

類似関係データに基づく関連アイテムの可視化手法の提案

黒田 脩平[†] 奥 健太[†]

[†] 龍谷大学理工学部 〒520-2194 滋賀県大津市瀬田大江町横谷 1-5
E-mail: [†]t150482@mail.ryukoku.ac.jp, ^{††}okukenta@rins.ryukoku.ac.jp

あらまし クラウドソーシングから得られたアイテム間の類似関係データに基づき、アイテムを2次元空間に可視化する手法を提案する。本研究では、アイテムのドメインとして映画を取りあげる。クラウドソーシングにより映画の世界設定の観点から類似関係データを収集する。得られた類似関係データに基づき2次元のアイテム空間を構築する。構築されたアイテム空間において、ユーザが興味のあるアイテムに関連するアイテムを探索することを支援する。実験結果より、人の主観により挙げられた類似関係データを基にすることで、互いに関連のあるアイテム同士が近くに配置されるように、類似関係マップを適切に生成できることを確認した。

キーワード 類似関係データ, クラウドソーシング, 推薦システム, 可視化

1. はじめに

映画や音楽、小説、漫画など、世の中には膨大なコンテンツで溢れかえっている。例えば、IMDB^(注1)には400万件もの映画情報が登録されている。推薦システムは、このような膨大なコンテンツ群の中から、ユーザの嗜好に合ったコンテンツを提示する。推薦システムの基本方式として、大きく次の三つの方式がある；(a) 協調ベース推薦[1]、(b) 内容ベース推薦[2]、(c) 知識ベース推薦[3][4]。(a)および(b)の方式には、大量の履歴データが必要となる。また、(a)では新規ユーザおよび新規アイテムのコールドスタート問題がある。(b)では、新規アイテムのコールドスタート問題については解消されるものの、新規ユーザのコールドスタート問題は依然として残る。これらに対し、(c)は履歴データは必要とせず、新規ユーザに対しても推薦可能である。

本研究では、履歴データをもたない新規ユーザにも推薦可能なシステムを目指し、知識ベースのアイテム推薦システムを提案する。知識ベース推薦には、制約ベース推薦方式と事例ベース推薦方式がある。制約ベース推薦は、ユーザが提示した制約を満たすアイテムを推薦する。事例ベース推薦は、ユーザが例示したアイテムや属性に類似するアイテムを推薦する。推薦されたアイテムを参考にしながら、ユーザが条件を修正（批評）していくことで、ユーザのアイテム探索を支援する。事例ベース推薦方式により推薦システムを実装する際には、アイテム間の類似性尺度の定義と批評インタフェースの設計が必要となる。

本研究では、人の主観的な印象によるアイテム間の類似性尺度に着目する。具体的にはクラウドソーシングにより、ワーカーに主観的に類似すると思うアイテム同士を選んでもらう。ここで、ワーカーに類似性を判断してもらう際には、一定の基準となる観点を与える必要がある。本研究では、映画ドメインでの適用を想定し、世界設定が類似しているかという観点でアイテムを選択してもらう。ただし、世界設定とは、ジャンルとは異なり、

時代や舞台、登場人物の種族、世界の法則などに関する設定を指すが、類似性の最終的な判断はワーカーの主観に委ねている。こうして得られたアイテム間の類似関係データに基づき、アイテム間の類似度を算出する。また、アイテム間の類似性の関係を2次元特徴空間上にマッピングすることで可視化したインタフェースを設計する。ユーザはこのインタフェース上で興味アイテムとの類似性を閲覧しながら批評することができる。

2. 関連研究

2.1 映画推薦システム

内容に基づく映画推薦システムにおいては、映画の特徴表現が重要となる[5]。高次の映画特徴としては、ジャンルや監督、俳優などが用いられる。これらは、コンテンツ作成者によって登録されたり、Webを介してタグによって与えられたりする。低次の映画特徴としては、映画コンテンツの映像の色や動作、明るさなどが挙げられる。これらは、ミゼンセーヌ(mise-en-scène)[6]特徴とよばれ、動画ファイルから抽出される。

本研究では、人の主観的な印象による映画間の類似性尺度に着目する。具体的にはクラウドソーシングにより、ワーカーに主観的に類似すると思う映画同士を選んでもらう。ここで、映画の類似性の判断基準として世界設定に着目している。ただし、世界設定とは、ジャンルとは異なり、時代や舞台、登場人物の種族、世界の法則などに関する設定を指すが、類似性の最終的な判断はワーカーの主観に委ねている。

