COMPUTER-AIDED DESIGN & APPLICATIONS, 2017
VOL. 14, NO. 4, 422–435
http://dx.doi.org/10.1080/16864360.2016.1257185

Measuring 3D process plant model similarity based on topological relationship distribution

Rui Wen [a], Weiqing Tang [b] and Zhiyong Su [c]

[a]School of Computer Science and Engineering, Nanjing University of Science and Technology; [b]Institute of Computing Technology, Chinese Academy of Sciences; [c]School of Automation, Nanjing University of Science and Technology

ABSTRACT

Current research on 3D model similarity mainly concentrates on shape feature extraction, as well as the measurement with respect to individual objects. However, a quintessential Process Plant Model (PPM) consists of thousands of geometric solids, such as reaction vessels, pipelines and supports. These solids are interconnected following specific engineering rules. When measuring PPM similarity, geometric features as well as engineering attributes should both be taken into account. Therefore, existing shape based methods are inapplicable. As topological relationships are the core of PPM, this paper applies graph similarity into PPM and presents a new similarity measurement based on Topological Relationship Distribution (TRD) feature. First, a Relation Tree (RT) model for extracting TRD is proposed. The RT model attains and stores a PPM's relationship statistics by traversing all attributes and topological relationships of components. Second, as to achieve the comparable feature vector, standardization is performed via mapping relationship statistics into vector space. Last, a hybrid similarity function combining both directional and numerical differences in feature vectors is proposed to evaluate PPM similarity. Due to the exploitation of topological features and engineering attributes rather than raw directions and positions, the TRD based method embraces the properties of translation, rotation and similarity transformation consistency. Experimental results demonstrate the feasibility and accuracy of the proposed framework.

KEYWORDS

Process plant; 3D model; similarity measure; feature extraction; graph similarity; topological relationship distribution

1. Introduction

Computer aided process plant design involves a wide range of specializations and lots of manpower, during which many intermediate models are generated. With the expanding investment in process plant, the number of Process Plant Models (PPMs) is exploding. In practice, experienced engineers usually construct a plant based on realized plants or already developed projects. Good solutions from old projects may serve as examples for future work and could even be stored electronically [4]. Thus, model retrieval is of great significance in shortening design cycle and improving efficiency.

Due to the difficulty of an accurate 3D model text acquisition method and lacking of standardization, a filename based model retrieval system can no longer meet enterprise-level needs. Comparatively, through directly exploiting the internal characteristics of 3D models, content-based retrieval techniques serve as a useful tool for locating existing models. Furthermore, it can assist designers in maximizing management and reuse of existing resources in future design:

- By referring to the retrieved historical models, designers are able to enhance model reuse and build engineering models in a more efficient way;
- By using revised models for reference, some unnecessary mistakes can be avoided and model quality is better ensured;
- Model retrieval techniques make engineering design no longer merely rely on enterprise or individual experiences, which could be a major boost to the development of design competence.

The underlying problem in model retrieval is similarity measurement. Current researches on 3D model similarity [25,21] focus on shape metric. In contrast, PPM similarity presents the following variants, making it more complex:

(1) **Primary Topology**. Process plant design concentrates on the structural and topological information of targets, demanding model's capability to precisely describing all components' spatial positions

CONTACT Rui Wen wenruity@163.com

and mutual relations. Consequently, PPM similarity requires less of component's geometrical shapes [9]. In other words, even though there are some differences in geometry (pipeline sizes, for example), two models should still be considered highly similar as long as they have the same topology.

(2) **Extra engineering attributes.** A PPM can be described by three kinds of information [22]: geometrical information for describing shapes and 3D spatial positions, topological information for describing adjacency relations among components, and engineering attributes for describing design constraints, engineering disciplines, etc. Accordingly, besides geometric features and topology structures, engineering attributes should also be considered in PPM similarity measure.

Unfortunately, similarity research towards PPMs is relatively scarce. As PPM is a parametric data structure and its core lies in topology, we represent a PPM as a spatial attribute graph. Components can be expressed as nodes while topological relationships as edges. In that sense, the problem of PPM similarity boils down to attribute graph similarity.

Process plants with different application backgrounds could have differences in engineering design rules, which are mainly reflected in topological relationships between components. As a result, their distributions of topological relationships should be different accordingly. Considering these characteristics, we attempt to map the frequency of each relationship type to vector space and exploit Topological Relationship Distribution (TRD) as PPM's feature. In this step, the attribute graph similarity is boiled down to a vector similarity/distance measure.

Our algorithm is consisted of three parts, listed as follows:

(1) **Feature extraction.** A Relation Tree (RT) model is proposed to transform the attribute graph into a tree. It is responsible for storing relationship statistics and extracting TRD features.

(2) **Feature standardization.** Because models' relationship types can have different scales, feature standardization is needed in order to make feature vectors comparable.

(3) **Similarity calculation.** According to the specialized characteristics of PPMs, we propose a hybrid similarity function to calculate vector similarity in both numerical and directional aspect.

The rest of this paper is organized as follows. Related work is reviewed in Section 2; Section 3 gives a brief introduction to PPMs; Section 4 describes the proposed framework; experiment and performance of the proposed framework are presented in Section 5, followed by conclusions and future work in Section 6.

2. Related work

Similarity research on general 3D models and product CAD models makes significant contributions to the development of model retrieval techniques. Since a PPM can be represented as a spatial attribute graph, existing algorithms for graph similarity can also be introduced. In this section, we give a brief review of the state-of-the-art similarity research in general, as well as in CAD and graph areas.

2.1. Similarity measure for general 3D models

General 3D models are models that human can make contact with in daily life, such as furniture 3D models, animal 3D models, etc. Their shapes and topologies are complex and built-up by triangular meshes. After years of research, the similarity measurements for general 3D models mainly fall into three categories: feature based methods, graph based methods and geometry based methods [25].

According to the types of shape features used, feature based methods can be divided into global features, global feature distributions, spatial maps and local features. These methods all focus on the appearing shapes of 3D models. Zhang and Chen [30] describe methods to compute global features, such as volume, area, statistical moments, and Fourier transform coefficients efficiently. Osada *et al.* [19] introduce and compare shape distributions, which measure properties based on distance, angle, area and volume measurements between random surface points. Ankerst *et al.* [1] use shape histograms as a means of analyzing the similarity of 3D molecular surfaces. Shum *et al.* [6] use a spherical coordinate system to map the surface curvature of 3D objects to the unit sphere.

Graph based methods can fall into model graphs, Reeb graphs and skeletons. They attempt to extract a geometric meaning from a 3D shape by analyzing how shape components are linked together. Elinson *et al.* [11], Cicirello and Regli [7] compare the similarity of solid models by comparing their associated manufacturing plans. Biasotti *et al.* [3] introduce a framework for the matching of 3D shapes represented by Reeb graphs. Sundar *et al.* [24] encode the geometric and topological information in the form of a skeletal graph and use graph matching techniques to match the skeletons and to compare them.

Geometry based methods are divided into view based similarity, volumetric error based similarity, weighted

point set similarity and deformation based similarity. A major property of geometry based methods is their robustness to coordinate rotation and transformation. Loffler [17] applies view based similarity to retrieve 3D models using a 2D query interface. Novotni and Klein [18] describe a geometry similarity approach to 3D shape matching based on calculating a volumetric error between one object and a sequence of offset hulls of the other object. Dey *et al.* [10] present a method to obtain a descriptor of a shape, given by a point sample, by first decomposing the shape into its components. Also, a number of methods [2] compare a pair of 2D shapes by measuring the amount of deformation required to register the shapes exactly.

2.2. Similarity measure for CAD models

In CAD area, mechanical product similarity is capturing the most attention. Different from PPMs, mechanical models are comprised of curves and surfaces. Consequently, its retrieval system concentrates on the research of surface feature computation, free-form surface parameterization and mechanical specific standardization.

