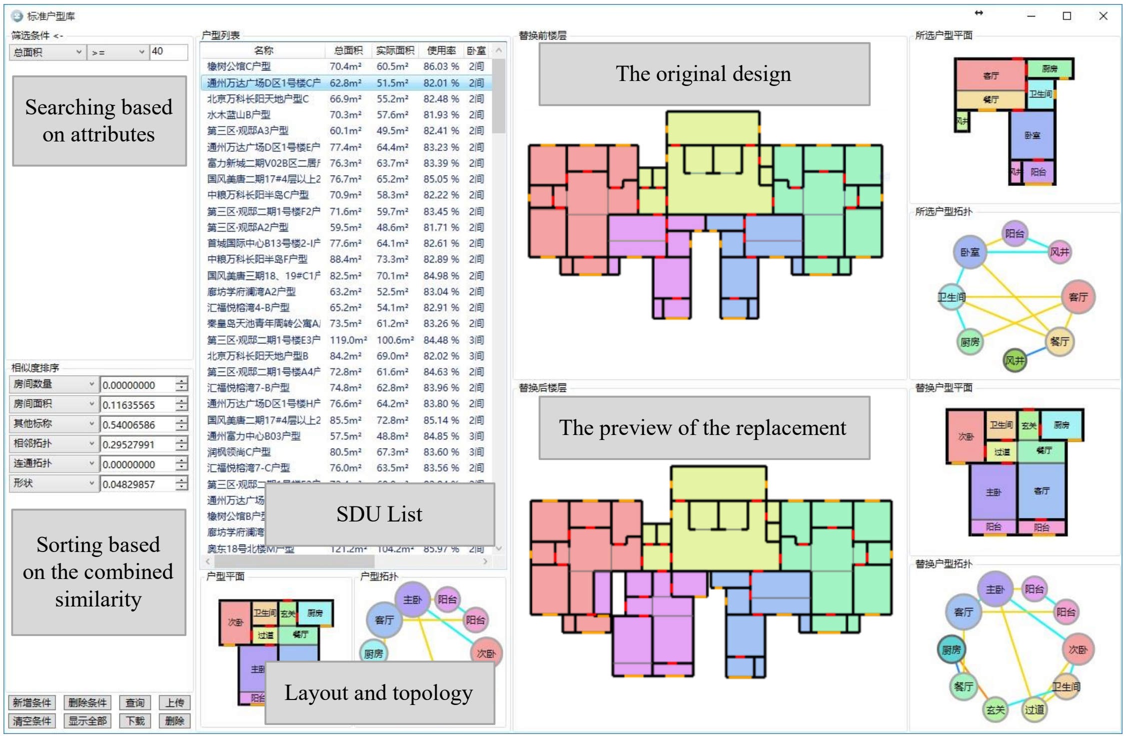


Fig. 8. The interface of the design tool based on the similarity of SDUs

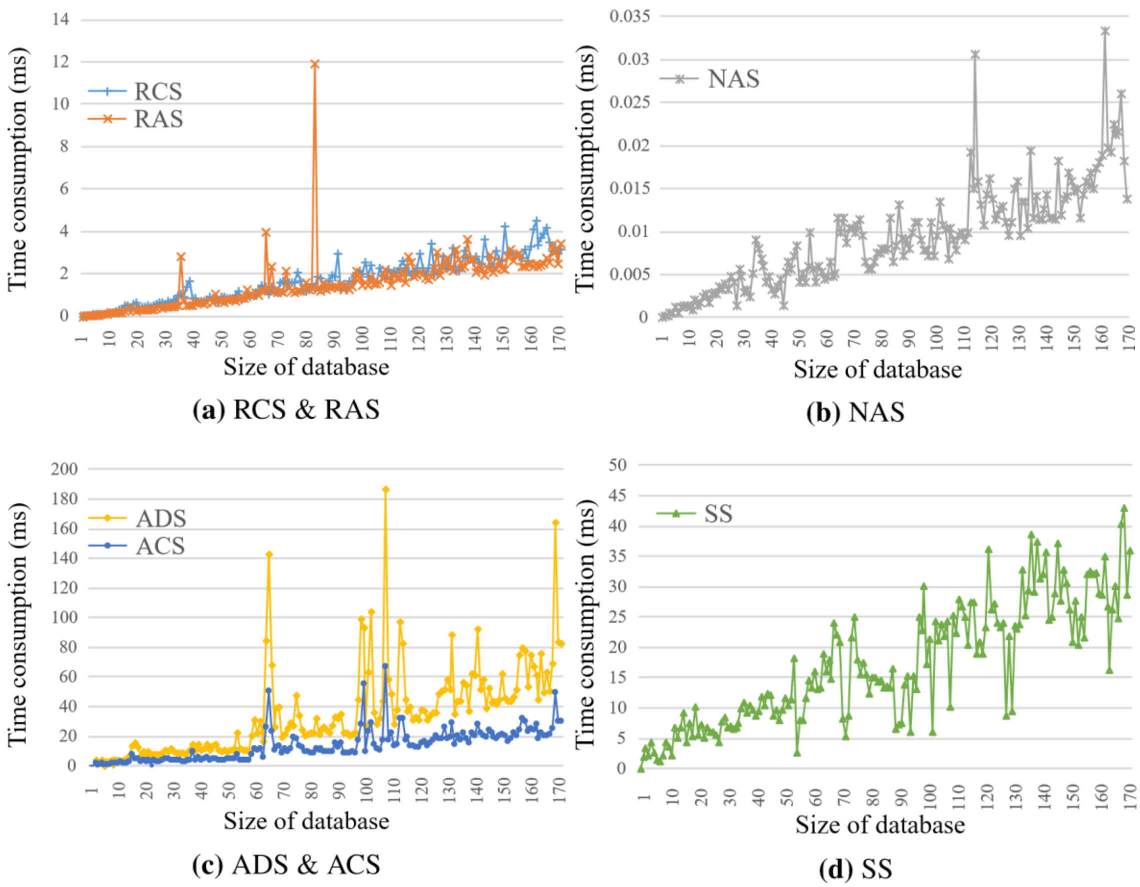
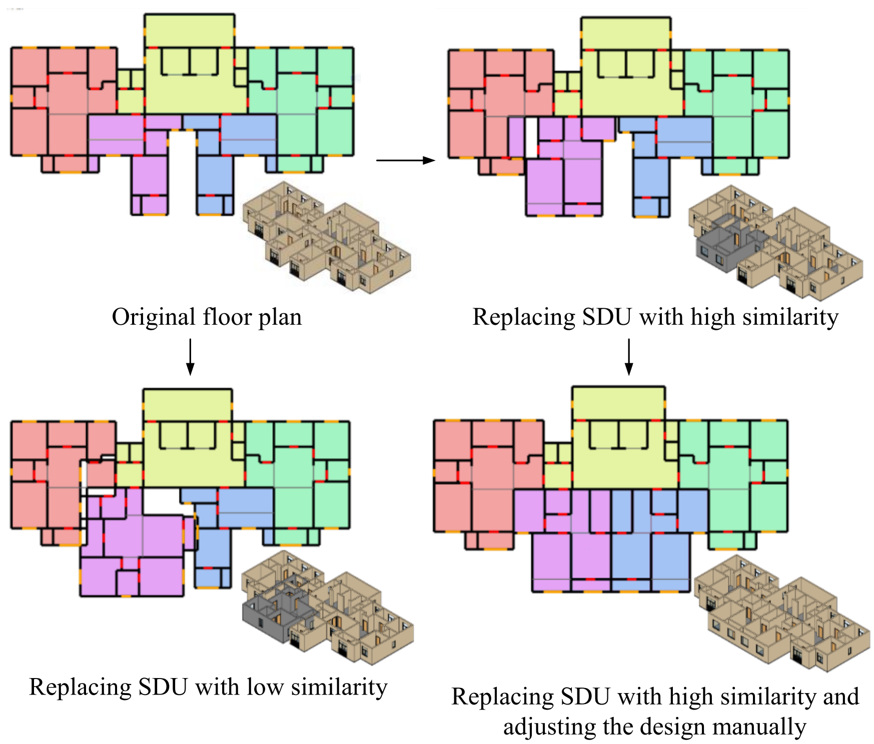


Fig. 9. Time consumption for calculating similarities as the size of the SDU database increases



(a) SDU search and replacement

Values of Search Results Rank of

Bedroom Values of count 1

foundRCS 26 2 Not

Attribute-based search

Living room count 1 1 33 SDUsRAS 29

Toilet count 1 1 33 SDUsNAS 2

Elevators per dwelling unit 1/4 1/4 17 SDUsADS 29

Orientation South South 21 SDUsACS 10

Region Warm Cold Not foundSS 2

Building height High High 29 SDUs**Combined similarity9**

Area 44 m^2(S) 63 m^2(M) Not found

Similarity-based search

(b) Attribute-based search **(c)** Similarity-based search

Fig. 10. An example of the design recommendation in the floor-plan design