Body measure-aware fashion product recommendations: evaluating the predictive power of body scan data

Alexander Piazza
University of Erlangen-Nuremberg
Institute of Information Systems
Lange Gasse 20
Nuremberg 90403
alexander.piazza@fau.de

Jochen Süßmuth
University of Erlangen-Nuremberg
Chair of Computer Graphics
Cauerstraße 11
Erlangen 91058
jochen.suessmuth@fau.de

Freimut Bodendorf
University of Erlangen-Nuremberg
Institute of Information Systems
Lange Gasse 20
Nuremberg 90403
freimut.bodendorf@fau.de

ABSTRACT

Fashion product consumer are faced with large and fast changing product offerings. The fashion purchase decision process is complex, as the consumer has to consider various influencing factors like current fashion trends, what fashion products fit to their personality, and what products fit to their physical appearance like hair colors or body measures. Based on novel technologies, 3D body avatars can be reconstructed from 3D or 2D data. From these avatars, body measures can be determined. The objective of this research is to investigate the predictive performance of body measures extracted from a 3D body scanner for predicting fashion item preferences. Therefore, item preferences and body scans from 200 users were collected. From the body scans, 11 body measures are extracted and integrated into a prediction model using Factorization Machines. The results from a cross-validation show, that including body measurements significantly improves the prediction performance of the recommendation model, especially in new user scenarios, when no information about the fashion product preferences of the active user is known.

KEYWORDS

fashion product recommendation, body scan data, factorization machines, feature-driven recommendation

1 INTRODUCTION

Fashion products have the highest turnover of all product categories sold via e-commerce worldwide. Online fashion retailer offer their consumers large product ranges. For instance, the German online retailer Zalando indicates to have 150,000 products from 1,500 brands in their assortment [1].

Consumer behavior research indicates, that having a too broad product offer can cause a choice overload for the consumer, what leads to a delayed or no final purchase decision [7], and can decrease the consumers' satisfaction regarding their purchase decision [17]. This effect has also been identified within fashion online purchase decisions [11]. To avoid choice overload, fashion online stores should limit the options for consumers in a way, so they are displayed with a sufficient product variety but in the same time, they can make informed decisions [11]. Especially fashion purchase decisions are complex, as multiple aspects of the user are influecing factors like the personality, emotion, or general fashion trends as well as their physical appearance like the height [9], skin color, or body type [12]. The user has to decide not only based on what fashion products she or he in general likes, but also e.g., what colors fit his or her hair and skin color, and what cut of the clothes emphasizing or covers certain body parts.

Usually, recommender systems are applied to filter relevant products for the visitors of online stores, leveraging techniques like collaborative filtering or content-based filtering. The central assumption of collaborative filtering is, that users having demonstrated similar preferences in the past will have similar preferences in the future [3]. Therefore, conventional collaborative filtering based recommender systems only consider the product preferences per user in the form of ratings, views, or purchases and consequently derive similarities between users on these data. However, due to the wide adoption of new technologies like social networks, or mobile phones by the users as well as novel technologies to store and processing of large datasets, rich side information about users, items, and their interaction are available. Therefore, further research is needed to investigate mechanisms for integrating such rich side information into recommendation systems, as well as to assess the impact of such side information on the prediction quality [18].

Based on novel technologies, information about the users' physical appearance can be acquired. For instance, 3D models of bodies or faces can be constructed based on low-cost 3D scanners [21] [22], or reconstructed from 2D photos [8] [5]. User studies indicate that they perceive 3D scanner as useful for online shopping in general [20], as well as in the fashion product context [12], as long as data protection is ensured. However, according to a recent literature review about apparel product recommendation research, there is a research gap regarding the potential of body scan measurements to enrich consumer profiles in apparel recommendation [4].

The objective of this research is to investigate the predictive power of body measures for predicting fashion product preferences of individual users. Therefore, a dataset containing fashion product preferences as well as body measurements extracted from body scans are collected, and offline experiments are conducted to compare the impact of these measurements on the prediction accuracy. In this paper, first the data collection and the resulting data set is illustrated in Section 2, and preliminary results are demonstrated in Section 3. Finally, in Section 4 the results are discussed as well as planned next steps for further research are illustrated.

[1] https://www.zalando.de/presse_zahlen-und-fakten/

ComplexRec 2017, Como, Italy.