3. 定義

定義1：アイテム。 映画や小説、ゲームなどの作品をアイテムとよび $a_i \in A$ と表す。本稿では、アイテム種別として映画を扱う。

定義2：アイテム間共起行列。 アイテム間共起行列を $R = [r_{ij}]_{|A| \times |A|}$ と表す。共起行列 R は $|A| \times |A|$ の対称行列である。 r_{ij} はアイテム i とアイテム j が共起した回数を表す。

定義3：アイテム間距離行列。 アイテム間距離行列を $D =$

(注1) : <https://www.imdb.com/>

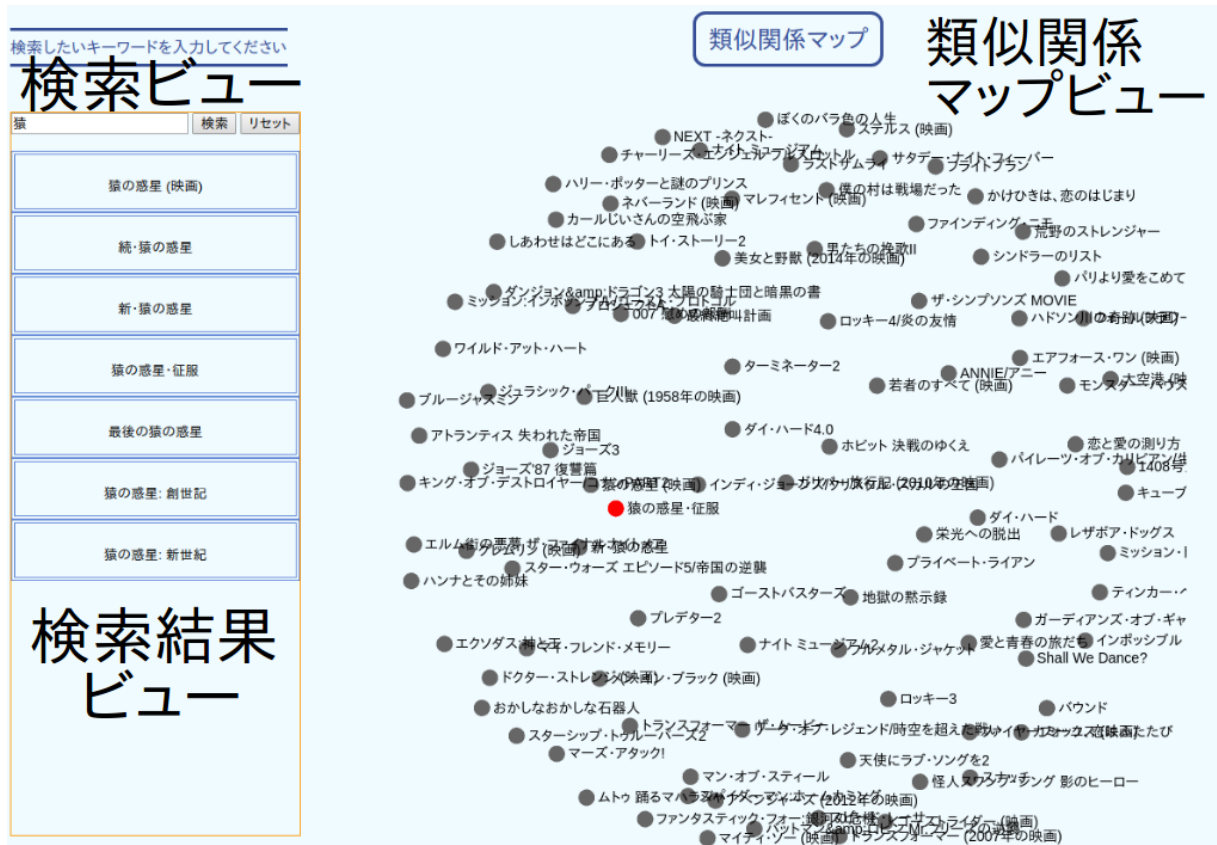


図 1 システムインタフェース。検索ビューと検索結果ビュー，マップビューから構成される。類似関係マップビューとして，選択したアイテムの近辺に世界設定が類似した作品が提示される。一方で，選択したアイテムから離れるほど類似していない作品が提示される。類似関係マップビュー上で気になるアイテムをクリックすると，選択アイテムの Wikipedia ページを閲覧することができる。図では，例として、『猿の惑星・征服』を選択した結果を示している。

