ICIC Express Letters
Part B: Applications
Volume **13**, Number **5**, May **2022**

 ISSN 2185-2766
pp. 463–470

FASHION OUTFIT RETRIEVAL USING HASHTAG SEARCH AND VISUALLY ASSISTED BROWSING ON JOINTED MANIFOLD MODELS

Shuri Hirowatari[1], Takuro Ishida[1], Tohru Iwasaki[2] and Tetsuo Furukawa[1,*]

[1]Graduate School of Life Science and Systems Engineering
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan
{ hirowatari.shuri637; ishida.takuro249 }@mail.kyutech.jp
*Corresponding author: furukawa@brain.kyutech.ac.jp

[2]ZOZO Technologies, Inc.
1-15-16 Midori-cho, Inage-ku, Chiba 263-0023, Japan
tohru.iwasaki@zozo.com

Received September 2021; accepted December 2021

Abstract. *In the fashion domain, the demand for outfit retrieval on social media and e-commerce sites is increasing. The most popular approach is searching outfits using hashtags, which has several limitations. First, users do not always know the appropriate hashtags for their search intention. Second, the hashtags determined by the posters are typically incomplete and inconsistent. Finally, it is not easy to browse when a large number of outfits share a hashtag. The purpose of this study is to develop a method for visually assisted fashion outfit retrieval. Specifically, the method enables users to search outfits using visually arranged hashtags in two-dimensional space, that is, the tag map, and browse outfits visually arranged in the outfit map. For this purpose, we use an approach in which jointed manifolds in style and hashtag spaces model the outfit data. The quantitative experimental result demonstrated that the proposed method preserves more mutual information between outfits and hashtags than the conventional method, which indicates that the proposed method is suitable for visual exploration.*

Keywords: Manifold model, Fashion outfit, Image retrieval, Tag search, Embedding method

1. **Introduction.** In the fashion domain, the demand for outfit retrieval on social media and e-commerce sites is increasing [1]. A vast number of outfit images are posted on social media, such as Instagram and Twitter, and users can now seek their favorite outfits. Furthermore, users can purchase items from these outfits, which are linked to e-commerce sites. It has been reported that 81.7% of female college students in Japan search for outfits on social media [2].

The most popular means of outfit retrieval is to search outfits using keywords. In particular, users search their favorite outfits by specifying hashtags (hereafter referred to as *tags*), which are attached to the outfits by the posters. However, such tag-based outfit retrieval is insufficient for the following reasons. First, users typically do not know the appropriate tags for their search intention. In some cases, users cannot express their target outfit styles or tastes verbally. Even if a list of tags is provided, it is difficult for users to choose an appropriate one from the huge number of available tags. Second, the tags determined by the posters are typically incomplete and inconsistent. Often, some necessary tags are missing, and different tags are used to express the same meaning. Finally, after executing a tag search, users are faced with a huge number of outfits that

DOI: 10.24507/icicelb.13.05.463

match their specified tag. For users, it is difficult to locate favorites from unsorted outfits or identify additional tags for search refinement.

The aim of this study is to develop a method for fashion outfit retrieval. The proposed method visually assists users in outfit retrieval using the style and tag, and it assists users in identifying appropriate tags. The key idea is to model the outfit data using jointed manifold models in the style and tag spaces. These manifold models embed the outfits and tags into two-dimensional spaces, that is, *the outfit map* and *tag map*, where users can browse them visually and interactively.

A conceptual illustration of the proposed method is shown in Figure 1. This method organizes two maps, *the outfit map* and *tag map*, where similar outfits/tags are arranged close to each other. By specifying *the target tag of interest (TTOI)* in the tag map, the system magnifies the tag map around the TTOI. By moving the TTOI, users can trace the continuous change of tags. Simultaneously, the system indicates the areas of the outfit map related to the TTOI using grayscale (described using blue balloons). Similarly, when the user selects *the target outfit of interest (TOOI)* in the outfit map, the system magnifies the area around the TOOI and displays the outfits in that area. Simultaneously, the system indicates the corresponding areas in the tag map related to the TOOI (described using orange balloons). Thus, the proposed method allows the visual exploration of outfits without the user knowing the exact tag.

