2019 22nd International Conference on Control Systems and Computer Science (CSCS)

An Intelligent Personalized Fashion Recommendation System

Cristiana Stan
Computer Science Department
University Politehnica of Bucharest
Bucharest, Romania
cristiana.stan@stud.acs.upb.ro

Irina Mocanu
Computer Science Department
University Politehnica of Bucharest
Bucharest, Romania
irina.mocanu@cs.pub.ro

***Abstract*—Creating an outfit is a problem that is based on the preferences of each person and it can be difficult even for the best experts. This paper presents an automated system that can recommend a full outfit based on a cloth item considering also user's preference. Two convolutional neural networks based on the AlexNet model are used to identify cloth items and attributes associated with each item. After that, two types of scores are used in order to evaluate the user's preference for combination of different items, that are continuously updated in order to obtain recommendations that are more suitable for each user.**

***Index Terms*—fashion recommendation, convolutional neural network, user preferences**

I. Introduction

From the oldest times there are fashion recommendations for dressing according to each period. Over time there have been various fashion periods that are devoted to certain types of clothing. To be able to create an outfit there must be tacking into account many more factors, such as what is fashionable, but also the style and personal preferences of each person.

The notion of style is a subjective one, depending on the interests and personality of each one. Thus, there is a need for an application that makes recommendations on the basis of the user's preferences.

The purpose of this paper is to recommend an outfit, starting from a clothing apparel, by reducing the uncertainties as to which would be the most appropriate clothes to complete the clothing outfit.

The proposed recommendation system will recommend a full outfit based both on a cloth item (provided as an image) and on user's preferences. Thus, first of all, the type of the clothing item must be computed together with its associated attributes. After that, it is necessary to have a set of rules that specify the degree of compatibility of the garments, taking into account their properties. The objective is to assign a series of scores to various combinations of clothing types and styles in order to make the most relevant recommendations that can be dynamically improved and also taking into account the user's preferences.

This work was founded by the CCDI UEFISCDI and of the AAL Programme with co-funding from the European Union's Horizon 2020 research and innovation programme project "INCARE – Integrated Solution for Innovative Elderly Care", project number AAL-2017-059-INCARE.

Also, another important factor that must be considered is the personality of each user. To address this problem, a personal wardrobe is created and a set of scores are assigned based on both fashion rules and on user's preferences. These scores are dynamically updated according to the user's choices.

The rest of the paper is organised as follows: Section II describes related work. Section III gives the description of the proposed system. Evaluation of the proposed method is explained in Section IV. Conclusions and future work are given in Section V.

II. Related Work

Article [1] provides a way to make recommendations focusing on the notion of how it is composed. The proposed way to get the best results is to use only images that contain the entire human body, so a complete dress with all clothing pieces that make up it. The selected images are used in a learning process that is based on their visual characteristics. There are more visual details that are used in the training process, including the textures and colors of the elements that make up the gowns. Another factor that is essential to this approach is to determine, from the image, the component parts, namely the upper part and the lower part of the outfit. The described method gives a recommendations for completing a piece that already has a piece in its composition. In other words, on the basis of an image offered as input by the user, it is desired to propose clothing items that matches the already existing one. Individual pieces of clothing are recognised by categories and attributes. No other metadata for clothes are needed, all information are automatically extracted by the application either in the recommendations that are referring to visual features.

Article [2] focuses on a good understanding of the concept of holding and how well it is formed, an assessment that is made on the basis of a system of scores. The notion of wearing is defined as a set of pieces of clothes that fit not only in terms of stylistic rules but also as a creative combination. The creative component is a target quite difficult to achieve because it is based on the subjective side of the fashion industry. In addition, each piece of clothing is characterized by a large number of attributes that are difficult to capture by a common classification algorithm. Taking these into account, the solution proposed by the authors is to learn the composition of a score

2379-0482/19/$31.00 ©2019 IEEE
DOI 10.1109/CSCS.2019.00042

based on items' scores and not on individual characteristics of the clothing.

The solution proposed in this paper will recommend a set of items clothes in order to create a full outfit - it recommends a set of other pieces of clothing that best fit it in order to obtain the highest scores. In article [1] the scores are static in the sense that once the model has been trained, the learned scores are no longer altered, and the same predictions based on the attributes learned about the holdings are always used. The solution proposed in this paper considers scores as values that update dynamically, taking into account the user's preferences, given that all actions made by the user are recorded by constantly updating both the scores per cloth item and the combinations of clothes categories. Also, the scores show a measure of the user's preferences comparing with the solution from paper [1] where the scores represents a general mark for each outfit.