$[d_{ij}]$ と表す。距離行列 D は $|A| \times |A|$ の対称行列である。 d_{ij} はアイテム i とアイテム j の間の距離を表す。距離が近いほど，アイテム同士が類似することを表す。

定義 4: アイテム間類似関係マップ。 アイテム間類似関係マップは，アイテム間の類似関係を 2 次元特徴空間に可視化したものである。類似しているアイテム同士は近く，類似していないアイテム同士は離れて配置される。

4. システム

本章では，提案システムである類似関係データに基づく関連アイテムの可視化システムについて説明する。

4.1 システム概要

本節では，まず，システムのインタフェースについて説明し，システム構成について説明する。

4.1.1 インタフェース

図 1 に本システムのインタフェースを示す。インタフェースは，検索ビュー，検索結果ビュー，類似関係マップビューから構成される。

ユーザは検索ビューから検索キーワードを入力する。「検索」ボタンをクリックすることで，検索キーワードと部分一致するアイテムが検索結果ビューに提示される。検索結果ビューの一

覧からアイテムを一つ選択することで，選択したアイテムの類似関係マップビューが表示される。選択したアイテムは赤丸で強調表示される。この類似関係マップビューでは選択したアイテムの近辺に世界設定が類似した作品が提示される。図 1 では，例として『猿の惑星・征服』を選択した結果を示している。

ユーザは類似関係マップビューにおいて，操作（移動，ズームイン，ズームアウト）を行うことができ，周辺のアイテムを探索することができる。さらに，気になるアイテムをクリックすることでダイアログを表示させ，リンクをクリックすることで選択したアイテムの Wikipedia ページを別タブで開くことができ，詳細を閲覧することができる。

4.1.2 システム構成

図 2 にシステム構成図を示す。以下，各処理について説明する。項目番号は図中の番号に対応する。

- (1) クラウドソーシングによりアイテム間の類似関係データを収集する（4.2 節）。
- (2) 収集した類似関係データに基づき，アイテム間距離行列を作成する（4.3 節）。
- (3) アイテム間距離行列を基に，多次元尺度構成法を用いてアイテム集合を 2 次元特徴空間に写像し，アイテム間類似関係マップを作成する（4.4 節）。

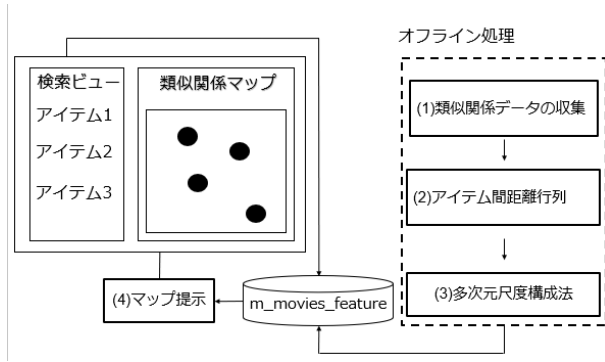


図2 システム構成. (1) クラウドソーシングを用いて、類似関係データを収集する. (2) 類似関係データに基づき、アイテム間距離行列を作成する. (3) 多次元尺度構成法によりアイテム間距離行列を2次元特徴空間に写像し、アイテム間類似関係マップを作成する. (4) 選択されたアイテム周辺のアイテム間類似関係マップをマップビューに提示する.

$$R = \begin{matrix} & \begin{matrix} 873 & 136 & 822 & 1518 \end{matrix} \\ \begin{matrix} 873 \\ 136 \\ 822 \\ 1518 \end{matrix} & \begin{bmatrix} 1 & 4 & 1 & 1 \\ 4 & 1 & 3 & 4 \\ 1 & 3 & 1 & 0 \\ 1 & 4 & 0 & 1 \end{bmatrix} \end{matrix}$$

図3 アイテム間共起行列.

(4) 選択されたアイテム周辺のアイテム間類似関係マップをマップビューに表示する.

4.2 アイテム間類似関係データの収集

クラウドソーシングによりアイテム間類似関係データを収集する. 類似関係データの収集手順は下記のとおりである:

(1) クラウドソーシングのワーカが、アイテムリスト A の中から、任意のアイテム a_i を一つ選択する. ここで、選択するアイテムはワーカが利用したことがある (例; 映画の場合は観たことがある), あるいは知っているアイテムとする.