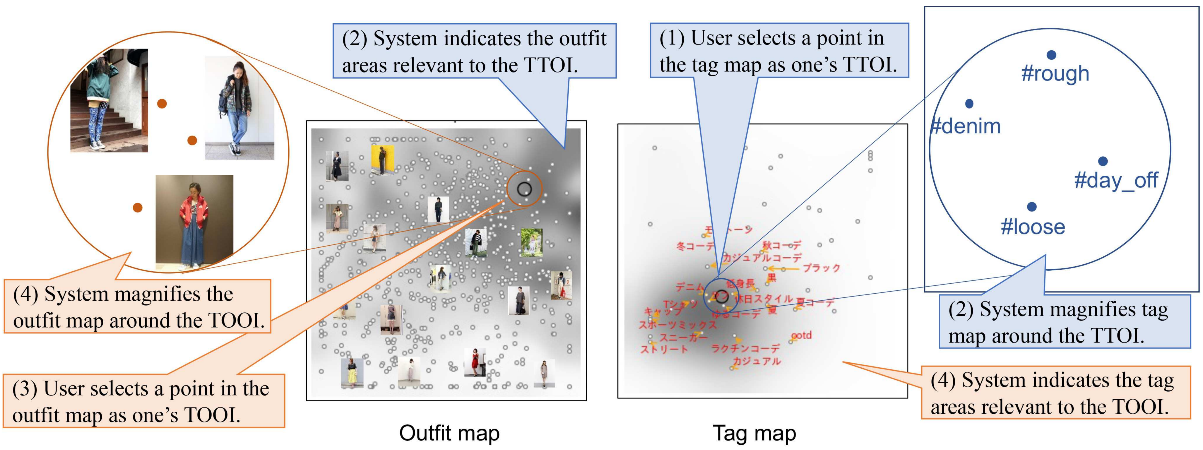

Figure 1. (color online) Conceptual illustration of the proposed method

The remainder of this paper is organized as follows. In Section 2, the background and related work are introduced. The proposed method is presented in Section 3, and some application results are presented in Section 4, followed by a discussion in Section 5. The paper is concluded in Section 6.

2. Background and Related Work.

2.1. Data-driven approach in the fashion domain. As in other domains, the data-driven approach is rapidly becoming widespread in the fashion domain. A typical example is fashion retrieval, which aims to specify items or outfits suitable for users' search intentions [3, 4]. For this purpose, annotation information (e.g., styles of outfits) is generally required for each item/outfit. Style annotation is typically performed manually; however, style estimation and automated annotation are increasingly being considered [5, 6]. Another important issue is fashion retrieval using a query image, where users indicate an image as their query instead of using keywords [3].

Item/outfit recommendation is also an important and challenging theme in the fashion domain. Typically, the compatibility matching of items is the main issue in outfit recommendation [7]. In particular, personalized recommendation in which the system is required to recommend outfits that suit each user is a challenging issue [8, 9].

While most of previous studies focused on automatic retrieval or recommendation, the aim of the present study is to develop visually assisted outfit retrieval using query tags; that is, the target system is *not* aimed at discovering outfits on behalf of users; rather, it aims to *assist* users in discovering favorite outfits by themselves. This approach allows users to explore outfits freely in the outfit space. It also gives users the opportunity to discover unexpected outfits through serendipity.

2.2. **Manifold models and visual analytics for relational data.** The key idea of this work is to model outfit data using jointed manifold models in probability spaces. This method is based on two existing methods: the tensor self-organizing map (TSOM) [10] and unsupervised kernel regression (UKR) [11]. The TSOM is an extension of the conventional self-organizing map (SOM) from a multivariate dataset to a relational dataset. The TSOM also provides a visual exploration method in a set of low-dimensional spaces, called the conditional component plane. In this study, we regard tag information as the relational dataset of the outfits and tags, and apply the TSOM approach. The difference between our method and the TSOM is that the TSOM must discretize the low-dimensional spaces, whereas the proposed method treats them as continuous spaces. For this purpose, we introduce UKR, which is another extension of the SOM from a discretized low-dimensional space to a continuous space. Thus, our method can be regarded as an extension of UKR from a multivariate dataset to a relational dataset. Additionally, the proposed method measures errors between the data and model using the Kullback-Leibler (KL) divergence, whereas the TSOM and UKR use the Euclidean distance in the data space.