2 DATA COLLECTION

2.1 Data collection process

For data collection, volunteers were recruited from the School of Business at the University of Erlangen-Nuremberg. To obtain one rather large sample instead of two smaller ones, data collection was concentrated on only one gender. The decision was made to focus on female participants, as a higher variation in body shapes are expected as well as female participants are assumed to have stronger opinions according to their fashion preferences. The data collection process consists out of two steps. In the first step, participants were scanned using a low-cost body scanner which creates three-dimensional polygonal mesh data. Based on the mesh data, 11 key body measurements per participant were extracted. In the second step, the participants filled out an online questionnaire. In the online questionnaire, the participants gave information about their height and weight as well as indicated the color of their hair, eyes, and skin. Also, they classified their body shape into one of eight shapes. Finally, every participant indicated their preference towards 36 fashion products by answering the question, whether they would buy the displayed clothing or not on a seven-point Likert scale (1=do not want to buy it at all; 7= would buy it definitely). For the set of fashion products, upper an lower apparel that rather accentuate or mask certain body features were selected.

2.2 Resulting data-set

In total, 200 persons participated in the survey and body scanning. The distribution of the eleven body measures are illustrated in Figure 3. From the 200 valid scans, the average age is 22.35 and the standard deviation 2.94 years.

3 EVALUATION

3.1 Evaluation protocol

For determining the prediction quality, the rating is converted from the seven-point Likert scale to a binary rating, indicating whether a consumer would buy the product or not, by transforming ratings ≤ 5 as not interested (=0), and ratings > 5 as interested (=1). In the following, the term rating prediction is therefore interpreted as a classification problem. During the evaluation, the data is split up in a test and training set using a 10-fold cross-validation, as it is illustrated in Figure 4. Each user is randomly assigned to one of the ten test and train user-sets. Furthermore, to simulate the new user scenarios, the 36 items are randomly divided into 24 test and 12 train item-sets. The split is based on a random selection of items, which is illustrated in Figure 3, where the darker gray bars indicate the test items belonging to the test-set. For each iteration through the ten folds, the algorithm is provided with all 36 ratings of all of the users in the train user-set. Furthermore, to investigate the new user scenarios, the algorithm was provided with no, two, or five item ratings randomly selected from the train item-set. For all of the three scenarios, the ratings for the same 20 products of the test item-set are predicted.

For determining the prediction quality, the metrics precision (Equation 1), recall (Equation 2), and the F-measure (Equation 3) are used. In machine learning research, various aggregation approaches are used to aggregate the F-measures from the individual cross-validation results, which lead to diverging results. In this research, we use the formulation of the F-measure suggested by Forman and Scholz (2010) which resulted in the least bias [2].

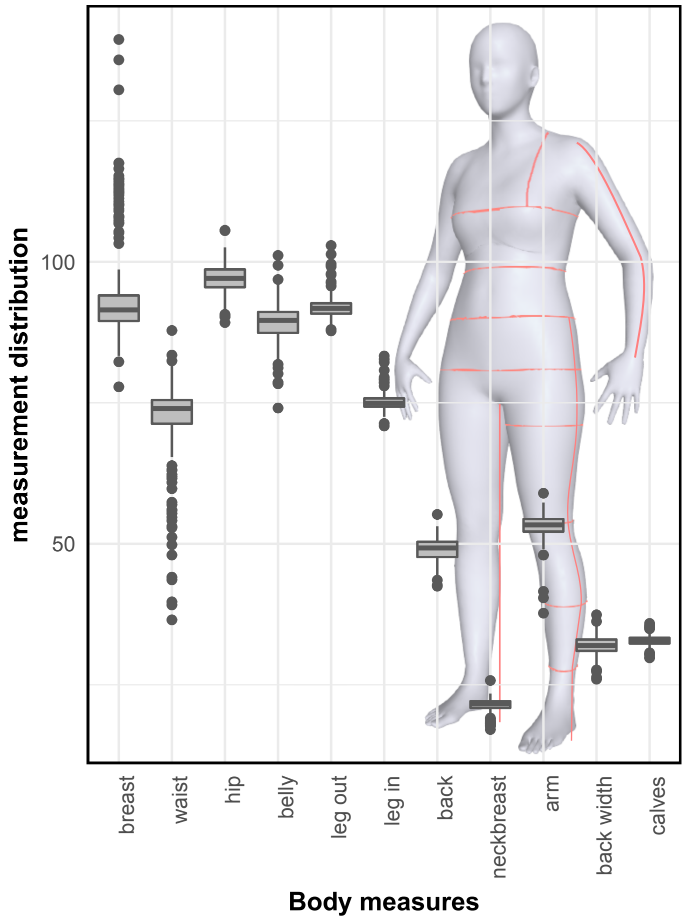

Figure 1: Distribution of the eleven body measures extracted from the body scans (N=200)