Cicirello *et al.* [7] propose a feature-based matching method for mechanical models. Firstly, machining feature extraction is performed to map the solid model to a set of STEP AP 224 machining features; Then, a model dependency graph from the set of machining features is constructed; at last, the nearest neighbors to the query graph is found using an iterative improvement search across a database of other models.

El-Mehalawi *et al.* [12,13] introduce a geometry and topology based mechanical model database system. In this paper, a model representation is defined using Standard for the Exchange of Product (STEP). Through this model, attribute graph construction, model indexing, retrieval and matching can be efficiently implemented.

As indicated by Zhang *et al.* [31] and Wang *et al.* [26], each CAD model can be represented as an attribute adjacency graph by extracting B-Rep information. Thus model-wise local similarity is measured using their common sub-graphs.

2.3. Similarity measure for graphs

Graph similarity can be classified into two categories [16]:

(1) **With known node correspondence**. In this situation, adjacency relation based difference in each dual node is used to measure graph structural similarity. The best approaches are Graph Edit Distance [29,32] and Maximum Com mon Sub-graph [20].

Zheng and *et al.* [32] retrieve similar graphs by a given query graph under the constraint of the minimum edit distance. They derive a lower bound, branch-based bound, which can greatly reduce the search space of the graph similarity search. At the same time, they also propose a tree index structure, namely b-tree, to facilitate effective pruning and efficient query processing.

Raymond and *et al.* [20] propose a graph distance metric based on the maximal common subgraph. Given a minimum similarity coefficient, this procedure consists of an initial screening process to determine whether it is possible for the measure of similarity between the two graphs to exceed the minimum coefficient. Then a rigorous maximum common edge subgraph detection algorithm is proposed to compute the exact degree and composition of similarity.

Nonetheless, computing the edit distance is a Nondeterministic Polynomial (NP) problem [8] and maximal common sub-graph is also NP-complete, implying the infeasibility to efficiently compute existing metrics for general graphs.

(2) **With unknown node correspondence.** This situation highlights a much higher complexity, which has been addressed by two solutions: (i) Constructing a point-to-point mapping for nodes with the same label in both graphs, then applying the known dual node methods; (ii) Applying certain identical feature extraction method to both graphs, then feature based difference is used against graph similarity.

For example, Yan and *et al.* [27] propose a feature-based graph indexing method. They investigate the issues of indexing graphs and propose a novel solution by applying a graph mining technique. Different from the existing path-based methods, their approach which is called gIndex, makes use of frequent substructure as the basic indexing feature. Frequent substructures are ideal candidates since they explore the intrinsic characteristics of the data and are relatively stable to database updates.

Zhao and *et al.* [33] introduce a new cost-effective graph retrieval method based on frequent tree-features of the graph database. They analyze the effectiveness and efficiency of tree as indexing feature from three critical aspects: feature size, feature selection cost, and pruning power. At last, they verify that tree is a better choice than graph for retrieval purpose.

Accordingly, feature based similarity research directions toward feature extraction and similarity computation. Especially for large graphs, feature based graph similarity has an advantage of calculation efficiency, instead of the complicate comparisons of each dual node.

3. Process plant models

A process plant consists of a set of reaction vessels, pipelines and supports, which are designed for making

chemical or physical manufactured products [9]. We will first give a brief introduction to the composition, topology and engineering attribute of PPMs.

3.1. Composition of process plant models

A typical PPM consists of thousands of basic components, including equipment, pipelines (i.e. pipes and piping components), valves, instruments, etc. [15], as shown in Fig. 1. For example, the model shown in Fig. 1(d) consists of 41,569 components. These components are normally composed of fourteen basic entities, such as cylinder, scylinder, prism, econe, concone, squcir, squcone, box, torus, squtorus, sphere, wedge, saddle and oval, as shown in Fig. 2. A PPM is designed to be parametric [23]. Consequently, all components follow the same serialized standard which records information such as geometric parameters, types and other engineering attributes, rather than raw meshes as in general 3D models.

Besides the above basic types, there are some other specialized components oriented toward different engineering applications, such as electrical appliance, civil engineering, etc. Many engineering CAD softwares are equipped with specialized database management systems.

3.2. Topology of process plant models

Topology is the kernel of PPMs. In order to satisfy construction, operation, maintenance and safety requirements, a PPM's topology is concentrated with the most

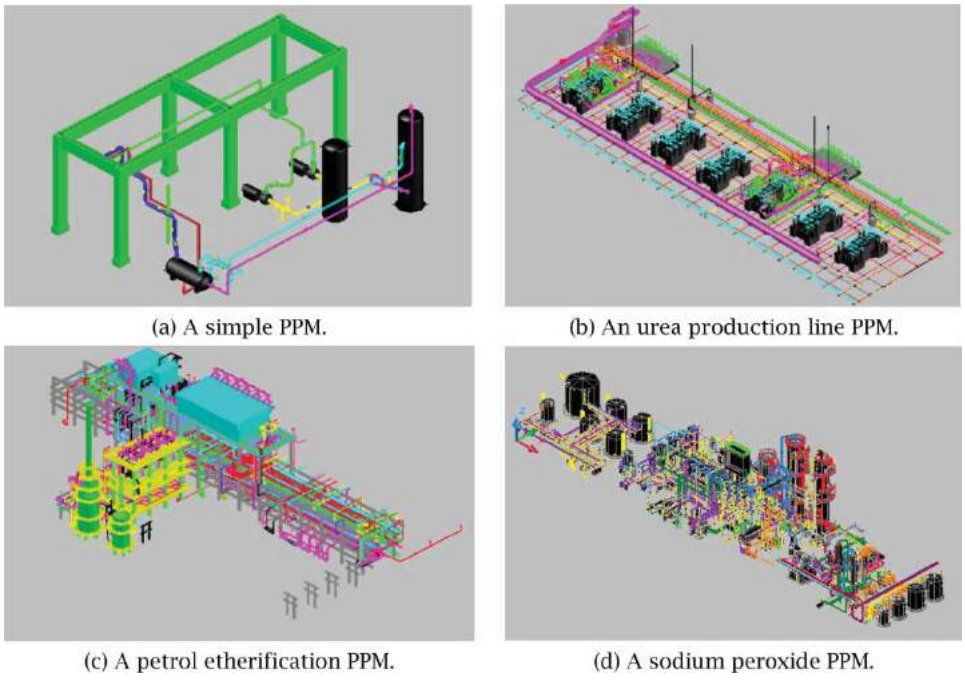

Figure 1. PPMs from different engineering applications.

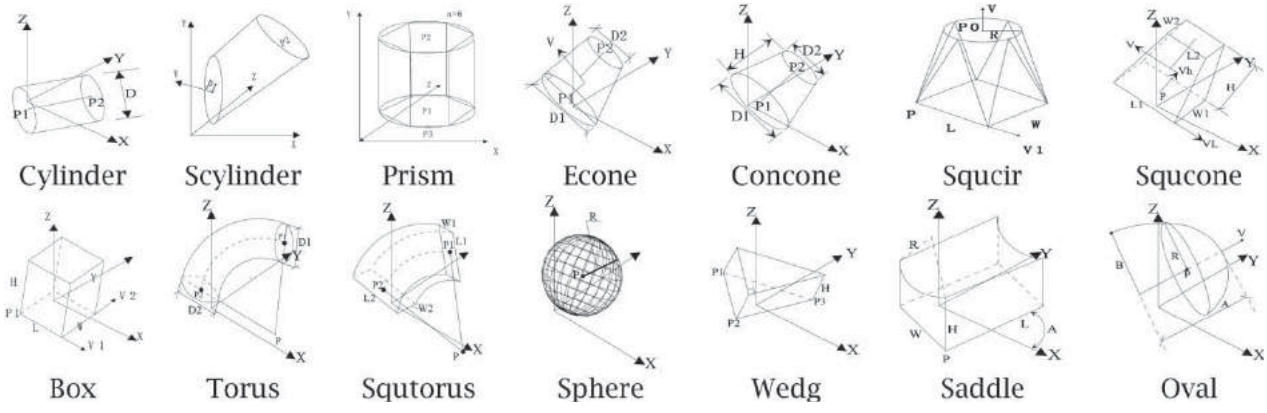

Figure 2. Basic entities used in PPMs.