The proposed method can also be regarded as the visual analytics of outfit data. Visual analytics is a concept in visually assisted interactive data exploration [12]. At the core of visual analytics is a human-centered knowledge discovery process, where users are the central players in knowledge discovery, and the role of the system is to assist the players. Watanabe et al. proposed a visual analytics method for team formation support based on manifold models [13]. Although the problem settings and data structures of team formation support are quite different from those of fashion outfit retrieval, and thus the method of Watanabe et al. cannot be used in our study directly, we propose a similar approach inspired by that method.

3. **Proposed Method.** In this section, we first define the problem setting, that is, the input and output of the system. Then, we outline the proposed method. We also explain how a visual search is performed using the proposed method.

3.1. **Problem setting.** Suppose that we have a set of outfits posted on social media, each of which has tags attached by the posters. Let $\mathcal{O} = \{outfit_1, \ldots, outfit_N\}$ be the set of outfits and let $\mathcal{T} = \{tag_1, \ldots, tag_M\}$ be the set of tags. Then, the tag information is represented by a binary matrix $\mathbf{T} = (T_{nm}) \in \{0,1\}^{N\times M}$, where $T_{nm} = 1$ if $outfit_n$ has tag_m, and $T_{nm} = 0$ otherwise. In the proposed method, $\mathbf{T}$ is transformed to $\mathbf{Q} = \left(Q_{nm}\right) \in [0,1]^{N\times M}$, where $Q_{nm} = T_{nm}/\sum_{m'} T_{nm'}$. Thus, Q_{nm} represents the empirical probability that $outfit_n$ has tag_m, that is, $Q_{nm} = Q(tag_m \mid outfit_n)$.

In addition to tags, we assume that each outfit has a style annotation given by stylists, such as "*formal*" and "*casual*". Suppose that $\mathcal{S} = \{style_1, \ldots, style_L\}$ is the set of style labels and the style annotation of $outfit_n$ is an L-dimensional vector $\mathbf{p}_n = (P_{n1}, \ldots, P_{nL})$, where $P_{nl} \in [0,1]$ and $\sum_l P_{nl} = 1$. We assume that an outfit can be classified as having multiple styles. For example, if $outfit_n$ is classified as "*natural*" and "*casual*" with the same degree, then $P_{nl} = 0.5$ for these two styles, and $P_{nl} = 0$ for other styles. Therefore, P_{nl} represents the empirical probability $P_{nl} = P(style_l \mid outfit_n)$. Consequently, the entirety of the annotation information is represented by the matrix $\mathbf{P} = (P_{nl}) \in [0,1]^{N \times L}$. To summarize, $\mathbf{Q}$ and $\mathbf{P}$ are the inputs of the proposed method.

By contrast, the outputs are low-dimensional vectors of the outfits and tags, which are referred to as latent variables. Let $\mathbf{Z}^{outfit} = \left(\mathbf{z}_1^{outfit}, \ldots, \mathbf{z}_N^{outfit}\right)$ and $\mathbf{Z}^{tag} = \left(\mathbf{z}_1^{tag}, \ldots, \mathbf{z}_M^{tag}\right)$, where $\mathbf{z}_n^{outfit} \in \mathcal{Z}^{outfit}$, $\mathbf{z}_m^{tag} \in \mathcal{Z}^{tag}$ are the latent variables of the outfits and tags, respectively. For the sake of visualization, we assume that the latent spaces are $\mathcal{Z}^{outfit/tag} = [0,1]^2$ and the prior of the latent variables is the uniform distribution on $\mathcal{Z}$. In addition to the latent variables, the method estimates two probability densities, $p\colon \mathcal{Z}^{outfit} \times \mathcal{S} \to \mathbb{R}_+$ and $q\colon \mathcal{Z}^{outfit} \times \mathcal{Z}^{tag} \to \mathbb{R}_+$, which represent the conditional probability $p\left(style \mid \boldsymbol{\zeta}^{outfit}\right)$ and joint probability $q\left(\boldsymbol{\zeta}^{outfit}, \boldsymbol{\zeta}^{tag}\right)$, respectively. To summarize, the outputs of the proposed method are the latent variables $\mathbf{Z}^{outfit}$ and $\mathbf{Z}^{tag}$, and probability densities p and q. The entire model is illustrated in Figure 2, and represented by jointed manifolds.

