Paper:

Product Selection Support System based on Ordered Structure by Formal Concept Analysis

Kazuki Aikawa, Hajime Nobuhara

Intelligent Interaction Technologies, Graduate School of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan
E-mail: aikawa@cmu.iit.tsukuba.ac.jp, nobuhara@cmu.iit.tsukuba.ac.jp

Abstract. With the spread of electronic commerce, it is necessary to find, compare, and select products from a wide variety of products. In a real shop, we can ask staff, but in an online shop, you have to investigate and compare by yourself. In a large-scale net shop, it is possible to accumulate behavior data of many users and make a proposal, but it is difficult to use a similar algorithm in a small-scale net shop because the amount of data is small. In this paper, we propose a recommendation algorithm for small datasets by focusing on the ordered structure obtained by formal concept analysis. Experiments on two datasets confirmed that some recommendation is possible and that performance is improved by using the sequential structure obtained by formal concept analysis.

Keywords: Recommend system, Formal concept analysis

1. Introduction

With the spread of electronic commerce, it is necessary to find, compare, and select products from a wide variety of products. In a real shop, we can ask staff, but in an online shop, you have to investigate and compare by yourself. In a large-scale net shop, it is possible to accumulate behavior data of many users and make a proposal, but it is difficult to use a similar algorithm in a small-scale net shop because the amount of data is small. In this paper, we propose a recommendation algorithm for small datasets by focusing on the ordered structure obtained by formal concept analysis.

The total value of online retail sales in Japan exceeds 19 trillion yen [1], and for example, Amazon sells over 400 million products [2]. Since there are no shop assistants in online shops, customers often choose their own products. There are many products that can be compared, and a detailed investigation is possible if you take time. However, there are so many that it is difficult to judge which products are appropriate. Many recommendation systems have been studied to solve this problem. Many recommendation systems use the algorithm using the phase structure. While phase structures have proven to perform well, they require a lot of data to operate, which makes them unsuitable for small or new retailers. In addition, it is difficult for humans to understand the phase structure of complex neural networks, so it is difficult to explain the reason for the recommendation. In this study, we propose an algorithm for recommendation by focusing on the ordered structure, which is easy for a human to understand. The user's thinking is structured as a concept lattice according to the degree of abstraction by the form concept analysis and used for recommendation. The goal is to recommend multiple products based on a combination of multiple products, such as a wish list.

2. Related research

2.1. Formal concept analysis

Formal concept analysis [3] is a technique to discover the latent relationships between objects from objects (objects) and their properties (attributes). A context K = (G, M, I) is a set of sets in which an object set G and an attribute set M are represented by a binomial relation I Ledger G × M and represents a relationship between an object and an attribute. If the object has attributes in the context table, then x is given. To generate a concept lattice from a context table. An example context table is shown in Table 1 An example concept lattice is shown in Figure. 1.

Table 1. Context Table Example

	cake	sweet	fruit	chocolate
A	X		X	X
B	X	X	X	
C	X			X

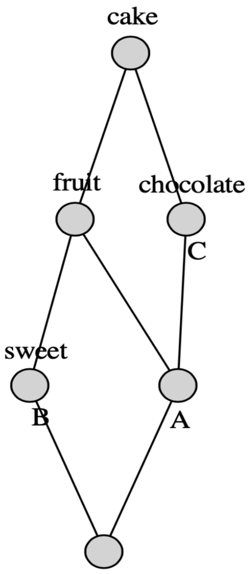

Fig. 1. Concept lattice Example

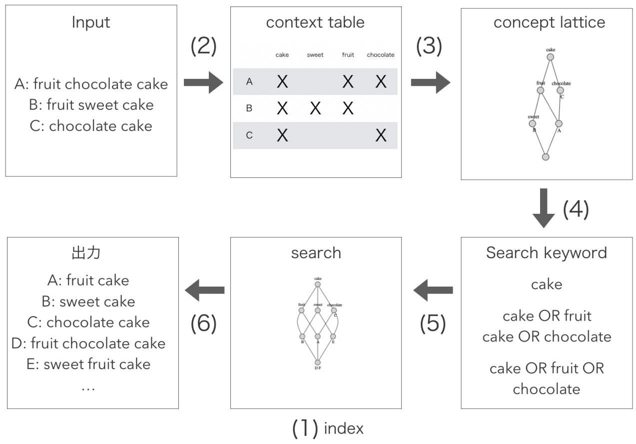

Fig. 2. Propose system: Product selection support system

2.2. A video recommendation method using formal concept analysis to clarify the reason for recommendation

Kitamura et al. [4] proposed a method to solve the problem of the small amount of information displayed on the recommendation screen in the video recommendation method of the video-sharing site YouTube. Specifically, we focus on the narrowness of the concept of tags attached to YouTube videos and apply formal concept analysis to these to recommend videos from various conceptual hierarchies to eliminate bias in a recommendation. In this study, an idea to make a recommendation in a limited small amount of data was obtained.