(2) ワーカが、アイテムリスト A の中から、その選択したアイテム a_i と類似していると思うアイテムをすべて選択する. ここで、アイテム同士が類似しているか否かの判断は、ワーカの主観によるものとする.

(3) 他の任意のアイテムについても、(1)–(2) を繰り返す. ここで、本稿では、アイテム種別として映画を扱う. このとき、類似アイテムを判断する際には、映画の世界設定の類似性に注目している. ワーカには「世界設定が類似すると思う作品」を選択するように指示した.

4.3 アイテム間距離行列の作成

4.3.1 アイテム間共起行列の作成

得られたアイテム間類似関係データを基に、アイテム間共起行列 $R = [r_{ij}]$ を作成する. ここでは、アイテム i とアイテム j が類似していると回答したワーカ数が r_{ij} となる. ただし、対角要素はすべて $r_{ii} = 1$ とする.

共起行列 R の例を図3に示す. この例では、例えば、ID873

$$D = \begin{matrix} & \begin{matrix} 873 & 136 & 822 & 1518 \end{matrix} \\ \begin{matrix} 873 \\ 136 \\ 822 \\ 1518 \end{matrix} & \begin{bmatrix} 0 & 0.2 & 0.5 & 0.5 \\ 0.2 & 0 & 0.25 & 0.2 \\ 0.5 & 0.25 & 0 & 1 \\ 0.5 & 0.2 & 1 & 0 \end{bmatrix} \end{matrix}$$

図4 アイテム間距離行列.

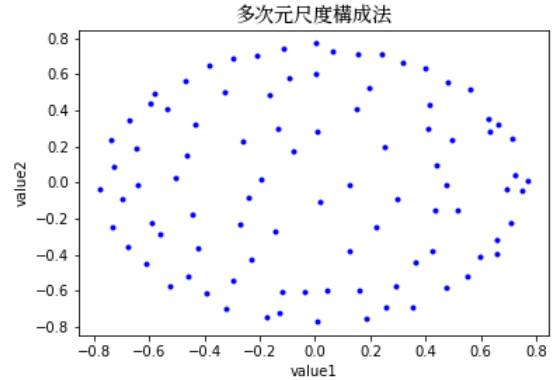


図5 アイテム間類似関係マップの例. value1 は削減された次元の第1軸, value2 は削減された次元の第2軸を表す. 各点はアイテムを表す.

のアイテムと ID136 のアイテムが4名のワーカにより類似していると判断されたことを表しており, $r_{873,136} = 4$ となる.

ここで、他と共起するアイテムが存在しないアイテムについては削除しておく.

4.3.2 アイテム間距離行列の作成

アイテム間共起行列 R をアイテム間距離行列 $D = [d_{ij}]_{|A| \times |A|}$ に変換する. アイテム i とアイテム j の間の距離 d_{ij} は共起回数 r_{ij} の逆数とする. ただし、0 による除算を避けるため分母に1を加える. すなわち、距離 d_{ij} は次式により算出する:

$$d_{ij} = \frac{1}{r_{ij} + 1} \quad (1)$$

ただし、対角要素はすべて $d_{ii} = 0$ とする.

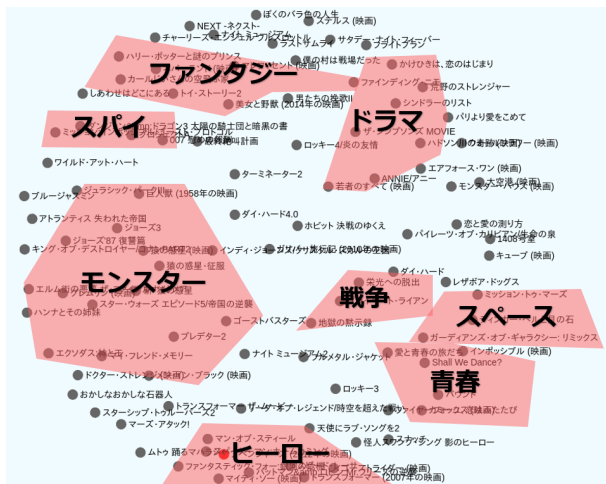
距離行列 D の例を図4に示す. この例では、例えば、ID873のアイテムと ID136 のアイテムの間の距離が0.2であることを表しており, $d_{873,136} = 0.2$ となる.