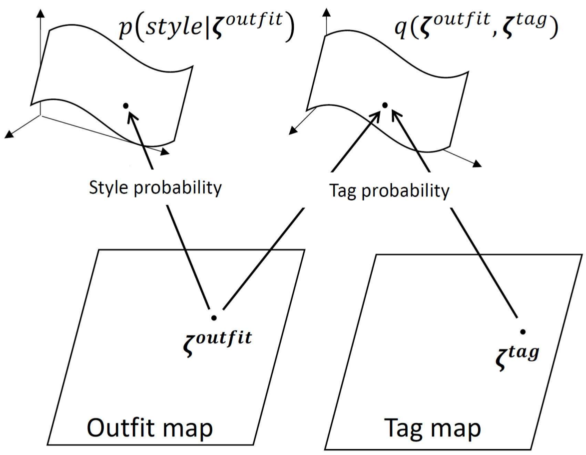

FIGURE 2. Jointed manifold models of the outfits and tags

3.2. **Learning method.** In the proposed method, the model distributions $p\left(style \mid \boldsymbol{\zeta}^{outfit}\right)$, $q\left(\boldsymbol{\zeta}^{outfit}, \boldsymbol{\zeta}^{tag}\right)$ are obtained using kernel density estimation as follows:

$$p\left(style_l \mid \boldsymbol{\zeta}^{outfit}\right) = \frac{1}{K\left(\boldsymbol{\zeta}^{outfit}\right)} \sum_{n=1}^{N} k\left(\boldsymbol{\zeta}^{outfit} \mid \mathbf{z}_n^{outfit}\right) P\left(style_l \mid outfit_n\right), \tag{1}$$

$$q\left(\boldsymbol{\zeta}^{outfit}, \boldsymbol{\zeta}^{tag}\right) = \frac{1}{N} \sum_{n=1}^{N} \sum_{m=1}^{M} k\left(\boldsymbol{\zeta}^{outfit} \mid \mathbf{z}_n^{outfit}\right) k\left(\boldsymbol{\zeta}^{tag} \mid \mathbf{z}_m^{tag}\right) Q\left(tag_m \mid outfit_n\right), \tag{2}$$

where $K\left(\boldsymbol{\zeta}^{outfit}\right) := \sum_n k\left(\boldsymbol{\zeta}^{outfit} \mid \mathbf{z}_n^{outfit}\right)$ and $k(\cdot \mid \cdot)$ is the kernel function. In this study, we use the Gaussian kernel $k(\boldsymbol{\zeta} \mid \mathbf{z}) := \mathcal{N}(\boldsymbol{\zeta} \mid \mathbf{z}, \sigma^2 \mathbf{I})$. Note that p and q are represented by $\mathbf{Z}^{outfit}$ and $\mathbf{Z}^{tag}$ non-parametrically.

To determine $\mathbf{Z}^{outfit}$ and $\mathbf{Z}^{tag}$, we define the objective function as the KL-divergence between the empirical and model distributions. The objective function with respect to styles is defined as follows:

$$\begin{aligned} F_{\text{style}} &:= \frac{1}{N} \sum_{n=1}^{N} D_{\text{KL}}\left[P\left(style \mid outfit_n\right) \middle\| p\left(style \mid \mathbf{z}_n^{outfit}\right)\right] \\ &= -\frac{1}{N} \sum_{n=1}^{N} \sum_{l=1}^{L} P_{nl} \log p\left(style_l \mid \mathbf{z}_n^{outfit}\right) - const. \end{aligned} \tag{3}$$

Similarly, the objective function with respect to tags is defined as $D_{\mathrm{KL}}\left[Q \,\middle\|\, \tilde{Q}\right]$, where $\tilde{Q}(\mathit{outfit}, \mathit{tag})$ is the joint distribution of the outfits and tags reconstructed from q:

$$\begin{aligned}
&\tilde{Q}(\mathit{outfit}_n, \mathit{tag}_m) \\
&:= \iint p\left(\mathit{outfit}_n \mid \boldsymbol{\zeta}^{\mathit{outfit}}\right) p\left(\mathit{tag}_m \mid \boldsymbol{\zeta}^{\mathit{tag}}\right) q\left(\boldsymbol{\zeta}^{\mathit{outfit}}, \boldsymbol{\zeta}^{\mathit{tag}}\right) d\boldsymbol{\zeta}^{\mathit{outfit}} d\boldsymbol{\zeta}^{\mathit{tag}} \\
&= \iint \frac{k\left(\boldsymbol{\zeta}^{\mathit{outfit}} \mid \mathbf{z}_n^{\mathit{outfit}}\right)}{\sum_{n'} k\left(\boldsymbol{\zeta}^{\mathit{outfit}} \mid \mathbf{z}_{n'}^{\mathit{outfit}}\right)} \frac{k\left(\boldsymbol{\zeta}^{\mathit{tag}} \mid \mathbf{z}_m^{\mathit{tag}}\right)}{\sum_{m'} k\left(\boldsymbol{\zeta}^{\mathit{tag}} \mid \mathbf{z}_{m'}^{\mathit{tag}}\right)} q\left(\boldsymbol{\zeta}^{\mathit{outfit}}, \boldsymbol{\zeta}^{\mathit{tag}}\right) d\boldsymbol{\zeta}^{\mathit{outfit}} d\boldsymbol{\zeta}^{\mathit{tag}}. \qquad (4)
\end{aligned}$$

Instead of minimizing $D_{\mathrm{KL}}\left[Q \,\middle\|\, \tilde{Q}\right]$, the proposed method minimizes its upper bound. By applying Jensen's inequality, we obtain the upper bound as follows:

$$D_{\mathrm{KL}}\left[Q \,\middle\|\, \tilde{Q}\right] \leq I\left[\mathit{outfit}; \mathit{tag}\right] - I\left[\boldsymbol{\zeta}^{\mathit{outfit}}; \boldsymbol{\zeta}^{\mathit{tag}}\right]. \qquad (5)$$

Thus, the objective function F_{tag} is defined by the negative mutual information between $\boldsymbol{\zeta}^{\mathit{outfit}}$ and $\boldsymbol{\zeta}^{\mathit{tag}}$ as

$$F_{\mathit{tag}} := \sum_{n=1}^{N} \sum_{m=1}^{M} Q(\mathit{outfit}_n, \mathit{tag}_m) \log\left[\frac{q(\mathbf{z}_n^{\mathit{outfit}}) q(\mathbf{z}_m^{\mathit{tag}})}{q\left(\mathbf{z}_n^{\mathit{outfit}}, \mathbf{z}_m^{\mathit{tag}}\right)}\right]. \qquad (6)$$

The entire objective function is $F := F_{\mathit{style}} + F_{\mathit{tag}}$, and the latent variables $\mathbf{Z}^{\mathit{outfit}}$ and $\mathbf{Z}^{\mathit{tag}}$ are estimated using a gradient-based method.

3.3. **Visually assisted outfit retrieval.** After the latent variables are estimated, the outfits and tags are mapped to the corresponding latent spaces similar to topographic maps, that is, *outfit map* and *tag map*. The proposed method provides various means for exploring these two maps visually and interactively.

(1) Exploring the outfit map. In the outfit map, the thumbnail images of representative outfits are displayed, where users can browse the outfit map. Suppose that a user wants to see the detail of a point in the outfit map. We refer to the point as *the TOOI*, denoted by $\boldsymbol{\zeta}_{\mathrm{TOOI}}^{\mathit{outfit}}$. When the user indicates the TOOI (e.g., clicks the point using a pointing device), the system magnifies the area around the TOOI and shows the outfit images in that area. By moving the TOOI continuously in the outfit map, the user can trace the continuous change of outfits.

(2) Visualization of style information. The outfit map can be colored according to the model distribution $p\left(\mathit{style} \mid \boldsymbol{\zeta}^{\mathit{outfit}}\right)$. For example, if a user wants to see the area related to style (say, style^*), then the outfit map is colored using grayscale according to $p\left(\mathit{style}^* \mid \boldsymbol{\zeta}^{\mathit{outfit}}\right)$. If the user wants to see the distribution of all styles simultaneously, the system can color-code the outfit map using the dominant style (Figure 3).