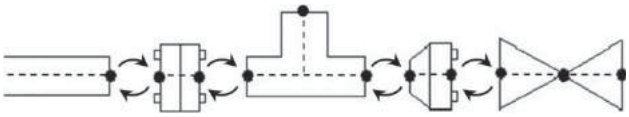

Figure 3. A simple duality point and smart line based pipeline example.

economical spatial arrangement and interconnections of process vessels and equipment.

Because of the variety of design systems, there exist several topology representations of PPMs. Duality point and smart line based approach, as well as branch based approach, are widely used now. In the former approach, the duality point is used to describe the topological adjacency between piping components, such as pipe segment, tee, valve, equipment, etc. The smart line is a special line which reflects topological attributes of the pipes and components. Usually, the smart line is represented by the centerline of the pipeline. A simple example of duality point and smart line is illustrated in Fig. 3. In this figure, the black point is duality point and the dashed is smart line. Note that a pipeline may contain a certain number of pipes and components.

In the branch based approach, each pipeline is designed to have several branches and these branches are connected by concrete pipes. A simple branch based pipeline is illustrated in Fig. 4.

Notably, although design systems come with different topology representations, they all have the same essence, that is: a PPM is structured data which can be understood as a complex spatial graph.

3.3. Engineering attribute of process plant models

The ultimate goal of computer aided process plant design is to automatically generate isometrics, orthographic and other construction documents. These documents are directly exchanged with 3D models and are used to guide the actual engineering constructions. Engineering attribute plays an important role in generating accurate construction document. The reason is that it contains information about design constraints, engineering disciplines, etc., which are essential to the precise descriptions of components and relationships.

There are many kinds of engineering attributes. Type attribute stands out as the component's unique identifier in engineering databases. Direction flow attribute characterizes the special constraint of topological relationships. Other attributes, like material, pipeline level, facing type, wall thickness, etc., fully describe the process plant design and make it better understandable for builders. In addition, it's worth emphasizing that engineering attributes should be used in accordance with the specific applications.

4. Proposed similarity measure framework

4.1. Background

As mentioned in Section 3, because of its rigorous topology and parametric data structure, a PPM can be expressed as a spatial attribute graph. Each component is represented as a node and topological relationship as an edge. Now a PPM M can be described as $M = \{\mathbf{C}, \Sigma_\mathbf{C}\}$, where $\mathbf{C}$ is the component set and $\Sigma_\mathbf{C}$ is the relationship set. What's more, $\Sigma_\mathbf{C}$ can also be described as

$$\Sigma_\mathbf{C} = \{(c_i, c_j, A_{ij}) | c_i \in \mathbf{C}, c_j \in \mathbf{C}, 1 \leq i, j \leq n\} \quad (4.1)$$

where A_{ij} is the connection attribute of c_i and c_j, and $n = |\mathbf{C}|$. Based on this representation, PPM similarity can be evaluated by the similarity of their corresponding attribute graphs.

As presented in Tab. 1, each model contains large numbers of components and topological relationships. To ensure the efficiency of similarity measure, we try to tackle this problem through feature extraction, as described in the second solution.

In process plant, design rules reflecting specialized knowledge and application background are embodied in topological relationships. Components should be connected following the specific engineering rules.

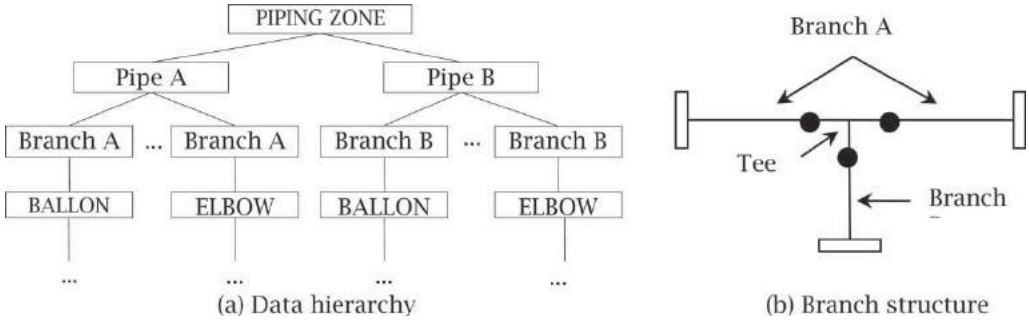

Figure 4. A simple branch based pipeline example.

Table 1. Statistics of models in Fig. 1.

Model	Components	Component types	Relationships	Relationship types
Fig. 1(a)	308	22	602	33
Fig. 1(b)	9,795	137	19,044	173
Fig. 1(c)	22,596	187	45,373	224
Fig. 1(d)	41,569	774	83,359	838

Therefore, PPMs from different application backgrounds should be different in the distributions of relationships. Consequently, we use Topological Relationship Distribution (TRD) as the major feature and propose a TRD based similarity measure method. In this method, we map the frequency of each relationship type into vector space and calculate the feature vector similarity to measure PPM similarity.

In order to record a PPM's relationships and compute the relationship distribution similarity, we present a Relation Tree (RT) model and a hybrid similarity function respectively. Meanwhile, because of the uncertainty and appearance sequences of relationship types, we need to standardize TRDs and acquire comparable feature vectors.

4.2. Feature extraction

With the purpose of extracting TRD feature, a RT model is proposed. Type attribute is the unique identifier for distinguishing different geometric components in engineering databases. Also, it is designers' major principle in selecting the target component. Accordingly, we uniquely identify each node by its component's type resided. Through traversing a PPM's components and topological relationships, a RT can record the amounts of relationship types.

As depicted in Fig. 5, the root node is denoted as *Root* and a node identified by *Type* as node(*Type*). A RT contains four aspects of information:

(1) $Level_1$ is the set of component types in the current model, let's say $Type_i$;
(2) $Level_2$ includes component types which connect with $Type_i$ in $Level_1$, let's say $R - Type_{ij}$;
(3) The linking edge A_{ij} between $Type_i$ and $R - Type_{ij}$ contains some design constraints, flow direction for example;
(4) Each node in $Level_2$ is equipped with a counter $N(Type_i, A_{ij}, R - Type_{ij})$.

Also, we can regard the RT as a 3-dimensional sparse matrix:

- $axis(i)$: containing all node types;
- $axis(j)$: containing all different kinds of attributes;
- $axis(k)$: containing all node types.

In this matrix, each entry $N(i, j, k)$ stores the number of occurrences of the corresponding relationship.

Given a PPM $M = \{\mathbf{C}, \Sigma_\mathbf{C}\}$, its RT construction process is a simple traversal of all relationships to populate the sparse matrix. That is, for each topological relationship $(c_i, c_j, A_{ij}) \in \Sigma_\mathbf{C}$, if the type of c_i is $Type_i$, c_j is $R - Type_{ij}$ and their connection attribute is A_{ij}, we increment its counter $N(Type_i, A_{ij}, R - Type_{ij})$ by one. After this traversal, the TRD of current PPM can be represented as a vector V_{RT}:

$$\begin{aligned} V = (&N_{11}, N_{12} \ldots, N_{1m_1}, \\ &N_{21}, N_{22} \ldots, N_{2m_2}, \\ &\ldots\ldots \\ &N_{n1}, N_{n2} \ldots, N_{nm_n}). \end{aligned} \tag{4.2}$$

4.3. Feature standardization

During RT construction, types and sequences of topological relationships are both non-deterministic. Assuming vectors V_{RT_x} and V_{RT_y} are derived from RTs of PPMs M_x andM_y, with a corresponding length of l_x and l_y. As shown in the left of Fig. 6, components with the same position of V_{RT_x} and V_{RT_y} may stand for different

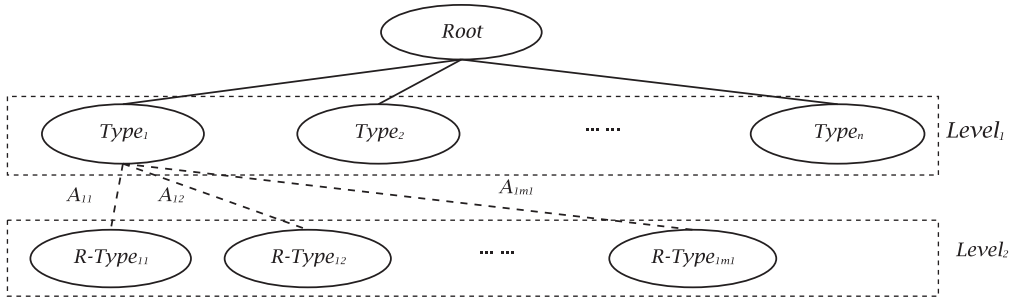

Figure 5. A simple diagram of relation tree.