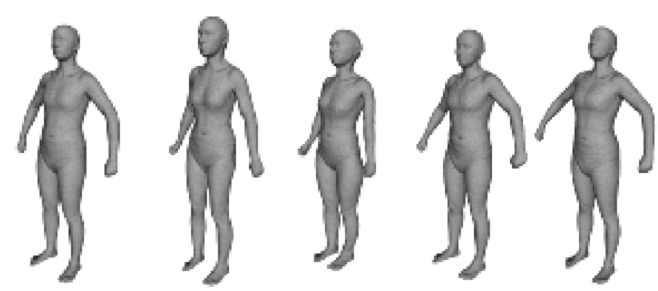

Figure 2: Exemplary polygonal meshes resulting from the body scans.

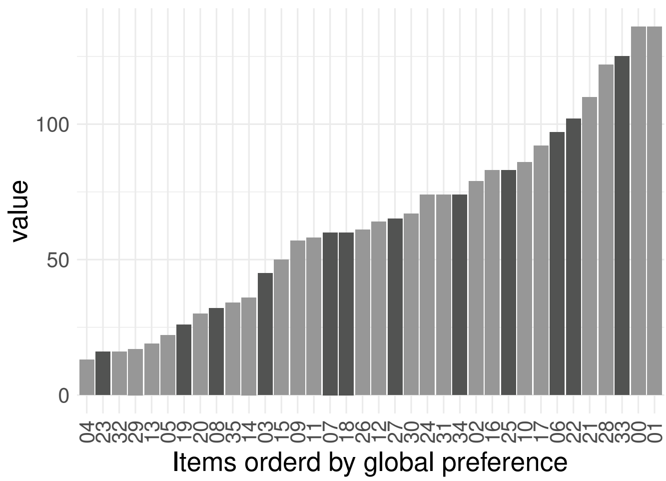

Figure 3: All 32 fashion products ordered by global preference. The products in darker gray color are selected for the test-set.

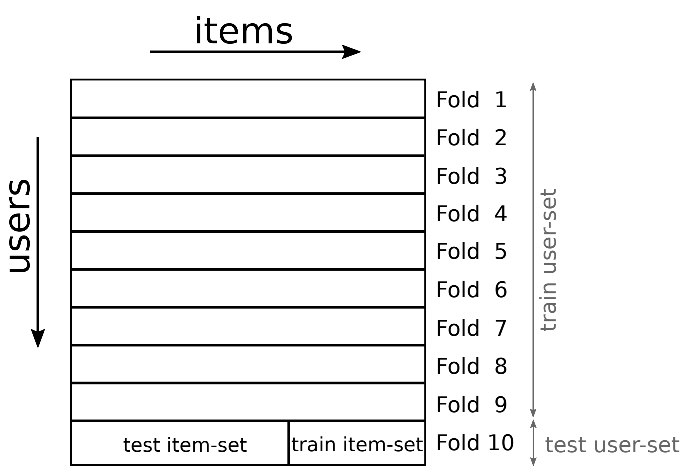

Figure 4: Illustration of the applied 10-fold cross-validation schema.

$$\text{precision} = \frac{\text{True Positive}}{\text{True Positive} \times \text{False Positive}} \tag{1}$$

$$\text{precision} = \frac{\text{True Positive}}{\text{True Positive} \times \text{False Negative}} \tag{2}$$

$$\text{F} = \frac{2 \times \text{True Positive}}{2 \times \text{True Positive} + \text{False Positive} + \text{False Negative}} \tag{3}$$

3.2 Factorization Machines

In recommender systems research, various approaches were suggested to consider additional information besides the ratings of users per item. The additional information can be integrated via pre-filtering or post-filtering, where conventional recommendation algorithms are applied and the input data or the results are filtered by the additional information. Another approach is to build multidimensional models, also referred as contextual models, where the additional information is directly integrated into the prediction model [1]. Recently, especially tensor decomposition approaches gained attraction, enabling the direct modeling of multi-modal information. In previous research, the focus was especially on the Tucker decomposition [19], the Parallel Factor Analysis (PARAFAC) [6], and Pairwise Interaction Tensor Factorization (PITIF) [16], which demonstrated high prediction qualities. However, the main drawback of these methods is that their adaption to non-categorical factors is difficult and error-prone [14].