(3) Exploring the tag map. Similar to the outfit map, users can browse the tag map, where similar tags are mapped close to each other; hence, users can find appropriate tags easily. When a user selects *the TTOI* in the tag map, the system magnifies the tag map around the TTOI. It is also possible to color-code the tag map according to styles by evaluating $\mathbb{E}_{q\left(\boldsymbol{\zeta}^{\mathit{outfit}} \mid \boldsymbol{\zeta}^{\mathit{tag}}\right)}\left[p\left(\mathit{style} \mid \boldsymbol{\zeta}^{\mathit{outfit}}\right)\right]$.

(4) Outfit search via tags. When a user indicates the TTOI $\left(\text{say, } \boldsymbol{\zeta}_{\mathrm{TTOI}}^{\mathit{tag}}\right)$ in the tag map, the system indicates the relevant area in the outfit map according to $q\left(\boldsymbol{\zeta}^{\mathit{outfit}} \mid \boldsymbol{\zeta}_{\mathrm{TTOI}}^{\mathit{tag}}\right)$ (Figure 1). When the user moves the TTOI in the tag map, the user can see the continuous change of the corresponding area in the outfit map. This search method enables users to search outfits without knowing the correct tags attached to the target outfits. The

proposed method also enables users to conduct an inverse search from outfits to tags. In this case, the tag map is colored according to $q\left(\zeta^{tag} \mid \zeta^{outfit}_{\mathrm{TOOI}}\right)$.

The advantages of the proposed method are that it provides various visualization methods and has a simple and unified user interface; all users have to do is to indicate their points of interest via a TOOI or TTOI.

4. Experimental Results.

4.1. Method. We used 737 outfit data posted on WEAR, which is one of the largest fashion social media platforms in Japan[1]. The number of styles was 9 (*office, girly, sweet, street, natural, feminine, formal, basic, and mode*), and all outfits were annotated by 62 stylists. For these outfits, 1,585 tags were used. We removed tags that appeared rarely and chose 50 typical tags for the experiment.

We performed quantitative analysis by evaluating the mutual information between the latent variables of the outfits and tags. Thus, we evaluated $I\left[\zeta^{outfit}; \zeta^{tag}\right]$ using (6). We also evaluated the mutual information for the conventional matrix decomposition method, and compared it with the proposed method. Because the joint distribution should be positive, we chose non-negative matrix factorization (NMF) for the comparison. Note that the larger $I\left[\zeta^{outfit}; \zeta^{tag}\right]$ the better; its upper limit was $I\left[outfit; tag\right] = 2.71$ in this experiment.

4.2. Result. Figure 3 shows the obtained outfit map, which is color-coded by the dominant style. In the outfit map, outfits classified as having similar styles (e.g., *basic* and *natural*) are arranged close to each other, and the outfit map explains the continuous change of outfit styles. When the user selects a border point of two styles, the system shows the outfits with intermediate styles.

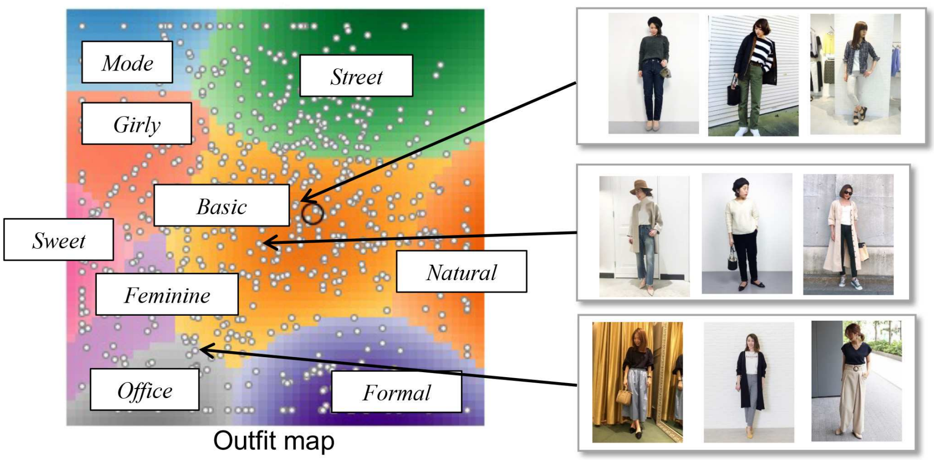

FIGURE 3. Outfit map color-coded by style. The color represents the dominant style of the outfits.