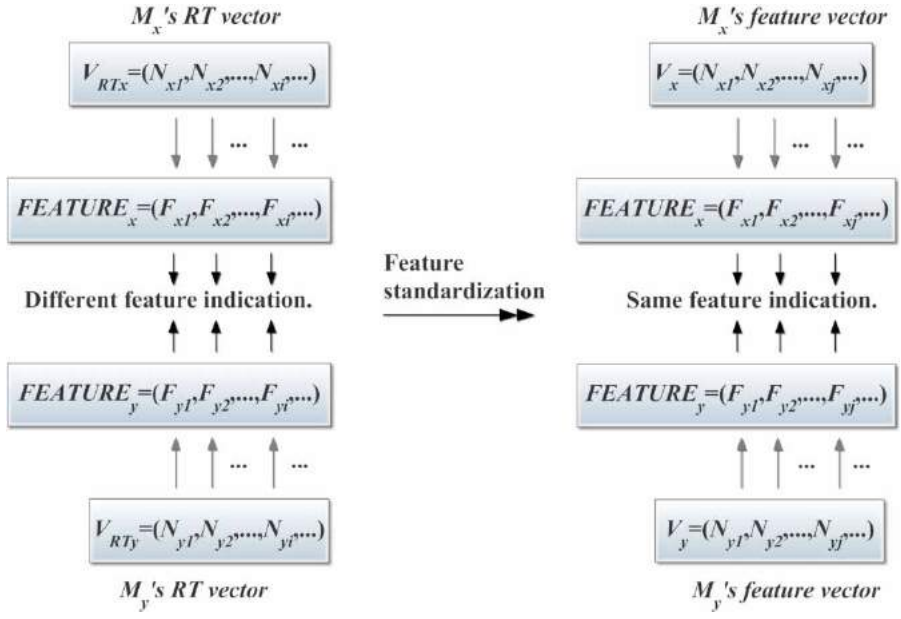

Figure 6. A diagram of feature standardization.

features. Consequently, these two vectors are not directly compared.

In order to make features comparable, feature standardization is indispensable [14]. Let l'_x and l'_y be the new lengths of feature vectors V_x and V_y. S is the quantity of relationship types of both M_x and M_y. As the right of Fig. 6 shows, after feature standardization, the corresponding components of V_x and V_y indicate the same relationship type, resulting in $F_{xi} = F_{yi}$ and $l_x{}' = l_y{}' = S$.

The procedure of feature standardization is described as Fig. 7 shows. Feature vectors V_x and V_y are initialized to empty (i.e. length = 0). A set **T** is denoted to record all relationships in M_x and M_y. The size of **T** is S which is initialized to zero.

Step 1: Traverse nodes in RTs T_x and T_y:

- **Step 1.1:** If the traversal of T_x is finished, go to Step 1.2; otherwise, if the current relationship doesn't exist in **T**, we add this relationship to **T** and mark it with 1, then go to Step 1.1;
- **Step 1.2:** If the traversal of T_y is finished, go to Step 2; otherwise, if the current relationship doesn't exist in **T**, we add this relationship to **T** and mark it with 2, then go to Step 1.2; if the current relationship exists in **T** and is marked by 1, we modify its mark to 3 and go to Step 1.2;

Step 2: Traverse **T**and execute one of the following operations (i-iii) according to each relationship's mark, then go to Step 3:

i. If the current relationship is marked by 1, which implies that this relationship exists only in M_x, we assign its counter's value in V_{RT_x} to the S-th position of V_x, while 0 to the same position of V_y;
ii. If the current relationship is marked by 2, which implies that this relationship exists only in M_y, we assign its counter's value in V_{RT_y} to the S-th position of V_y, while 0 to the same position of V_x;
iii. If the current relationship is marked by 3, which implies that this relationship exists both in M_x and M_y, we assign their counter's values in V_{RT_x} and V_{RT_y} to the S-th positions of V_x and V_y respectively;

Step 3: After the above operation, we add 1 to S. If **T**'s traversal is finished, the standardization is completed; otherwise go to Step 2.

4.4. Similarity function

Various distance/similarity measurements are applicable in comparing vector similarity, among which Euclidean distance and Manhattan distance belong to the former and Cosine method belongs to the latter [5]. Euclidean distance has an amplification effect on greater componential differences. Manhattan distance summates each componential difference and has a fair treatment on each component. Cosine method, which calculates the cosine value of two vectors, is sensitive to their direction similarity.

Fig. 8 shows the distributions and similarity results of simple vectors V_x and V_y under different measurements. Results indicate that:

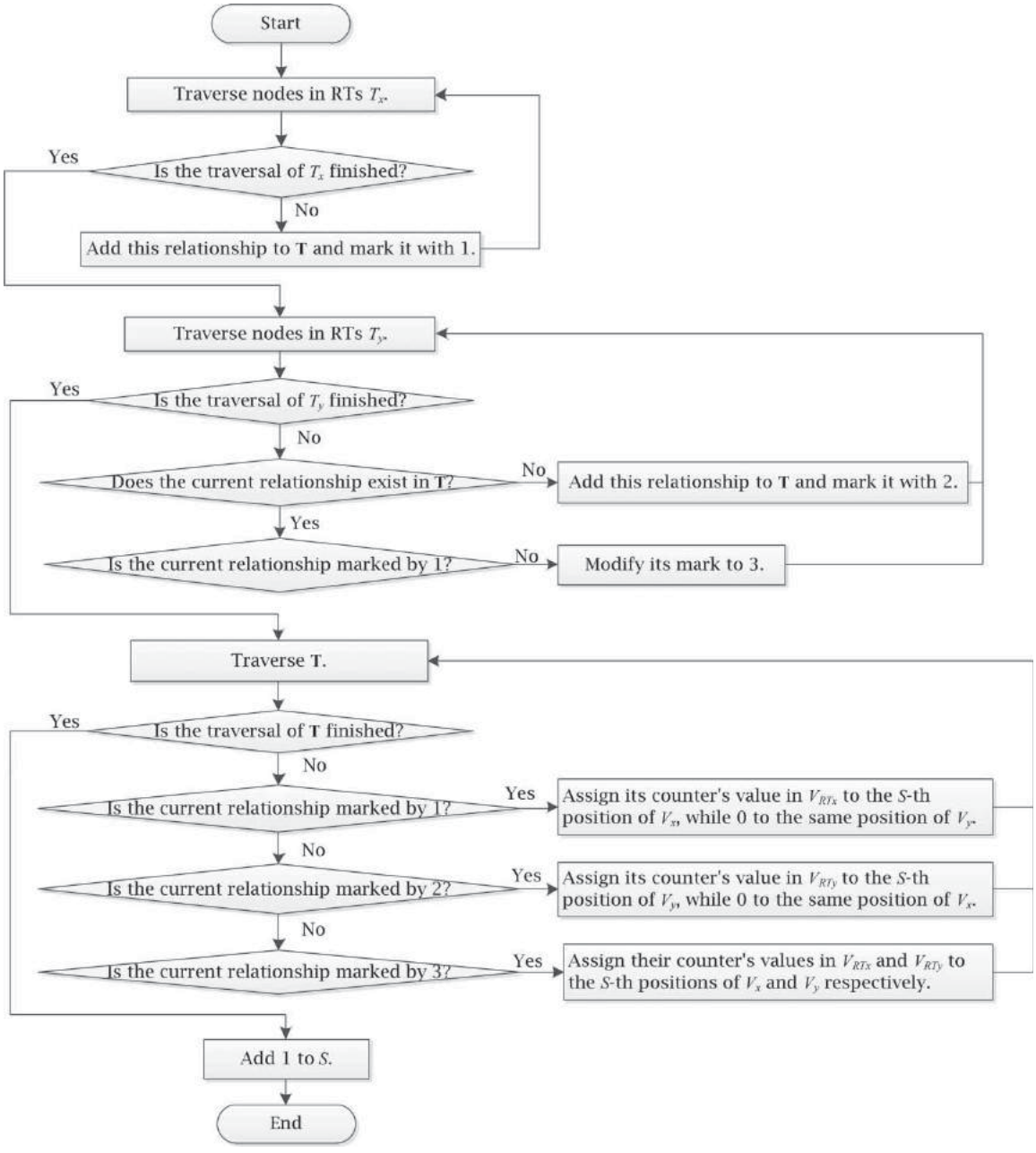

Figure 7. The flowchart of feature standardization.

(1) Cosine value reflects the proportional distribution similarity. In V_x and V_y, the proportions are both 2:4:3, which means the proportional distributions of components are the same. Thus, the proportional distribution similarity between V_x and V_y results in $\cos(V_x, V_y) = 1.0$.

(2) However, cosine similarity cannot reflect PPM similarity. In this paper, a PPM's TRD feature is represented by a vector, and each componential value indicates the frequency of corresponding relationship type. From V_x and V_y, we can obviously find that topological relationship types have significant differences in magnitude, but their cosine similarity is still 1.0. That is because the cosine value only measures how similar two vectors are based on their directions, but regardless of magnitude differences.

In conclusion, both numerical and directional differences cannot be discarded. Moreover, every topological relationship should be treated equally. To resolve this problem, we propose a hybrid similarity measurement to calculate vector similarity both in their magnitudes and directions.

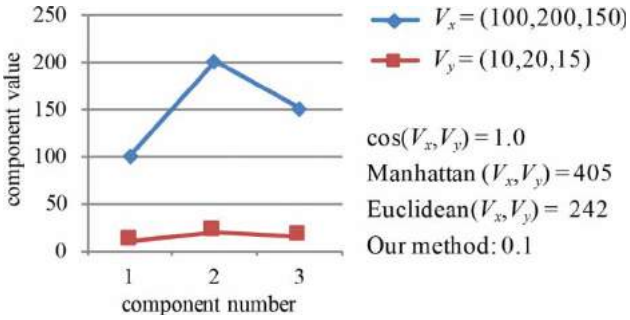

Figure 8. The component distributions of vectors V_x and V_y.

After the feature standardization described in Section 4.3, the final feature vectors of M_x and M_y are represented as $V_x = \{x_1, x_2, \ldots, x_S\}$ and $V_y = \{y_1, y_2, \ldots, y_S\}$ with an equal length S. Firstly, we use d_i to express the numerical difference ratio of components x_i and y_i,

$$d_i = \frac{|x_i - y_i|}{\max\{x_i, y_i\}} \tag{4.3}$$

Secondly, average over all components in V_x and V_y to indicate their numerical dissimilarity $dissim_M$, which is

$$dissim_M = \frac{1}{S}\sum_{i=1}^{S} d_i \tag{4.4}$$

Afterwards, calculate the cosine value of V_x and V_y which indicates the directional similarity sim_C as

$$sim_C = \cos(V_x, V_y) = \frac{(V_x, V_y)}{\|V_x\| \cdot \|V_y\|} = \frac{\sum_{i=1}^{S} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{S} x_i^2 \cdot \sum_{i=1}^{S} y_i^2}} \tag{4.5}$$

Lastly, through a blending coefficient φ, the similar degree between M_x and M_y is

$$sim(V_x, V_y) = (1-\varphi)sim_M + \varphi sim_C = (1-\varphi)(1-dissim_M) + \varphi sim_C \tag{4.6}$$

Table 2. Statistics of models in Fig. 1.

Model	Component types	Relationship sets
Fig. 1(a)	22	307
Fig. 1(b)	137	9,638
Fig. 1(c)	187	22,474
Fig. 1(d)	774	41,137

5. Experiment results and discussion

5.1. Preprocessing

Preprocessing and representation of a PPM should follow the following principles:

- Each component type in both models is uniquely identified.
- Each model's components are uniquely identified.

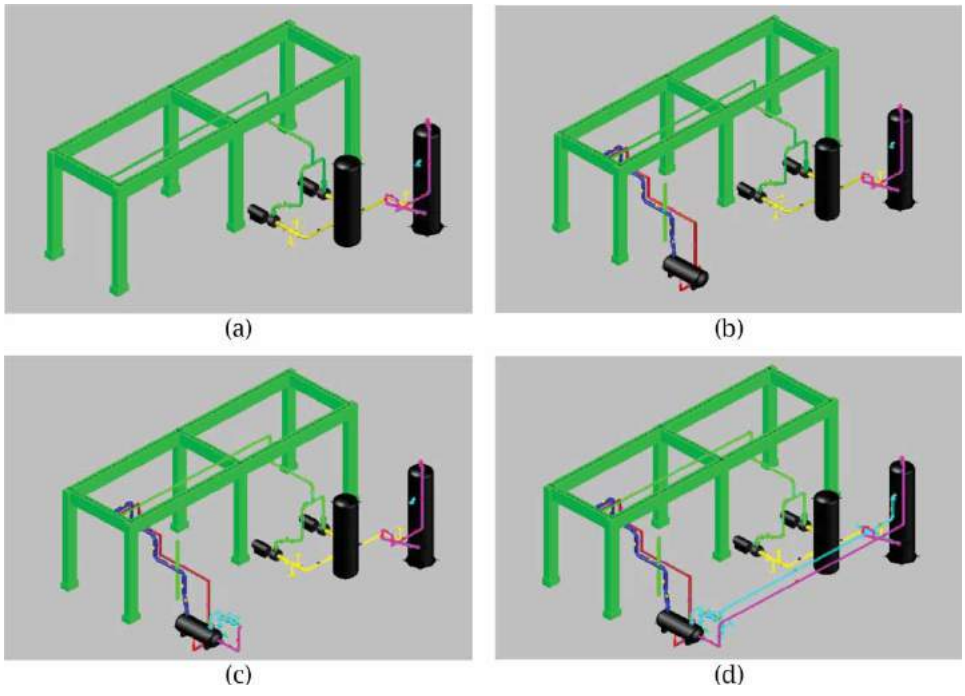

Figure 9. Intermediate PPMs in the process design of Fig. 1(a).

Table 3. Statistics of models in Fig. 9 and their relation trees.

Model	Components	Relationships	Nodes of $Level_1$	Nodes of $Level_2$
Fig. 9(a)	172	332	14	23
Fig. 9(b)	224	430	18	29
Fig. 9(c)	270	522	21	33
Fig. 9(d)	308	602	21	33

Table 4. Similarity measure results between PPMs in Fig. 9 ($\varphi = 0.1$).

	Fig. 9(a)	Fig. 9(b)	Fig. 9(c)	Fig. 9(d)
Fig. 9(d)	0.69103/15 ms cosin: 0.99246	0.83938/16 ms cosin: 0.99654	0.97522/16 ms cosin: 0.99877	1.0/16 ms cosin: 1.0

- Each component should be labeled by one type and one type only.
- Each component has at most one relationship set. A component has none relationship set if it is isolated or does not have any linking relationship.

All of our experiments are conducted in an Intel dual core 2.1 GHz CPU and 3G memory laptop. We use PDSOFT® 3DPiping and Visual Studio 2008 as the development environment to preprocess PPMs. Table 2 gives the quantities of component types and relationship sets.