As an alternative, Factorization Machines (FM) were introduced which combine the advantages of the tensor decomposition approaches to make predictions in highly sparse and multi-modal conditions, with the ability of support-vector-machines (SVM) to be a general predictor [13]. With FM, the user, rating, and additional information can be modeled as feature vectors, having categorical or continuous values. The target variable y can be a real-valued rating or a binary value [13]. FM have been demonstrated to be an effective approach to integrate contextual information, like mood, into movie recommender systems [15]. The key aspect of FM is, that the interactions between the input variables are not calculated directly, but a low-rank approximation is used. Within this paper, the FM model shown in Equation 4 is considered, which models binary interactions between the low-rank approximations $V = \langle v_i, v_j \rangle \in \mathbb{R}^{n \times k}$. The variable $k \in \mathbb{N}_0^+$ represents the number of latent variables. Appropriate values of k have to be determined empirically. On the one side, the value should be large enough so relevant interactions in the data can be captured. On the other side, restricting the value of k, and therefore its expressiveness might lead to better generalization of the model. [13]

$$\hat{y}(x) := w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j \tag{4}$$

In this paper, the reference implementation of FM within the LibFM library is used [14]. For learning models based on FM, optimization models based on Stochastic Gradient Descent (SGD), Alternative Least Squares (ALS), and Markov-Chain Monte Carlo (MCMC) are proposed, and implemented in LibFM. We decided to use the MCMC optimization, as this approach has the less hyperparameter which have to be tuned.

4 RESULTS AND DISCUSSION

The resulting F-measure values are illustrated in Figure 5. The models were built having latent variables values of $k \in \{16, 32, 64, 128\}$. As mentioned in subsection 3.1, three new user scenarios are investigated, in which no, two, or five ratings of the user are known. In the first scenario, the models having the measurement information of the user have a considerable better performance than the models without this information. In this scenario, the best model without information has a F-value of 0.40, whereas the best model with F-value a value of 0.49. This clearly better performance shrinks in case of the second scenario, where two items are known and vanishes in the last scenario, where five items are known. The precision of the models having body measurement information is in all cases higher, but at the same time, the recall is lower compared to

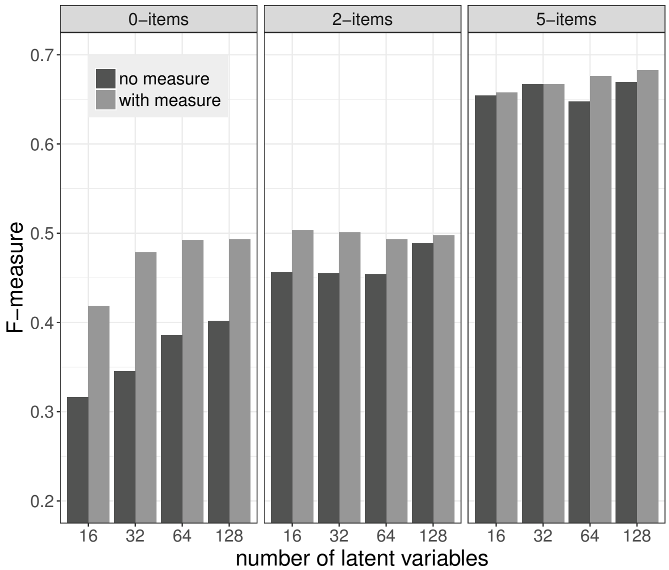

Figure 5: Resulting predictive performance per new user scenario.

the model having no measure information. In total, the empirical results indicate, that body measures possess significant predictive power in the context of apparel recommendation, especially in new users scenarios, where no previous user product preference information are available. In practice, one situation could be when consumers are getting scanned in a store the first time.

Further research is needed to identify which specific body measures have predictive power and which rather introduce noise to the model. From the eleven measures used, it can be expected, that some measures like the hip or belly measure give more information to the model than for example the calves measure. Another possibility to integrate body scan information is to assign each scan to a distinctive body shape class [10], or use the principal components from the morphable model approach [22]. In addition, the impact of the hair and skin color nuances on the predictive performance will be investigated in further research.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2008. Context-aware Recommender Systems. In *Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys '08)*. ACM, New York, NY, USA, 335–336. DOI: http://dx.doi.org/10.1145/1454008.1454068

[2] George Forman and Martin Scholz. 2010. Apples-to-apples in Cross-validation Studies: Pitfalls in Classifier Performance Measurement. *SIGKDD Explor. Newsl.* 12, 1 (2010), 49–57. DOI: http://dx.doi.org/10.1145/1882471.1882479

[3] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. 1992. Using Collaborative Filtering to Weave an Information Tapestry. *Commun. ACM* 35, 12 (1992), 61–70. DOI: http://dx.doi.org/10.1145/138859.138867

[4] Congying Guan, Shengfeng Qin, Wessie Ling, and Guofu Ding. 2016. Apparel recommendation system evolution: an empirical review. *International Journal of Clothing Science and Technology* 28, 6 (2016), 854–879. DOI: http://dx.doi.org/10.1108/IJCST-09-2015-0100

[5] Peng Guan, Alexander Weiss, Alexandru O. Balan, and Michael J. Black. 2009. Estimating human shape and pose from a single image. In *Computer Vision, 2009 IEEE 12th International Conference on*. 1381–1388.