4.4 アイテム間類似関係マップの作成

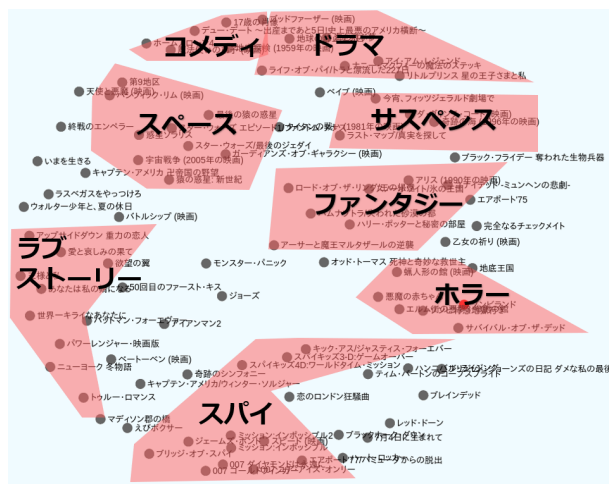
多次元尺度構成法を用いて、アイテム間距離行列 D を基に、アイテム集合 A を2次元特徴空間に写像する. 多次元尺度構成法により、2次元空間上において、類似したアイテム同士を近く、類似していないアイテム同士が遠くなるように、アイテム集合が配置され、アイテム間の関係を視覚化できる. こうして得られた2次元空間をアイテム間類似関係マップとする.

多次元尺度構成法による、アイテム集合 A の2次元特徴空間への写像は以下の手順で行う:

(1) アイテム間距離行列 D にヤング・ハウスホルダー変換を行う.



(a) 1 組目の類似関係マップ：ファンタジー、スパイ、モンスター、ヒーロー、戦争、ドラマ、スペース、青春の 8 つのグループに分類される。



(b) 2 組目の類似関係マップ：コメディ、ドラマ、スペース、ラブストーリー、ファンタジー、サスペンス、ホラー、スパイの 8 つのグループに分類される。

図 6 類似関係マップ。

(2) ヤング・ハウスホルダー変換で得られた行列をスペクトル分解し、固有値と固有ベクトルを求める。

(3) 上位 2 個の固有値に対応する固有ベクトル e_1, e_2 を得る。

(4) 2 個の固有ベクトルを基に基底行列 $P = [e_1, e_2]$ を得る。

以上の手順により得られた基底行列 P の i 番目のベクトルがアイテム a_i の 2 次元削減後の特徴ベクトルとなる。

図 5 は、アイテム間類似関係マップの例を表す。図中の value1 は削減された次元の第 1 軸、value2 は削減された次元の第 2 軸を表し、各点はアイテムを表す。

5. 結果

提案システムの有用性を検証するために評価を行った。5.1 節では、本実験で用いるデータセットについて説明する。5.2 節では、このデータセットを用いて、システムを実行したときの結果について定性的に評価する。

5.1 データセット

4.2 節で述べたとおり、クラウドソーシングによりアイテム間の類似関係データを収集した。

本実験では、映画のドメインを対象とし、Wikipedia の「映画作品一覧」からリンクが張られている 1,813 件の映画作品を扱った。この 1,813 件を 200 件ずつの 9 組に分割した。ただし、最後の 9 組目のみ 213 件とした。組ごとにクラウドソーシングによりアイテム間の類似関係データを収集した。ここで、前処理として、自分自身のアイテムとしか類似しておらず、他と類似するアイテムが存在しないアイテムはデータから削除した。得られた有効データ数は、全組合計で映画 819 件となった。組ごとに、アイテム間類似関係マップを作成した。例えば、「猿の惑星」という映画は 1 組目に含まれるため、対応する類似関

係マップは 1 組目の類似関係マップとなる。

5.2 定性評価

本項では、出力された類似関係マップに提示されている各アイテムの周辺に類似性のあるアイテムが提示されているか確認をする。まず、全体を俯瞰した後、例として、a) 図 6(a) の「ドラマ」、b) 図 6(a) の「ヒーロー」、c) 図 6(b) の「ホラー」付近の類似関係マップについて定性的に分析する。