III. Proposed Solution

The user selects an image with a fashion item in order to receive recommendation for completing the outfit. Figure 1, describes the architecture of the fashion recommendation system.

The fashion recommendation system is composed of two main modules:

- *recognition module* that performs recognition of a cloth item together with its attributes; this classification divides each cloth item into 9 classes: bleizer, blouse, coat, dress, jacket, trousers, skirt, sweater, T-shirt, that are relevant both for the identification of the type of item (trousers, blouse, etc.), but also for the style of the cloth item (jeans jacket, leather jacket, etc.). Each item has associated a set of attributes, such as:
  - for bleizer, the attributes are: with front buttons or long
  - for blouse, the attributes are: with ribbon, made by lace or made by silk
  - for coat, the attributes are: made by fur or made by cotton
  - for cotton dress, the attributes are: midi, mini, or long
  - for jacket, the attributes are: bomber, leather or denim
  - trousers, the attributes are: straight, slim or leggings
  - skirt: the attributes are: mini, midi, long or A line
  - sweater: the attributes are: round collar or V Collar
  - T-shirt: the attributes are: round collar, V Collar or without sleeves

  Using these classification, a set of relationships between the different attributes of the clothes are created in order to define if the combination is suitable to be wear.
- *recommendation module* that provides fashion recommendation based on both fashion rules and user's preferences.

Each user is represented by a set of attributes such as: his wardrobe and the scores that describe his or her preferences. The clothes of the user are ranked when they are added into the application (are inserted into the *Cloth Database*). A matrix of scores: $score_{item_i,item_j}$ is computed for each pair of items and it is stored in the *User Preferences Database*). A score per each item: $item_i$ is used and it is stored into the *User Preferences Database*, too.

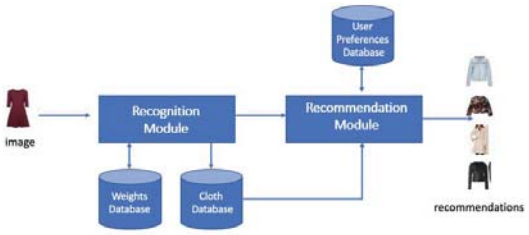

Fig. 1. The system architecture

The main steps of the *recognition module* are given bellow:

- the user selects an image with a clothing item
- the item is classified using a convolutional neural network as: bleizer, blouse, coat, cotton dress, jacket, trousers, skirt, sweater, T-shirt
- the attributes of the item are computed
- the classification and attributes extracted are computed using neural networks whose weights are saved in the *Weight Database*.

The *recommendation system* gives a set of recommendations to the user. Recommendations create an outfit based on a fashion item (specified through an image) using a set of predefined fashion rules combined with the user preferences. Two scores are used for the outfit recommendation:

- $score_{item_{user}}$ represents the score associated with a fashion item ($item_{user}$)
- $score_{item_i,item_j}$ represents the score associated of a combination of two fashion items ($item_i$ and $item_j$) that reflects the possibility of wearing together of the two items.

The main steps of the *recommendation system* are the following:

- attributes of the fashion item are extracted from the image provided by the user
- determine the elements from the *Cloth Database* that have similar attributes with the attributes of the fashion item provided by the user
- the selected items are sorted in descending order based on the score of combinations of fashion items, using $score_{item_i,item_j}$
- for each attribute, the selected items are sorted in descending order based on the user preferences (using $score_{item_{user}}$)
- for each attribute, the first *n* images are provided to the user

The user has the possibility to express their preference in order to select the most favourite outfit. The proposed method for this functionality follows the steps:

- the user selects the image with the preferred outfit
- the two scores: the score of the item pairs (item selected by the user and item recommended by the system) and the score for the recommended item associated to the user are updated (the values are increment by a threshold).

A. Recognition Module

Each piece of cloth must be recognised together with its associated attributes. The recommendations are based only on the garments owned by the user. Thus, the attributes of an item are identified when a new item is added to the wardrobe.