3. Propose system: Product selection support system

3.1. Outline of the proposed system

An overview of the proposed system is shown in Fig. 2. Input is one product group, and output is one product group. One product group created by the user in the application is an input, and the goal is to present one of the product groups created by other users as a candidate. As a preliminary preparation an already prepared product group is structured by using formal concept analysis and indexed to a search engine. For example, if there are 1000 existing product groups, a formal concept analysis is performed, and an index is performed on each of the 1000 product groups.

3.2. Preparation of the proposed system

To structure data in advance in order to make a recommendation system function. In the experiment, the search engine used Elasticsearch [5].

3.2.1. Create a context table from existing data(1)

A context table is prepared from the title and text of each item on one product group. The sentence is decomposed into words, and a context table is created, with each word appearing more than once as an attribute.

3.2.2. Creating a Concept Lattice(1)

Create a concept lattice from a context table.

3.2.3. Indexing into search engines(1)

Index concept lattice attributes to search engines. It is recorded in a layer 0 and a layer 1 ... in order from an upper node. Attributes may not correspond to layers in a concept lattice. In this case, the layer where the attribute appears first is set as layer 0. Each node is weighted for the scoring of retrieval. Multiply each layer down by the constant w ($0 < w < 1$).

Layer 0: 1
Layer 1: w
Layer 2: w^2
Layer 3: w^3
...
Layer 9: w^9

3.3. Implementation of recommendations by the proposal system

To retrieve and recommend data indexed to a retrieval engine.

3.3.1. Creating a Context Table(2)

The context tables are created in the same way as the preliminary preparation of the proposed system.

3.3.2. Creating a Concept Lattice(3)

The concept lattices in the same way as the preliminary preparation of the proposed system.

3.3.3. Determining Search Terms from a Concept Lattice(4)

To determine a search term according to the abstraction degree of a concept lattice. Search by the attributes of the concepts above the concept lattice. There are many possibilities in terms of how many attributes to use and how many to use. By searching with abstract concepts, it increases the possibility of recommending product groups that match the user's product group preferences.

3.3.4. Search(5)

Search is performed by a search engine. An example of a search query is shown below. In this method, a bool-should search by a combination of multiple multi_match queries.
"query":{ "bool" : { "should" : [{"multi_match": {"query": word, "fields" : layers,}}...] }}

3.3.5. Product group Presentation(6)

To present a corresponding product group to a user.

4. Experiment

4.1. Datasets

We used two datasets for this experiment. A. MovieLens 20M [6].
MovieLens 20M is a dataset of user reviews of movies. The results of 1 to 5 points for each movie watched by one user in increments of 0.5 points, user ID, movie title, and genre pairs are recorded. There is no fixed number of reviews created by a single user.
user example:
"userId","movieId","rating","timestamp"
1,2,3.5,2005-04-02 23:53:47
1,29,3.5,2005-04-02 23:31:16
1,32,3.5,2005-04-02 23:33:39
...
movie example
"movieId","title","genres"
1,"Toy Story (1995)","Adventure,Animation,Children,Comedy,Fantasy"
2,"Jumanji (1995)","Adventure,Children,Fantasy"
3,"Grumpier Old Men (1995)","Comedy,Romance"
...
We used movie names and genres for indexing and searching.
B. Yelp dataset [7]
Yelp dataset is a dataset of user reviews of businesses. The results of 1 5 points for each movie watched by one user, user ID, business name, and business category pairs are recorded. We used business names and business categories for indexing and searching.

4.2. Product group

The proposed method takes a product group as input and an output product group. The product group is used as a product selection support system by making the product group a favorite product group and proposing a similar favorite product group. For the MovieLens dataset, the list of movies rated 4.0 or higher by one user was one product group. For the Yelp dataset, a product group is a list of businesses with over 4.0 ratings per user.

4.3. Data Partitioning and Usage

In this experiment, data of 1000 users were indexed beforehand. We evaluated the output when entering data from another 1000 users. Both datasets use data from the near term. MovieLens divided the data sorted by user ID into ten parts and extracted more than 20 reviews with a rating of four or higher. Yelp divided the data sorted by Yelp's launch into three parts and extracted data with more than 20 reviews rated four or higher.