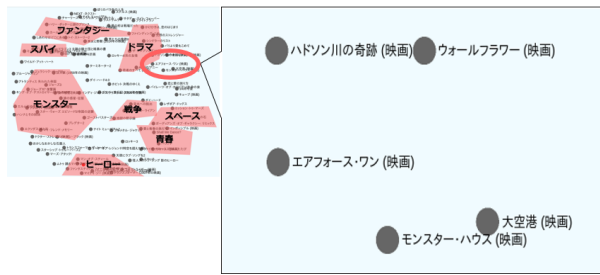
5.2.1 全体の俯瞰

図 6(a) は 1 組目、図 6(b) は 2 組目の類似関係マップを示している。全体の傾向を俯瞰するために、近傍のアイテム同士の関連性を踏まえ、大まかなグループを図中に示した。各グループのラベルは、TSUTAYA^(注2)の映画・映像 (DVD・ブルーレイ) のジャンルを参考に付与している。

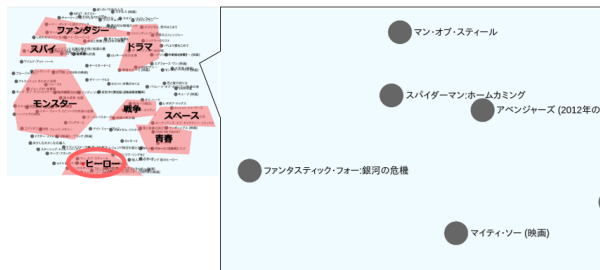
各グループに含まれるアイテムの内容について、著者が調べたところ、互いに関連するアイテム同士がそのグループ内で近くに分布していることを確認した。一方で、グループ間の位置関係は必ずしも予想どおりの結果が得られたとはいえない。例えば、図 6(a) では、「ドラマ」と「青春」の間に「戦争」や「スペース」が位置した。また、図 6(b) では、「ラブストーリー」が「コメディ」や「ドラマ」と離れた位置に提示された。

この問題が起きた理由として、本研究で利用した類似関係データには各アイテム同士の類似データしかなかったためである。このことから、各アイテムと極端に類似していないアイテムのデータを収集することで、距離が離れるほど類似していないアイテムが提示されるのではないかと考えられる。しかし、今回はクラウドソーシングにより世界設定の類似性の観点から得た類似関係データに基づいて類似関係マップを作成している。そのため、グループ間の位置関係が離れて提示されたとしても、

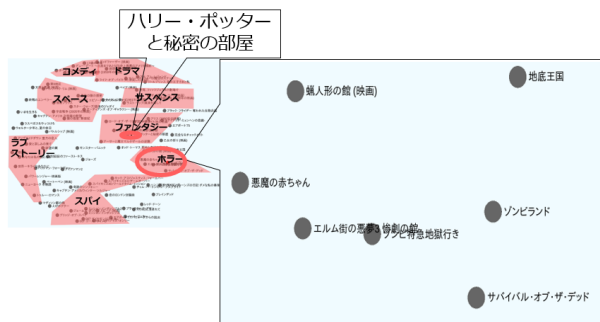
(注2) : <https://movie-tsutaya.tsite.jp/netdvd/dvd/hotrankingGenreWeekly.do?genreID=00001&subGenreID=00086&PT=0>



(a) 図 6(a)「ドラマ」の拡大図。



(b) 図 6(a)「ヒーロー」の拡大図。



(c) 図 6(b)「ホラー」の拡大図。

図 7 図 6(a) および図 6(b) の部分拡大図。

必ずしも類似性がないとは限らない。

5.2.2 細 部

ここでは、例として、a) 図 6(a) の「ドラマ」b) 図 6(a) の「ヒーロー」、c) 図 6(b) の「ホラー」付近の類似関係マップについて定性的に分析する。

a) 図 6(a) の「ドラマ」

図 7(a) は図 6(a) の「ドラマ」周辺を拡大したものである。

『ハドソン川の奇跡 (映画)』のグループは「ドラマ」であるが、周辺に提示されている『エアフォース・ワン (映画)』や『大空港 (映画)』にはグループの類似性がみられなかった。しかし、これらの映画はすべて飛行機が舞台の映画である。これは、グループではなく世界設定の観点で類似作品を提示できたといえる。また、『ハドソン川の奇跡 (映画)』と『ウォールフラワー (映画)』、『大空港 (映画)』と『モンスター・ハウス (映画)』