An example of tag search is shown in Figure 4. In this case, the user's TTOI was the tag `#elegant`. For this TTOI, the tags `#pump` and `#blouse` were arranged close to each other, which implies that these tags were often attached to `#elegant`. Simultaneously, the probability $p\left(\zeta^{outfit} \mid \mathbf{z}^{tag}_{\texttt{\#elegant}}\right)$ was indicated in the outfit map using grayscale. When the user selected a dark point in the outfit map as the TOOI, the user could see outfit images that may have the tag `#elegant`. In the figure, the TOOI is almost on the border of the"*office*" and "*feminine*" styles, which implies that the displayed outfits had the intermediate style.

[1]All data used were provided by ZOZO, Inc. and were used with permission.

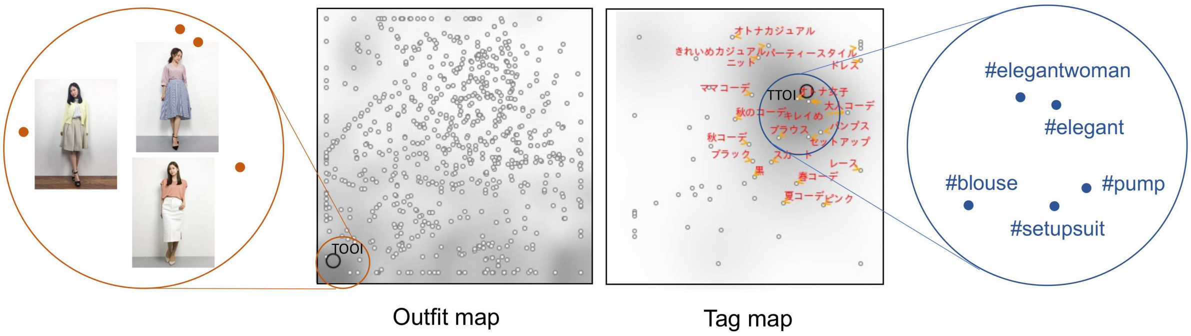

Figure 4. Example of a visual tag search

Figure 5 shows the mutual information between the latent variables of the outfits and tags for various dimensions of the latent spaces. As shown in the figure, the mutual information of the proposed method was greater than that of the NMF, which implies that the proposed method preserved more information than the NMF. Because the aim of this study is to develop a visual exploration system, it is desirable to preserve a large amount of mutual information, particularly in the two-dimensional case.

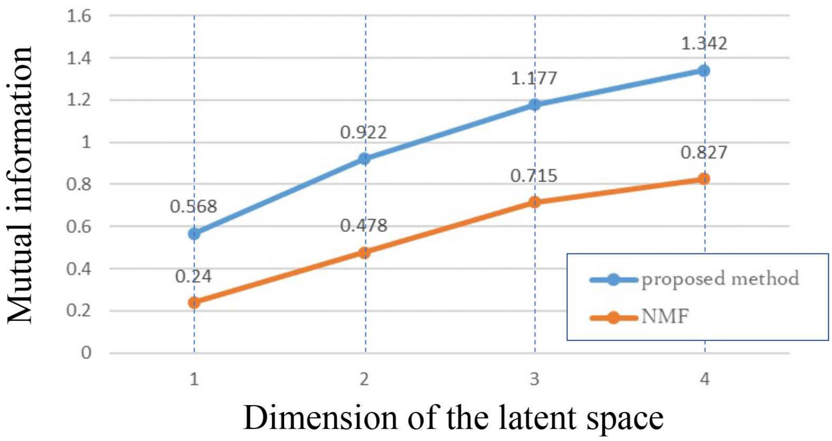

Figure 5. Mutual information of the latent variables estimated using the proposed method and non-negative matrix factorization (NMF)

5. **Discussion.** Conventionally, the aim of fashion retrieval is to determine outfits that match a user's query. By contrast, the proposed method assists users in finding outfits, but it does not determine fits on behalf of users. Additionally, the proposed system helps users to acquire knowledge about fashion through visualization. This is another useful aspect of the proposed method, particularly for users who are not familiar with fashion. Furthermore, the system gives users the opportunity to discover outfits serendipitously by browsing the outfit map, which is a challenging theme in the recommendation system field [14]. Therefore, the proposed method does not merely improve the performance of conventional methods, but also provides a new perspective in the field of fashion tech.

