5Y-08

間取りや広さをクエリとする直感的な不動産検索システム

大原康平　山崎俊彦　相澤清晴
東京大学

1 はじめに

従来の不動産検索システムはキーワードによる検索やAND/ORによる絞り込み検索であるが，この方法では各部屋の間取りや広さといった構造について検索することはできない．本研究では，不動産物件の間取りをグラフ構造で表現し，ユーザの入力するクエリグラフと類似度が高い物件を出力するシステム（Fig. 1）を提案する．さらに，各部屋の広さや，ユーザが特に重要だと感じる構造の重みを考慮するグラフ間類似度を設定することで，よりユーザの希望に沿う物件の出力を可能とする．

2 関連研究

2つのグラフの類似度や距離の計算において，2つのグラフのMaximum Common Subgraph（MCS，最大共通部分グラフ）を考える必要がある．MCSをFig. 2に示す．Fig. 2の矢印の上側の二つのグラフが与えられた場合，矢印の下側のグラフがMCSである．

Conteら[1]はMCSを求めるアルゴリズムを比較しており，ノード数やエッジ数，グラフの密度等によって最適なアルゴリズムが異なることが結果として得られている．また，

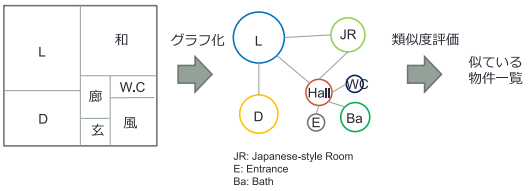

Fig. 1　目標とするシステム

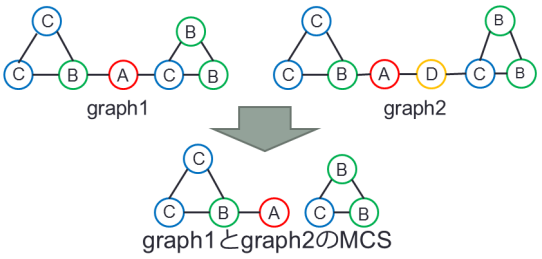

Fig. 2　二つのグラフのMCS（最大共通部分グラフ）

二つのグラフ間の距離の設定は様々であり，目的に応じて設定する必要がある[2] [3].

3 提案手法

本研究での検索システムでは，ユーザはクエリとして，あるグラフ $G = (V, E, L, W)$ を作成し検索する．ここで，V はノードの集合，E はエッジの集合，L はノードのラベルを返す写像，W は重み（部屋の広さに対応する値）を返す写像である．ユーザが作成したクエリグラフとデータベースの各グラフとの類似度が計算され，類似度の高いものから順に出力される．

グラフ同士のMCSの探索はMcGregorのアルゴリズム[1]を利用して行う．概念図をFig. 3に示す．グラフのノードはそれぞれidとlabelをもつとする．二つのグラフのうちノードが少ない方（Fig. 3でのgraph1）のノードidが小さい順に，同じlabelをもつノードをもう一方のグラフ（Fig. 3でのgraph2）から探し，存在していればMCSの現在の候補（graph1での部分グラフ，graph2での部分グラフ）にそのノードを加える．さらに，加えたノードと，MCSの候補であるgraph1，graph2の部分グラフとの間に共通して存在するエッジが存在すればそのエッジもMCSの候補に加える方法である．

クエリグラフ q とデータベースのグラフ g とのグラフ類似度 $sim(q, g)$ は，q と g のMCSを mcs として，以下の式を用いる．

$$sim(q,g) = \sum_{v \in V(mcs)} \gamma(v) + \sum_{e \in E(mcs)} \gamma(e) - \lambda \sum_{v \in V(mcs)} |W(v_q) - W(v_g)|$$

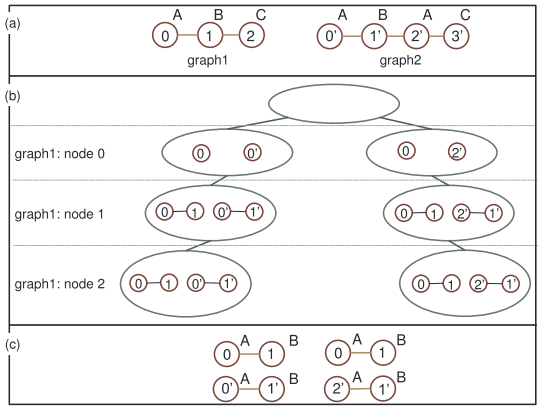

Fig. 3　McGregorのアルゴリズム：(a) 二つのグラフ (b) 探索木 (c) graph1とgraph2のMCSの二つの組

An Intuitive System for Searching Apartments using Floor Plans and Areas of Rooms
Kohei OHARA, Toshihiko YAMASAKI, Kiyoharu AIZAWA
The University of Tokyo

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.