にはグループに類似性があることを確認した。

b) 図 6(a) の「ヒーロー」

図 7(b) は図 6(a) の「ヒーロー」周辺を拡大したものである。

『アベンジャーズ (2012 年の映画)』はグループが「ヒーロー」であり、ニューヨークが舞台である。同様に、『スパイダーマン: ホームカミング』もグループが「ヒーロー」であり、ニューヨークが舞台である。このことから、グループ、世界設定がともに類似しているアイテムが提示されていることがわかる。また、周辺には『マン・オブ・スティール』など、世界の法則としてヒーローが登場する作品が提示されている。この結果から、『アベンジャーズ (2012 年の映画)』の周辺に類似作品を提示できたといえる。

c) 図 6(b) の「ホラー」

図 7(c) は図 6(b) の「ホラー」周辺を拡大したものである。

『ゾンビランド』は世界の法則として、ゾンビが登場する映画である。周辺には『サバイバル・オブ・ザ・デッド』、『ゾンビ特急地獄行き』といったゾンビが登場する作品が提示されている。また、『蟻人形の館 (映画)』や他の作品はゾンビが登場する作品ではないが、グループが『ホラー』であることから、『ゾンビランド』の類似作品として提示されている。

一方で、『地底王国』はクラウドソーシングの類似関係データから、『ハリー・ポッターと秘密の部屋』と類似しているはずであるが、「ホラー」周辺に提示された。これは、『ハリー・ポッターと秘密の部屋』と類似した他のアイテムの距離が『地底王国』の距離よりも短いため、少し離れた位置に『地底王国』が提示されてしまったと思われる。この問題を解決するには、他の類似しないアイテムとの距離を長くすることで類似していないアイテム周辺への提示を回避できると考えられる。

5.2.3 考 察

以上の結果から、人の主観により挙げられた類似関係データを基にすることで、互いに関連のあるアイテム同士が近くに配置されるように、類似関係マップを適切に生成できることを確認した。さらに、世界設定が類似しているアイテムも提示することがわかった。この類似関係マップを用いることにより、ユーザが興味のあるアイテムに関連するアイテムの探索を支援することが可能であるといえる。一方で、今回は 1,813 件の映画を 9 組に分割し、組ごとの独立した類似関係マップを作成するに留まった。今後は、組間の関係性も考慮し、映画全体を一つの空間で表現する方法について検討する。

6. おわりに

本稿では、クラウドソーシングから得られたアイテム間の類似関係データに基づき、アイテムを 2 次元空間に可視化する手法を提案した。今回の実験結果より、このシステムは得られた類似関係データに基づき構築されたアイテム空間において、ユーザが興味のあるアイテムに関連するアイテムを提示するシステムとしては機能していた。しかし、アイテムとの距離が離れるほど類似していないアイテムを提示するといったことができず、アイテムによっては近い距離に類似しない作品が提示され、システムとしては不完全な推薦が確認できた。今後の課題として、

クラウドソーシングによる類似関係データ数の増加や、類似アイテムだけでなく選択したアイテムと極端に類似していないアイテムのデータを適応させ、類似しているアイテム同士の距離を短く、類似していないアイテムの距離を長くすることで各アイテムの類似関係がより確認できるようにすることである。

謝 辞

本研究は JSPS 科研費 JP16H05932 の助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 5–53, 2004.
- [2] Iván Cantador, Alejandro Bellogín, and David Vallet. Content-based recommendation in social tagging systems.

In *Proceedings of the 4th ACM conference on Recommender systems*, p. 237, 2010.

- [3] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. *Recommender Systems Handbook*. Springer, 2011.
- [4] Marius Kaminskas, Ignacio Fernández-Tobías, Francesco Ricci, and Iván Cantador. Knowledge-based music retrieval for places of interest. In *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies - MIRUM '12*, pp. 19–24, New York, New York, USA, 2012. ACM Press.
- [5] Mehdi Elahi, Yashar Deldjoo, Farshad Bakhshandegan Moghaddam, Leonardo Cella, Stefano Cereda, and Paolo Cremonesi. Exploring the Semantic Gap for Movie Recommendations. In *Proceedings of the 11th ACM conference on Recommender systems*, pp. 326–330, 2017.
- [6] Yashar Deldjoo, Mehdi Elahi, Paolo Cremonesi, Franca Garzotto, Pietro Piazzolla, and Massimo Quadrona. Content-Based Video Recommendation System Based on Stylistic Visual Features. *Journal on Data Semantics*, Vol. 5, No. 2, pp. 99–113, 2016.