The proposed solution for recognition the category and attributes of a cloth item involves the use of neural convolutional networks organised in two layers. Both networks are based on the convolutional neural network using AlexNet network (see Figure 2) as described in [3], in which all AlexNet layers and their connections are graphically represented. Based on different experiments, we consider the first layer is a convolutional one having input dimensions of 224x224x3 and a filter equal to 96. The role of such a layer is to take a portion of the input data and convert it into a single number (the sum of all the input multiplied by the filter). The second layer is a MaxPooling type, followed by a convolution and it has an input size equal to 55x55x96. The role of a MaxPooling layer is to reduce the number of parameters within the model and to generalize the results of a convolutional filter. Next, there are three groups of such layers in the structure, followed by a six fully connected layers.

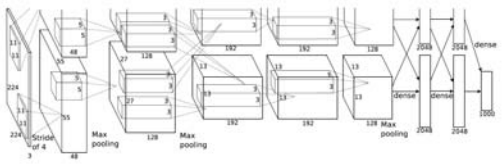

Fig. 2. Arhitectura modelului AlexNet [3]

The two layers of the convolutional neural network are:

- a convolutional network that it is used to obtain the category of the cloth item
- for each type of cloth we use another convolutional network to identify the attributes that characterized it.

Each network is trained specifically for its scope: item recognition, respectively attributes detection.

B. Recommendation Module

The recommendation system associates scores for pairs of cloth items in order to describe the degree of compatibility between two pieces of clothing and the user's preference for an outfit. We use integer values for scores and a larger value represents a higher probability for that item to be chosen within a recommendation that matches it.

Initially, only one type of score was used, representing the degree of compatibility between two clothing items. These scores are arranged in a square matrix where each cell corresponds to the matching between two fashion items. The matrix is populated with values that reflect the fashion rules between two items. They are distinct values. The only repeating scores are those that have a value equal to 0 and means that the two categories are not at all compatible, for example two subcategories of dresses (one person can not wear two dresses one over the other), so the recommendation would not make sense. When the user selects an image as preferred one, its score is updated, so that it becomes the maximum of all the recommended ones. In future, this attribute will be a priority for the recommendation.

As a result of the implementation of this functionality, it has been noticed that after several steps of improvement of these scores, for a particular piece of clothing, recommendations are made only of the preferred user attributes, without leaving it the chance to get other suggestions or change their preferences. In addition, the recommendations were no longer general and the outfit was not completed, for example a pair of trousers were suggested to be weared with a blouse, but not with a jacket. Thus, we consider a score associated with each element from the user's wardrobe. The proposed solution, selects a number of items according to the matrix scores (compatibility degree). The selected items, belonging to more than one category, are sorted by the score that reflects the user's preferences and then truncated to the required number of recommendations.

IV. Evaluation

The application is evaluated in two steps: i) first, it is evaluated only the *recognition module* and second the whole application is tested.

In order to evaluate the recognition module we use the dataset *DeepFashion* described in article [6] which contains categories of clothing together with its attributes, such as material, texture or print.

The data set contains cloth items that are annotated with attributes, parts of the localization of the different parts (for example, the location of the sleeve for a blouse or a collar of a shirt), but also attributes of clothing. In addition, for each image, a border is defined that exactly matches the piece of clothing in the picture, which is extremely useful, given that when using such images for training the background is not necessarily important only the central piece representing the purpose of learning. The data set consists of three distinct subsets: one used for category classification and the other for attribute prediction (*Category and Attribute Prediction Benchmark*), one for the store's stores (*In-Shop Clothes Retrieval* and one for similarities between buyer and shop clothes (*Consumer-to-Shop Clothes Retrieval*). We use only data from the (*Category and Attribute Prediction Benchmark*), being the most relevant, focusing on the user's personal wardrobe and makes recommendations only on the clothes he has at his

disposal without the need for new clothing purchases. A set of images that were used are given in 3.

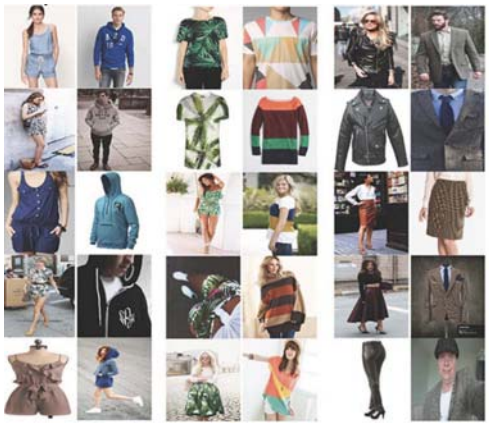

Fig. 3. Images from the dataset *Category and Attribute Prediction Benchmark*

The dataset consists of 63,720 pictures, covering 50 distinct categories and associating 1,000 different attributes. The accuracy of the annotations is high as the 50 categories have been manually attributed by field experts, and the attributes have been associated with an automated way, but have been checked against metadata data associated with the images. We use this information for creating categories and subcategories in which the images were divided for use in the training process, and only a few of the pictures were kept, selecting those that belonged to the most relevant categories for the work.