[6] Richard Harshman. 1970. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics* 16 (1970).

[7] Sheena S. Iyengar and Mark R. Lepper. 2000. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology* 79, 6 (2000), 995.

[8] Luo Jiang, Juyong Zhang, Bailin Deng, Hao Li, and Ligang Liu. 2017. 3D Face Reconstruction with Geometry Details from a Single Image. (2017). http://arxiv.org/pdf/1702.05619

[9] Michelle R. Jones and Valerie L. Giddings. 2010. Tall women's satisfaction with the fit and style of tall women's clothing. *Jnl of Fashion Mrkting and Mgt* 14, 1 (2010), 58–71. DOI: http://dx.doi.org/10.1108/13612021011025438

[10] Jeong Yim Lee, Cynthia L. Istook, Yun Ja Nam, and Sun Mi Park. 2007. Comparison of body shape between USA and Korean women. *Int Jnl of Clothing Sci & Tech* 19, 5 (2007), 374–391. DOI: http://dx.doi.org/10.1108/09556220710819555

[11] Komal Nagar and Payal Gandotra. 2016. Exploring Choice Overload, Internet Shopping Anxiety, Variety Seeking and Online Shopping Adoption Relationship: Evidence from Online Fashion Stores. *Global Business Review* 17, 4 (2016), 851–869. DOI: http://dx.doi.org/10.1177/0972150916645682

[12] Jaekyung Park, Yunja Nam, Kueng-mi Choi, Yuri Lee, and Kyu-Hye Lee. 2009. Apparel consumers' body type and their shopping characteristics. *Jnl of Fashion Mrkting and Mgt* 13, 3 (2009), 372–393. DOI: http://dx.doi.org/10.1108/13612020910974500

[13] Steffen Rendle. 2010. Factorization Machines. In *Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10)*. IEEE Computer Society, Washington, DC, USA, 995–1000. DOI: http://dx.doi.org/10.1109/ICDM.2010.127

[14] Steffen Rendle. 2012. Factorization Machines with libFM. *ACM Trans. Intell. Syst. Technol.* 3, 3 (2012), 57:1–57:22.

[15] Stefen Rendle, Zeno Gantner, Christoph Freudenthaler, and Schmidt-Thieme Lars. 2011. *Fast Context-aware Recommendations with Factorization Machines: 34th International ACM SIGIR Conference on Research and Development in Information Retrieval ; July 24 - 28, 2011, Beijing, China.* ACM, New York, NY.

[16] Steffen Rendle and Lars Schmidt-Thieme. 2010. Pairwise Interaction Tensor Factorization for Personalized Tag Recommendation. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM '10)*. ACM, New York, NY, USA, 81–90. DOI: http://dx.doi.org/10.1145/1718487.1718498

[17] Barry Schwartz. 2016. *The paradox of choice: Why more is less* (revised edition ed.).

[18] Yue Shi, Martha Larson, and Alan Hanjalic. 2014. Collaborative Filtering Beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges. *ACM Comput. Surv.* 47, 1 (2014), 3:1–3:45. DOI: http://dx.doi.org/10.1145/2556270

[19] Ledyard R. Tucker. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31, 3 (1966), 279–311. DOI: http://dx.doi.org/10.1007/BF02289464

[20] Christian Zagel and Jochen Süßmuth. 2013. Nutzenpotenziale maßgetreuer 3D Avatare aus Low-cost Bodyscannern. *HMD : Praxis der Wirtschaftsinformatik* 50 (2013), 48–57. DOI: http://dx.doi.org/10.1007/BF03342068

[21] Christian Zagel, Jochen Süßmuth, and Freimut Bodendorf. 2013. Automatische Rekonstruktion eines 3D Körpermodells aus Kinect Sensordaten. In *Wirtschaftsinformatik Proceedings 2013*. Paper 35. http://aisel.aisnet.org/wi2013/35

[22] Michael Zollhöfer, Michael Martinek, Günther Greiner, Marc Stamminger, and Jochen Süßmuth. 2011. Automatic Reconstruction of Personalized Avatars from 3D Face Scans. *Computer Animation and Virtual Worlds (Proceedings of CASA 2011)* 22, 2-3 (2011), 195–202.