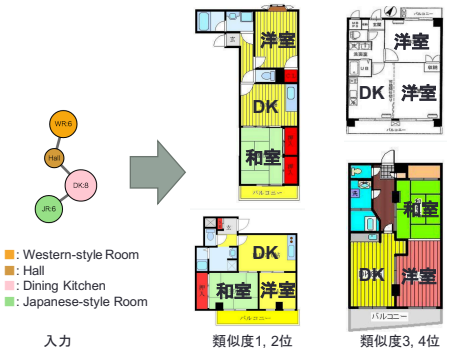

Fig. 4　クエリグラフの入力と検索結果の例 1

Fig. 5　クエリグラフの入力と検索結果の例 2

- $\gamma(v)$，$\gamma(e)$ はユーザの重要指定によって変わる正の値．本論文においては，v または e にユーザが重要指定をしていれば 100，そうでなければ 1 とニ値にしたが，x 種の値を用意すれば重要さの度合いを x 段階で指定することも可能である．
- $\lambda > 0$．本論文の実験では 0.1 とした．
- v_q，v_g はそれぞれ v に対応する q，g のノード

この式の第一項と第二項は，g において，q と同じ部屋や部屋同士の接続を持つ程類似度を大きくするとともに，ユーザが q にて重要だと指定している構造が g に存在すればさらに類似度を大きくする項である．第三項は，q と g の対応する部屋の広さの差が大きければ大きい程類似度を小さくする項である．

4　実験

SUMMO[4] の東京の賃貸物件データを対象とし，これをスクレイピングして取得した．いくつかの区・市の物件のデータをデータベースに格納した．この中から，家賃・管理費等・敷金・礼金・エリアが全く同じ物件は，同じ物件だが重複して登録されていた物件であると判断して除き，約 1 万件とした．この約 1 万件のうち約 1000 件の物件に対してグラフ構造を手動で入力した．グラフ構造が入力されていないグラフは，グラフ類似度の計算を行わず，システムの検索結果として表示されない．システムは Ruby on Rails 上で動作するよう実装し，不動産検索サイト同様ブラウザで検索できるようにした．本稿でのシステムでの検索結果の例として Figs. 4 ～ 6 を示す．

Fig. 4 は，従来の不動産検索システム同様間取りを 2DK 絞り込むと共に，間取り構造に関しても図の左側のような構造をクエリとした検索例である．類似度 1 位の物件はクエリの期待通りの結果が得られた．類似度が 2 位から 4 位の物件は，洋室が廊下に接続していない物件や和室が無い物件であった．これは，そのそれぞれの条件と「DK と廊下が接続する」ことが同スコアであるため，DK と廊下が接続していることで類似度が高く評価された結果である．これに対して，Fig. 5 では Fig. 4 の入力において，DK と和室が接続する事は重要であるとユーザが指定した場合の結果である．Fig. 4 と異なり Fig. 5 では類似度 3，4 位にお

Fig. 6　クエリグラフの入力と検索結果の例 3

いても DK と和室が接続する物件が得られた．Fig. 6 は，3LDK の物件の中で図の左のように，LD が 12J（帖・畳）の物件を検索した結果である．類似度 1 位から 4 位において，LD が 12J に近い物件が順に得られた．

5　結論

本研究ではグラフで表現できるような構造をもった不動産物件の間取りのレイアウトに着目し，直感的な検索システムの開発を目的とした．McGregor のアルゴリズムを拡張し，間取りに関するユーザの希望を反映した検索結果を返す不動産検索システムを開発した．今後は入力した間取り構造の活用を考えている．

参考文献

[1] D. Conte, et al. Challenging complexity of maximum common subgraph detection algorithms: A performance analysis of three algorithms on a wide database of graphs. *JGAA*, 2007.

[2] H. Shang, et al. Similarity search on supergraph containmen. In *ICDE*, 2010.

[3] Y. Zhu, et al. Finding top-k similar graphs in graph databases. In *EDBT*, 2012.

[4] `https://suumo.jp/`.

(謝辞) 本研究の一部は科学研究費助成事業 (26700008) の支援を受けて行われた．

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.