5.2. Experiment results

Constructing a process plant is a time consuming project. During this collaborative design process, new requirements may be added. Thus, before the completion of a plant design, lots of intermediate models are generated and stored. In order to retrieve these models for management and reuse, we add intermediate models shown in Fig. 9 and Fig. 11 to verify that our method has a positive performance in distinguishing whether the compared models come from an identical plant or different plants.

After preprocessing, according to the RT construction method described in Section 4.2, we build RTs for PPMs and generate their TRD features. In this experiment, we choose flow direction as edge attribute. Furthermore, we standardize features and acquire the comparable feature vectors as Section 4.3 mentioned. Last, for models shown in Fig. 1, Fig. 9 and Fig. 11, we measure their feature vector similarity using the function proposed in Section 4.4. The value of φ is chosen accordingly. Best performance is observed when φ is set to 0.1 in our experiments.

For models in Fig. 9, Fig. 11 and Fig. 1, statistics of their RTs are presented in Tab. 3, Tab. 5 and Tab. 7 respectively. Corresponding similarity measure results are given in Tab. 4, Tab. 6 and Tab. 8.

Table 5. Statistics of models in Fig. 11 and their relation trees.

Model	Components	Relationships	Nodes of $Level_1$	Nodes of $Level_2$
Fig. 11(a)	14,186	28,402	145	184
Fig. 11(b)	18,219	36,513	158	199
Fig. 11(c)	21,710	43,602	176	219
Fig. 11(d)	22,596	45,373	181	224

Table 6. Similarity measure results between PPMs in Fig. 11 ($\varphi = 0.1$).

	Fig. 11(a)	Fig. 11(b)	Fig. 11(c)	Fig. 11(d)
Fig. 11(d)	0.81112/62 ms cosin: 0.99910	0.88599/63 ms cosin: 0.99979	0.97862/63 ms cosin: 0.99995	1.0/63 ms cosin: 1.0

Table 7. Statistics of relation trees of models in Fig. 1.

Model	Nodes of $Level_1$	Nodes of $Level_2$
Fig. 1(a)	21	33
Fig. 1(b)	132	173
Fig. 1(c)	181	224
Fig. 1(d)	770	838

Table 8. Similarity measure results between PPMs in Fig. 1 ($\varphi = 0.1$).

model	Fig. 1(a)	Fig. 1(b)	Fig. 1(c)	Fig. 1(d)
Fig. 1(a)	1.0/16 ms cosin: 1.0	0.07934/31 ms cosin: 0.63490	0.09469/47 ms cosin: 0.79267	0.00218/203 ms cosin: 0.01223
Fig. 1(b)		1.0/47 ms cosin: 1.0	0.086317/78 ms cosin: 0.62019	0.00357/250 ms cosin: 0.01213
Fig. 1(c)			1.0/63 ms cosin: 1.0	0.00963/281 ms cosin: 0.06598
Fig. 1(d)				1.0/453 ms cosin: 1.0

5.3. Discussion

5.3.1. Measure results

Table 4 and Tab. 6 give the similarity measure results of PPMs in Fig. 9 and Fig. 11, from which we have the following observations:

(1) High similarities are obtained using both our proposed function and Cosine. Fig. 10 and Fig. 12 show the relationship distributions of intermediate PPMs in Fig. 9 and Fig. 11. We can see they share a similar TRD, which is in conformity with real engineering applications.
(2) Because Cosine method is a direction-based measurement without taking numerical difference of each individual feature into consideration, it results in much higher scores.
(3) Statistics in Tab. 3 and Tab. 5 indicate that the scales of RTs are significantly smaller than those of their corresponding models. It directly proves that our TRD based method can reduce the computational complexity.

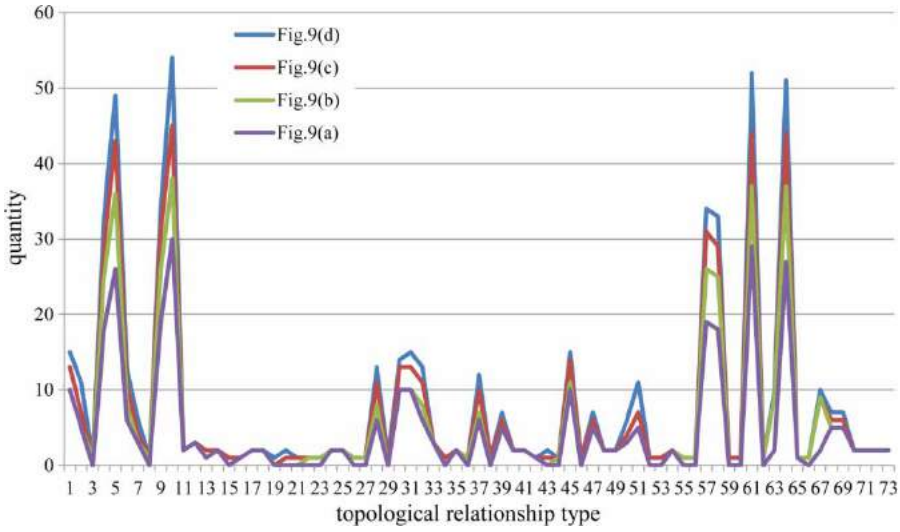

Figure 10. The topological relationship distributions of PPMs in Fig. 9.

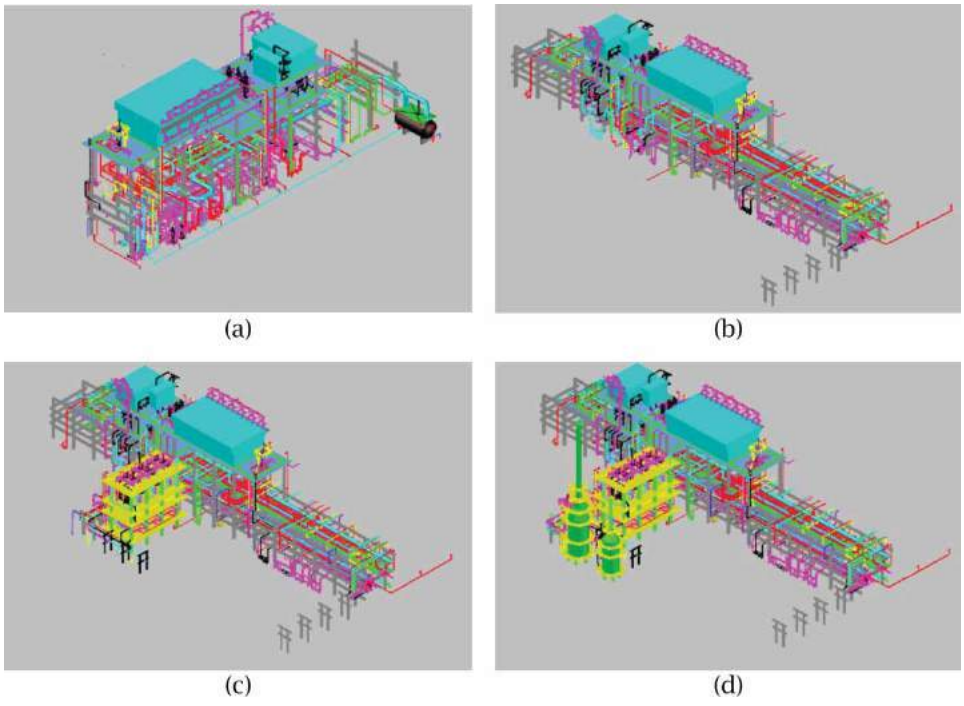

Figure 11. Intermediate PPMs in the process design of Fig. 1(c).

Tab. 8 gives measure results of four PPMs with different engineering applications presented in Fig. 1, from which we can see:

(1) Our measure results are sensitive to plant application backgrounds. Models in Fig. 1 come from different plants. The different design criteria lead to the big differences in both relationship types and quantities. Experiments in Tab. 8 also indicate that different background models result in a low similarity.