4.3.1. Metrics

The index r is the percentage of products expected in the proposed product group—the higher the r, the better the performance.
r = "probability that the product group in the input for the recommendation contains at least one product in the product group being recommended in the recommended product group"
The original product group has at least 20 products. Extract the input number of products (In this experiment, either 3, 5, or 10 was used.) from the original product group and create an input product group. Extracting ten products from the original product group that was not used in the input product group, if any of the ten products were included in the recommended product group, the process succeeded; otherwise, the process failed. The probability of success for each product group is r. During the product selection process, the product group that was recommended when some products (movie, business) were entered included products that were ultimately highly evaluated, and the recommendation was judged to be appropriate.

4.4. Experimental conditions

Weighting: Change the value of the weighting constant w to see how the results differ. Verify with w = 1.0 (no weighting), w = 0.8, w = 0.6. A number of items on the input product group: Change the number of products on the input product group to see how the results differ. The number of products is increased to 3, 5, and 10 for verification. To appropriately recommend even when the number of products is small.

Search words: In this experiment, or retrieval was carried out on each word in which the abstractness of concept lattice got by the form concept analysis became first attribute and second attribute.

4.5. Result

4.5.1. Metrics

The results are shown in Table2 and Table3. The vertical axis indicates the weight value, and the horizontal axis indicates the number of products to be recommended.

Table 2. MovieLens 20M Result

weight	three	five	ten
$w = 1.0$	55.5	55.6	**63.6**
$w = 0.8$	57.7	58.2	62.5
$w = 0.6$	**62.0**	**61.4**	62.4

Table 3. Yelp Result

weight	three	five	ten
$w = 1.0$	9.2	12.7	22.1
$w = 0.8$	10.5	15.9	29.5
$w = 0.6$	**10.7**	**17.5**	**31.3**

Since a certain degree of recommendation can be obtained from three products, it is considered that a certain degree of recommendation can be obtained even with a small amount of data by using formal concept analysis. With the exception of MovieLens 20 M' s 10 products inputs, weighting is a performance improvement over no weighting, so overall weighting works. As expected, the evaluation value increases as the number of input products increases. The benefit to users is expected to be higher than this figure because there is a good chance that a recommendation would be helpful. We also want to do a qualitative assessment of whether it's actually helpful to users.

4.6. Reason for a recommendation

The word used for the search was retrieved, as in the following example of the recommendation reason.
MovieLens: Adventure, Comedy, Action
Yelp: Restaurants, American, (Traditional), Pizza, (New), Burgers
This makes it possible to explain the reason such as, "Other people who are interested in 'Adventure', 'Comedy' and 'Action' see these products.". Since the past behavior history is not grasped in large quantity, it is considered that there may be an advantage from the privacy aspect.

4.7. Ordered structure

An example of an ordered structure (concept lattice) used in the system is shown in Fig. 3.

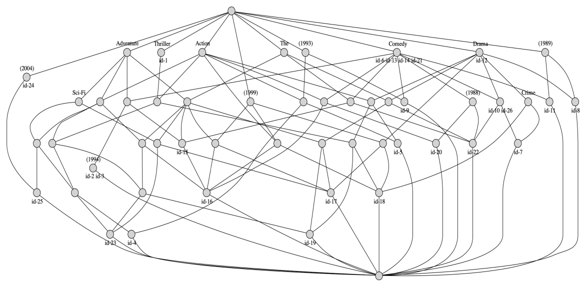

Fig. 3. Concept lattice Example (MovieLens)

5. Conclusion

In this paper, we propose a product selection support system for recommendation in limited data. It is considered that the possibility to obtain the recommendation result which the user expects was able to be indicated by using the ordered structure by the form concept analysis. In the future, we would like to conduct experiments on other datasets and verify them using actual users.

References:
[1] Ministry of Economy, Trade and Industry, Results of market research on e-commerce https://www.meti.go.jp/press/2020/07/20200722003/20200722003.html
[2] Amazon 2020 SME Impact Report released, https://amazon-press.jp/Top-Navi/RSS/Presse-release/amazon/jp/Corporate/20200812_Impact-Report/
[3] Wille R. (1982) Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. In: Rival I. (eds) Ordered Sets. NATO Advanced Study Institutes Series (Series C — Mathematical and Physical Sciences), vol 83. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-7798-3_15
[4] Formal Concept Analysis Based Movie Recommendation, Yutaro KITAMURA, Kazuhito SAWASE, Hajime NOBUHARA, Knowledge and information 2013 25-1 p.624-635
[5] Elasticsearch https://www.elastic.co/jp/
[6] MovieLens 20M Dataset https://grouplens.org/datasets/movielens/20m/
[7] Yelp open dataset, https://www.yelp.com/dataset