A. Evaluation of the Recognition Module

The proposed solution for recognizing the cloth items from an image involves the use of convolutional neural networks based on the AlexNet architecture. We used a pretrained network. The network is implemented using the PyTorch package [4]. The implementation uses two component packages: *torch* and *torhivision* from [4]. They provide both individual layers already implemented (convolutional, MaxPooling, etc.), but also models with all the already configured parameters and the layer architecture set. In addition to this, another package is *PIL* [5], which provides the features necessary for imaging operations: opening, saving and processing, resizing, cropping, etc.

The data set is divided into 20% test images and 80% images for training. As far as the test images are concerned, a score is calculated that reflects the percentage for which the predicted categories of the trained network have coincided with the known ones, represented by the labels already existing in the data set. This accuracy is 83%, which reflects the fact that the weights learned in this training phase are suitable for achieving the desired goal. The score obtained is good enough, compared with that obtained in the data set presentation.

We create a subset of the data set for the *recognition module*. The data set has been modified because there is no clear separation between what attributes belong to a category, but there were categories combined with attributes, such as evening dress, leopard print evening dress, etc. Thus, for only one very special class, there were very few training images, the average number being around 50. We want to differentiate between categories and attributes. Some images were removed, such as: removing part of the background or elements that are not part of the training and testing purpose. Images are labeled with categories and attributes. The clothes present in the dataset are classified into 9 categories and the associated subcategories (attributes), which can be seen in the Table I.

The data set thus obtained consists of nine main classes: jacket, blouse, coat, dress, jacket, trousers, skirt, sweater and t-shirt, each of which is in turn divided into several subcategories specific to the class. To get this result, several original categories were merged into one, for example, *round collar jersey and abstract print* was kept in the *T-shirt* and subcategory *round collar* category. The classes and sub-classes chosen as final were selected from the original so that they are as relevant as possible to the project purpose and cover a broad range of clothing pieces. Fashion articles have also been studied, specifying the most common categories of clothing, such as V-shaped collar jerseys, molded dresses, leather jackets and more.

After all these changes were achieved, an accuracy gain on test data was obtained by approximately 6-7 percent, increasing from 70% to 76%. To get the current accuracy, a data set was cleared, meaning some images were deleted or moved to other categories. There were mainly images where the piece of clothing that was intended to be classified was not completely visible or covered by another piece. Following this operation, a total of approximately 7,000 images were obtained, divided into 20% for testing and 80% for training. The final accuracy score obtained from all processing was 83%.

Several training processes were done: one for the general classification and one for each category, for computing the attributes of each item cloth. This is performed in virtual machine from Google, using the application Collaboratory [7]. 13GB of RAM and GPUs were made available for the faster Cuda. Each training had 50 epochs and lasted about 4 hours.

For the first layer of the network we obtained an accuracy of 83% for the testing data set, compared to the accuracy obtained in the presentation article, which has a value of 76.4%. The main changes that lead to this difference in scores are that the data set used in the project is cleared of images that are not relevant to the learning process and that more different clothes have been grouped under the same category. The results obtained for this first stage of training can also be seen in the confusion matrix obtained, shown in Figure 4.

The second level of network is trained to recognize attributes specific to each item cloth. All the images belonging to a category were divided into 80% and 20% respectively to facilitate the training process. Results are given in Table I.

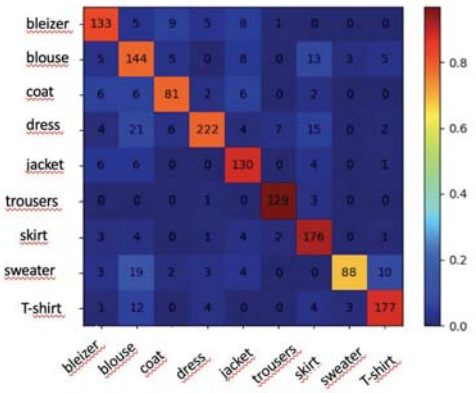

Fig. 4. Confusion matrix for each cloth item

TABLE I
ACCURACY FOR ATTRIBUTES DETECTION

Item Type	Accuracy	Attributes
bleizer	91.92%	with front buttons, long
blouse	65.57%	with ribbon. made by lace, made by silk
coat	93.2%	made by fur, made by cotton
cotton dress	92.17%	midi, mini, long
jacket	82.31%	bomber, leather, denim
trousers	79.25%	straight, slim, leggings
skirt	76.43%	mini, midi, long, A line
sweater	83.72%	round collar, V Collar
T-shirt	60.69%	round collar, V Collar, without sleeves

B. Evaluation of the Recommendation Module

The *recommendation module* uses the internal structures of the application (two types of scores) that have already been populated with scores that reflect user preferences. The scores are dynamically modified while an user uses the application.