(2) Cosine is not ideally suitable for TRD features. As shown in Tab. 8 and Fig. 13, although model Fig. 1(b) has significant differences with Fig. 1(c) in scale and TRD, Cosine still gives a higher similarity. Therefore, the higher similarity by Cosine cannot guarantee the higher similarity between their corresponding PPMs.

We conclude the experimental results shown in Tab. 4, Tab. 6 and Tab. 8 as follows:

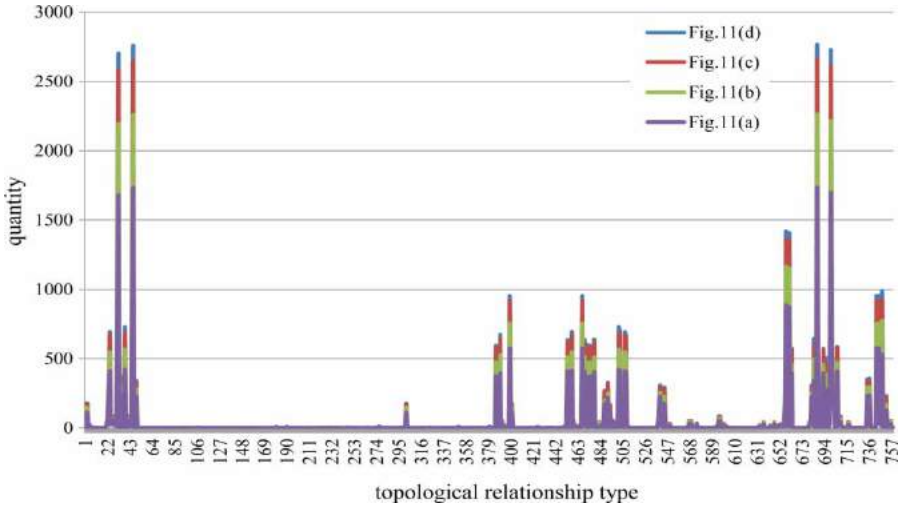

Figure 12. The topological relationship distributions of PPMs in Fig. 11.

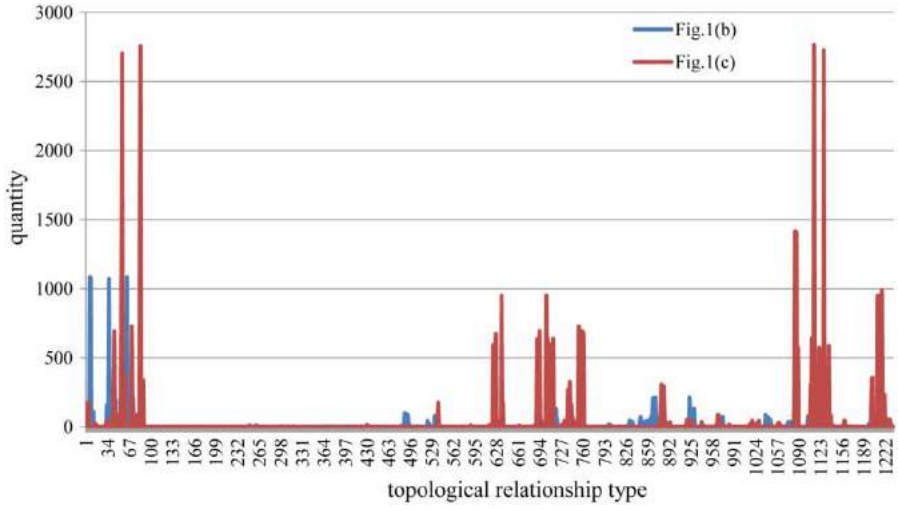

Figure 13. The topological relationship distributions of PPMs Fig. 1(b) and Fig. 1(c).

(1) A general similarity function should meet at least the following two conditions [28]:
- Normality: $0 < sim(X, Y) \leq 1.0$, $sim(X, Y) = 1$ only when $X = Y$;
- Symmetry: $sim(X, Y) = sim(Y, X)$.

Our hybrid similarity function confirms to the above conditions.

(2) Compared with Cosine, our hybrid function is more suitable for TRD based similarity measure. For models from identical applications as shown in Tab. 4 and Tab. 6, Cosine method gives positive results; however, for models from different applications as shown in Tab. 8, Cosine method has a negative discrimination. It proves that an ideal PPM similarity measure needs to incorporate numerical difference of feature vectors.

(3) The TRD method is adaptive to the engineering applications. Different design rules and project properties give rise to the differences of TRDs.

(4) The time cost of our method has a desired performance. The total run times of feature standardization and similarity calculation are presented in Tab. 4, Tab. 6 and Tab. 8. These time costs all lie in a proper range. As an example, the largest model Fig. 1(d) contains 41,569 components and 83,359 topological relationships, but its similarity measure times with other models are less than 453 ms. With the model scales of models shrinking, their measure time is reducing.

5.3.2. φ selection

Without further explanation, it's plain to see from Section 4.4 that the blending coefficient φ is a tradeoff

between sim_C and $dissim_M$ to the overall measuring emphasis of TRD. sim_C emphasizes the proportional distribution similarity of relationships, while totally regardless of magnitude differences. On the other hand, $dissim_M$ emphasizes more on the magnitude differences. In real applications, if we have a clear prior knowledge that proportional distribution matters more that magnitude, φ should be increased appropriately. Otherwise, φ should be reduced. If none prior knowledge is presented, we can still decide the best choice through a series of experiments. Take this paper for example, we use a line scan method letting φ ranging from 0 to 1 with step $= 0.1$. Results are cross compared with each value of φ. The best performance is observed when φ is set to 0.1.

5.3.3. Effectiveness discussion

A specialized characteristic of PPMs is that, two models with identical topology structures and engineering attributes but different component sizes should be regarded as highly similar. Results demonstrate that our TRD based method possesses a nice property that is only involves topological features and engineering attributes rather than geometrical information.

The time complexity of our proposed framework is $O(O_{RT} + O_{std} + O_{cal})$. The construction of a RT only needs to traverse the PPM once, consequently, $O_{RT} = O(|\mathbf{C}| + |\Sigma_{\mathbf{C}}|)$. The complexities of feature standardization O_{std} and similarity calculation O_{cal} are both linear to the total number of topological relationship types of two compared models.

As a statistical method, our measurement may lose accuracy when TRD features are the same but their corresponding models are actually different. However, after our investigation and experiment, we find that when comparing the global similarity of PPMs, the above circumstance is quite infrequent. The reason is that the characteristics of PPMs are large scale, complex topology and various engineering attributes.

5.3.4. Robustness discussion

Due to the exploitation without raw directions and position information, our TRD based measure embraces the advantages of translation, rotation and similarity transformation consistency. It effectively bypasses the problems of numerical error, low efficiency and expensive running cost brought by traditional 3D coordinate standardization.

In engineering applications of process plant, our measure results can be used for model retrieval system. A model retrieval system with a good performance is beneficial to the management and reuse of PPMs, further shortening design cycle and improving design efficiency. Moreover, as shown in Tab. 4 and Tab. 6, our retrieved results may contain intermediate models generated during the collaborative design process. This can avoid unnecessary repeated design and is rewarding for the following design.

6. Conclusions and future work

To our best knowledge, this paper is the first research to measure Process Plant Model (PPM) similarity. Due to the specialized characteristics of PPMs, existing shape feature based 3D model similarity methods cannot be applied to measure PPM similarity. In this paper, to solve this problem, we propose a Topological Relationship Distribution (TRD) based method. Firstly, to extract TRD feature, a Relation Tree (RT) is presented according to components' attributes and topological relationships. After building RT, the TRD feature can be extracted as a vector. When comparing TRD similarity, for their representative vectors, components with the same position may describe different relationships. Therefore, to acquire comparable feature vectors, we adjust the sequence of relationships through feature standardization. Afterwards, we proposed a hybrid similarity function to measure the similarity of feature vectors, and the computational results are used to measure PPM similarity. The experiment manifests a satisfied implementation and performance of our proposed framework. Compared to conventional methods, our solution is translation, rotation and similarity transformation consistent, bypassing the problems of numerical error and low efficiency brought by coordinates standardization.