Currently, the proposed method does not consider the compatibility of items. We believe that this is the most important issue for consideration in future work. To address this challenge, the concept of visual analytics for set data would be useful [13].

From a machine learning viewpoint, the proposed method is an extension of manifold modeling from multivariate data to joint probability data. Because this technique is not limited to outfit data, the proposed method can be applied in other fields. In particular, the maximization of mutual information provides a new viewpoint in co-occurrence data embedding.

6. **Conclusion.** In this paper, we proposed a method of fashion outfit retrieval using hashtag search. Unlike the conventional approach, the proposed method does not aim to determine fits according to user queries; rather, it assists users in determining fits in a human-centered manner. Because fashion has an aspect of social communication, such a human-centered approach will become more important than the traditional automation approach.

Acknowledgment. This work was supported by ZOZO Research. This work was partially supported by JSPS Kakenhi, Grant Number 21K12061.

REFERENCES

[1] S. Jaradat, N. Dokoohaki, K. Hammar, U. Wara and M. Matskin, Dynamic CNN models for fashion recommendation in Instagram, *Proc. of the 16th IEEE International Symposium on Parallel and Distributed Processing with Applications, 17th IEEE International Conference on Ubiquitous Computing and Communications, 8th IEEE International Conference on Big Data and Cloud Computing, 11th IEEE International Conference on Social Computing and Networking and 8th IEEE International Conference on Sustainable Computing and Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom2018)*, pp.1144-1151, 2018.

[2] A. Watanabe, *Research on the Fashion Awareness of Modern Youth through an Actual Survey of "Ryousangata" Fashion (Mass-Produced Fashion)*, Annual Bulletin, Department of the Science of Living, Kyoritsu Women's Junior College, 2018 (in Japanese).

[3] A. Kovashka, D. Parikh and K. Grauman, WhittleSearch: Interactive image search with relative attribute feedback, *International Journal of Computer Vision*, vol.115, no.2, pp.185-210, 2015.

[4] B. Zhao, J. Feng, X. Wu and S. Yan. Memory-augmented attribute manipulation networks for interactive fashion search, *Proc. of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2017)*, pp.6156-6164, 2017.

[5] K. Yamaguchi, M. H. Kiapour and T. L. Berg, Paper doll parsing: Retrieving similar styles to parse clothing items, *Proc. of the IEEE International Conference on Computer Vision*, pp.3519-3526, 2013.

[6] Z. Liu, P. Luo, S. Qiu, X. Wang and X. Tang, DeepFashion: Powering robust clothes recognition and retrieval with rich annotations, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.1096-1104, 2016.

[7] T. Iwata, S. Watanabe and H. Sawada, Fashion coordinates recommender system using photographs from fashion magazines, *International Joint Conference on Artificial Intelligence (IJCAI)*, pp.2262-2267, 2011.

[8] X. Li, X. Wang, X. He, L. Chen, J. Xiao and T.-S. Chua, Hierarchical fashion graph network for personalized outfit recommendation, *Proc. of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2020)*, pp.159-168, 2020.

[9] Y. Hu, X. Yi and L. S. Davis, Collaborative fashion recommendation: A functional tensor factorization approach, *Proc. of the 23rd ACM International Conference on Multimedia (MM2015)*, pp.129-138, 2015.

[10] T. Iwasaki and T. Furukawa, Tensor SOM and tensor GTM: Nonlinear tensor analysis by topographic mappings, *Neural Networks*, vol.77, pp.107-125, 2016.

[11] P. Meinicke, S. Klanke, R. Memisevic and H. Ritter, Principal surfaces from unsupervised kernel regression, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, no.9, pp.1379-1391, 2005.

[12] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer and G. Melançon, Visual analytics: Definition, process, and challenges, in *Information Visualization. Lecture Notes in Computer Science*, A. Kerren, J. T. Stasko, J.-D. Fekete and C. North (eds.), Berlin, Heidelberg, Springer, 2008.

[13] R. Watanabe, H. Ishibashi and T. Furukawa, Visual analytics of set data for knowledge discovery and member selection support, *Decision Support Systems*, vol.152, no.1, DOI: 10.1016/j.dss.2021.113635, 2021.

[14] D. Kotkov, S. Wang and J. Veijalainen, A survey of serendipity in recommender systems, *Knowledge-Based Systems*, vol.111, pp.180-192, 2016.