In order to make a recommendation, we use the two types of scores. For the user input, its attributes are determined from the associated wardrobe and the corresponding matrix for the outfits is extracted. Each line is populated with score values that represent how well you can combine the current attribute with the rest of the attributes on that line. The elements from the line are sorted and a set of relevant attributes are extracted based on the user's query. Then, for each attribute, the clothes from the user's wardrobe are extracted. These are sorted according to the user's score and the first n options are displayed to the user into a graphical interface.

All elements related to the graphical interface can be seen in Figure 5, which identifies three instances of the open application from the same user's account. The first example is adding a new element and the categories determined by the application: *Maxi Dress*.

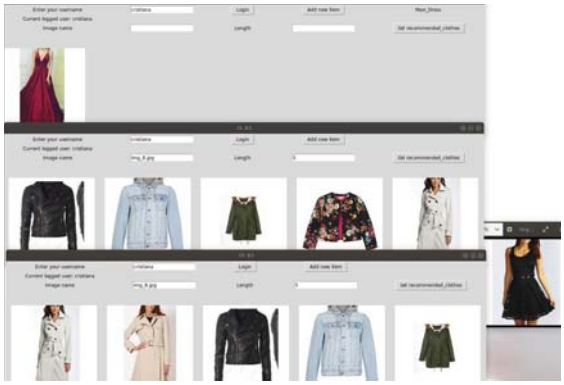

Fig. 5. Interface of the application

To test this application module, the personal side of different users was used. They have been subjected to creating a new account, starting from default values with standard values and requesting recommendations and selecting favorites combinations of fashion items. It is noted that recommendations are becoming more and more suited to users' preferences.

V. CONCLUSIONS AND FUTURE WORK

The paper presents a system for fashion recommendation, offering to the user a complete outfit based on a single fashion item (provided by the user through an image that contains a cloth item). The application can recognize 9 categories of clothes and different associated attributes that can be encountered.

Regarding the part of the recommendation, we considered recommendations provided by a set of users for creating a complete outfit considering the feedback provided by the user. In addition, the advances made are easy to note, as it is a matter of developing a solution that will help users to choose their choices more easily in their daily lives. The existing solution propose methods that are focused only on the composition of correct dresses from the point of view of the fashion rules of combining the garments.

The main drawback of using a two-layer convolutional neural network consists in having the classification process very time consuming. Two classifications (category and attributes) are required to determine the class and the attributes of a piece of clothing. Thus, two very large set of weights - 228MB weights - are loaded into the application. This operation makes the functionality of adding a new piece of clothing by the user to take about 10 seconds, that has to wait for the category of new clothing to be included. It is desirable to

determine optimizations that shorten the execution time of this functionality and a solution for retaining the weights involved in easier loading structures and smaller in size.

REFERENCES

[1] Iwata, Tomoharu and Watanabe, Shinji and Sawada and Hiroshi, Fashion Coordinates Recommender System Using Photographs from Fashion Magazines, Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, vol. 3, pp. 2262–2267, 2011

[2] Y. Li and L. Cao and J. Zhu and J. Luo, Mining Fashion Outfit Composition Using an End-to-End Deep Learning Approach on Set Data, IEEE Transactions on Multimedia, vol. 19, no. 8, pp. 1946–1955, 2017

[3] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E., ImageNet Classification with Deep Convolutional Neural Networks, Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1, pp. 1097–1105, 2012.

[4] PyTorch,https://pytorch.org/, Last accessed February 2019

[5] Pillow, https://pillow.readthedocs.io/en/5.1.x/, Last accessed February 2019

[6] Ziwei Liu, Ping Luo, Shi Qiu, Ziaogang Wang and Xiaoou Tang, DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1096–1104, 2016

[7] Colaboratory, https://colab.research.google.com/notebooks/welcome.ipynb, Last accessed February 2019