In our future study, by introducing the measurement algorithm in this paper, we will build a PPM retrieval system for practical engineering model management. And the partial retrieval of PPM is another research direction of our future work.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61300160). The PPMs used in this paper are courtesy of Beijing Zhongke Fulong Computer Technology Co., Ltd.

ORCID

Rui Wen http://orcid.org/0000-0002-0964-8730

References

[1] Ankerst, M.; Kastenmüller, G.; Kriegel, H. P.; Seidl, T.: 3D Shape Histograms for Similarity Search and Classification in Spatial Databases, Proceedings of the 6th International Symposium on Advances in Spatial Databases,

Springer-Verlag, 1999, 207–226. http://dx.doi.org/10.1007/3-540-48482-5_14

[2] Basri, R.; Costa, L.; Geiger, D.; Jacobs, D.: Determining the similarity of deformable shapes, Vision Research, 38(s 15–16), 1998, 2365–2385.

[3] Biasotti, S.; Marini, S.; Mortara, M.; Patanè, G.; Spagnuolo, M.; Falcidieno, B.: 3D Shape Matching through Topological Structures, Lecture Notes in Computer Science, 2886, 2003, 194–203. http://dx.doi.org/10.1007/978-3-540-39966-7_18

[4] Burdorf, A.; Kampczyk, B.; Lederhose, M.; Schmidt-Traub, H.: CAPD-computer-aided plant design, Computers & Chemical Engineering, 28(s 1–2), 2004, 73–81.

[5] Cha, S. H.: Comprehensive survey on distance/similarity measures between probability density functions, City, 1(2), 2007, 1.

[6] Chua, C. S.; Jarvis, R.: Point Signatures: A New Representation for 3D Object Recognition, International Journal of Computer Vision, 25(1), 1997, 63–85. http://dx.doi.org/10.1023/A:1007981719186

[7] Cicirello, V.; Regli, W. C.: Machining feature-based comparisons of mechanical parts, SMI 2001 International Conference on Shape Modeling and Applications, IEEE, 2001, 176–185.

[8] Cook, SA. The complexity of theorem-proving procedures, ACM Symposium on Theory of Computing. ACM, 1971, 151–158. http://dx.doi.org/10.1145/800157.805047

[9] Dai, X.: The research on AEC CAD modeling technology based on enhanced graph and multi-state model, Beijing: Institute of Computing Technology, Chinese Academy of Sciences, 2000.

[10] Dey, T. K.; Giesen, J.; Goswami, S.: Shape Segmentation and Matching with Flow Discretization, Algorithms and Data Structures. Springer Berlin Heidelberg, 2003, 25–36.

[11] Elinson, A.; Nau, D. S.; Regli, W. C.: Feature-based similarity assessment of solid models, Proceedings of the fourth ACM symposium on Solid modeling and applications, ACM, 1997, 297–310. http://dx.doi.org/10.1145/267734.267806

[12] El-Mehalawi, M.; Allen Miller, R.: A database system of mechanical components based on geometric and topological similarity. Part I: representation, Computer-Aided Design, 2003, 35(1), 83–94. http://dx.doi.org/10.1016/S0010-4485(01)00177-4

[13] El-Mehalawi, M.; Allen Miller, R.: A database system of mechanical components based on geometric and topological similarity. Part II: indexing, retrieval, matching, and similarity assessment, Computer-Aided Design, 2003, 35(1): 95–105. http://dx.doi.org/10.1016/S0010-4485(01)00178-6

[14] Guyon, I.; Elisseeff, A.: An Introduction to Feature Extraction, Feature Extraction, Springer Berlin Heidelberg, 2006, 1–25. http://dx.doi.org/10.1007/978-3-540-35488-8_1

[15] Huang, X.: Research on fast rendering techniques in engineering CAD, Ph.D. Thesis, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 2002.

[16] Koutra, D.: Large Graph Mining and Sense-making, Ph.D. Thesis, Microsoft Research, 2014.

[17] Loffler, J.: Content-based retrieval of 3D models in distributed Web databases by visual shape information, Proceedings of IEEE International Conference on Information Visualization, IEEE, 2000, 82–87. http://dx.doi.org/10.1109/iv.2000.859741

[18] Novotni, M.; Klein, R.: A geometric approach to 3D object comparison, Proceedings of international conference on shape modeling, 2001, 154–166. http://dx.doi.org/10.1109/SMA.2001.923387

[19] Osada, R.; Funkhouser, T.; Chazelle, B.; Dobkin, D.: Shape distributions, ACM Transactions on Graphics, 21(4), 2002, 807–832. http://dx.doi.org/10.1145/571647.571648

[20] Raymond, J. W.; Gardiner, E. J.; Willett, P.: Rascal: Calculation of graph similarity using maximum common edge subgraphs, The Computer Journal, 45(6), 2002, 631–644. http://dx.doi.org/10.1093/comjnl/45.6.631

[21] Stephan, M.; Cordy, J. R.: A survey of model comparison approaches and applications, Modelsward, 2013.

[22] Su, Z.; Li, W.; Kong, J.; Dai, Y.; Tang, W.: Watermarking 3D CAPD models for topology verification, Computer-Aided Design, 45(7), 2013, 1042–1052. http://dx.doi.org/10.1016/j.cad.2013.04.001

[23] Su, Z.; Zhou, L.; Li, W.; Dai, Y.; Tang, W.: Topology authentication for CAPD models based on Laplacian coordinates, Computers & Graphics, 37(4), 2013, 269–279. http://dx.doi.org/10.1016/j.cag.2013.02.009

[24] Sundar, H.; Silver, D.; Gagvani, N.; Dickinson, S.: Skeleton Based Shape Matching and Retrieval, Proceedings of International Conference on Shape Modeling and Applications, IEEE Computer Society, 2003, 130. http://dx.doi.org/10.1109/smi.2003.1199609

[25] Tangelder, J. W. H.; Veltkamp, R. C.: A survey of content based 3D shape retrieval methods, Multimedia tools and applications, 39(3), 2008, 441–471. http://dx.doi.org/10.1007/s11042-007-0181-0

[26] Wang, F.; Zhang, S.; Bai, X.; Chen, S.: Local Matching of 3D CAD Models Based on Subgraph Isomorphism, Journal of Computer-Aided Design & Computer Graphics, 2008, 20(8):1078–1084.

[27] Yan, X.; Yu, P. S.; Han, J.: Graph indexing: a frequent structure-based approach, International Conference on Management of Data, 2004, 335–346. http://dx.doi.org/10.1145/1007568.1007607

[28] Yu, Z.; Dong, L. Y.; Zhao, J. I.: Vector similarity measurement method, Technical Acoustics, 28(4), 2009, 532–536.

[29] Zeng, Z.; Tung, A. K. H.; Wang, J.Feng, J.; Zhou, L.: Comparing stars: On approximating graph edit distance, Proceedings of the Very Large Data Base Endowment, 2(1), 2009, 25–36.

[30] Zhang, C.; Chen, T.: Efficient feature extraction for 2D/3D objects in mesh representation, Processings of International Conference on Image Processing, IEEE, 2001, 935–938.

[31] Zhang, K. X.; Zhang, S. S.; Liang, L. I.: Partial matching algorithm of 3D CAD model, Computer Integrated Manufacturing Systems, 17(9), 2011, 1880–1886.

[32] Zheng, W.; Zou, L.; Lian, X.; Wang, D.; Zhao, D.: Graph similarity search with edit distance constraint in large graph databases, Proceedings of the 22nd ACM international conference on information & knowledge management, ACM, 2013, 1595–1600.

[33] Zhao, P.; Yu, J. X.; Yu, P. S.: Graph indexing: Tree + delta >= graph, Proceedings of international conference on very large data bases, 2007, 938–949.