Identifying Exceptional Descriptions of People using Topic Modeling and Subgroup Discovery

Andrew T. Hendrickson, Jason Wang, and Martin Atzmueller

Tilburg University, 5037AB, the Netherlands
{a.hendrickson, y.w.wang, m.atzmuller}@uvt.nl

Abstract. Descriptions of images form the backbone for many intelligent systems, assuming descriptions that randomly vary in construction and content, but where description content is homogeneous. This assumption becomes problematic being extended to descriptions of images of *people* [14], where people are known to show systematic biases in how they process others [19]. Therefore, this paper presents a novel approach for discovering exceptional subgroups of descriptions in which the content of those descriptions reliably differs from the general set of descriptions. We develop a novel interestingness measure for subgroup discovery appropriate for probability distributions across semantic representations. The proposed method is applied to a web-based experiment in which 500 raters describe images of 200 people. Our analysis identifies multiple exceptional subgroups and the attributes of the respective raters and images. We further discuss implications for intelligent systems.

1 Introduction

The fields of machine learning and computational linguistics are increasingly focused on building multi-modal intelligent systems that integrate visual and text-based information, e. g., answering questions and generating textual descriptions of images [2, 9, 16], identifying objects in images based on text descriptions [20], or improving the sentence parsing of descriptions by grounding them with visual information [10]. While the data rely on text descriptions of images written by people, people generate idiosyncratic descriptions that differ significantly based on the goals, biases, and expertise of the person writing the description [14]. Detecting and quantifying these idiosyncratic descriptions is necessary for systems to optimally select, weigh, or filter descriptions. In this paper we present a novel approach for identifying homogeneous subgroups of descriptions that reliably differ in their content from the general set of descriptions. Specifically, we extract a low-dimensionality representation of individual descriptions based on latent Dirichlet allocation (LDA) [8]. Using that representation, we apply *exceptional model mining* [3, 11, 18], a variant of subgroup discovery [3] that focuses on complex target properties, for detecting homogeneous subgroups of descriptions in the low-dimensional space. We define a novel quality function based on a subgroups' topic distribution and use it to identify exceptionally unique homogeneous subgroups of textual descriptions. We believe this is the first demonstration of subgroup discovery based on such textual description data.

Preprint of: Andrew T. Hendrickson, Jason Wang, and Martin Atzmueller. Identifying Exceptional Descriptions of People using Topic Modeling and Subgroup Discovery.
In: Foundations of Intelligent Systems - Proc. ISMIS 2018. Springer, 2018

The efficacy of the proposed approach is validated on a new dataset of descriptions of people. We present results of applying the LDA-based exceptional model mining method on that dataset and discuss the implications for intelligent systems based on descriptions generated by people in general, and descriptions of people in particular. The contribution of the paper is summarized as follows:

1. We present a novel approach for mining exceptional subgroups in descriptions of people using subgroup discovery on topic models using LDA.
2. We introduce a new interestingness measure for subgroups that compares the distribution across topics in subgroups to the overall (expected) distribution.
3. We present and discuss the results of applying the proposed novel methodology to a real-world dataset of descriptions of people collected online.

2 Method

The proposed approach consists of three phases. First, textual data is transformed into a low dimensional space using latent Dirichelet allocation. Second, a novel interestingness measure for subgroup discovery is used to define and search for exceptional subgroups. Finally, the resulting subgroups are evaluated with a human-in-the-loop, in order to facilitate their interpretation and validation.

2.1 Topic Modeling

The latent Dirichlet allocation model (LDA) [8] is the most popular method of topic modeling in natural language processing. It is a statistical model of how text documents are generated that relies on the assumption that a written text can be represented as a collection of topics where each topic consists of a probability distribution across all possible words. Formally, the generative model for the j_{th} word w_{ij} (and its topic z_{ij}) in document i, given a distribution of topics $\boldsymbol{\theta}_i$ in document i and distribution of words $\boldsymbol{\varphi}_k$ in topic k, is:

(1) $\boldsymbol{\theta}_i \sim Dirichlet(\boldsymbol{\alpha})$, (2) $\boldsymbol{\varphi}_k \sim Dirichlet(\boldsymbol{\beta})$, (3) $z_{ij} \sim Multinomial(\boldsymbol{\theta}_i)$, (4) $w_{ij} \sim Multinomial(\boldsymbol{\varphi}_{z_{i,j}})$, where the number of topics k and the vectors of identical values $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are hyperparameters of the model.

2.2 Subgroup Discovery

Formally, a *database* $D = (I, A)$ is given by a set of individuals I and a set of attributes A. For nominal attributes, a *selector* or *basic pattern* $(a_i = v_j)$ is a Boolean function $I \rightarrow \{0, 1\}$ that is true if the value of attribute $a_i \in A$ is equal to v_j for the respective individual. The set of all basic patterns is denoted by Σ. A subgroup is described using a description language, typically consisting of attribute–value pairs. Here, we focus on an exemplary conjunctive pattern description language. A *subgroup description* or (complex) *pattern* P is then given by a set of basic patterns $P = \{sel_1, \ldots, sel_l\}, sel_i \in \Sigma, i = 1, \ldots, l$, which is interpreted as a conjunction, i. e., $P(I) = sel_1 \wedge \ldots \wedge sel_l$, with $length(P) = l$.

A *subgroup* $S_P := ext(P) := \{i \in I | P(i) = true\}$, i. e., a *pattern cover* is the set of all individuals that are covered by the subgroup description P. The set of all possible subgroup description is then given by 2^Σ. The pattern $P = \emptyset$ covers all instances contained in the database. A *quality function* $q\colon 2^\Sigma \to \mathbb{R}$ maps every pattern to a real number reflecting its interestingness. In contrast to related approaches like methods for mining association rules [1] or algorithms from the field of formal concept analysis [13], subgroup discovery (and in particular exceptional model mining) allow the specification and efficient application of complex quality functions for estimating the interestingness of a pattern, e. g., [3,5,17,18].

In the case of topic models, we utilize Dirichlet distributions capturing the overall distribution of topics in the overall dataset (i. e., modeling the expected distribution) while the topic distribution contained in each subgroup is modeled by another Dirichlet distribution. In order to obtain the respective Dirichlet distributions $Dir(\alpha_\emptyset)$ and $Dir(\alpha_{S_P})$ for the overall population $S_\emptyset$ and the subgroup S_P, respectively, we can compute a maximum likelihood estimate (MLE) utilizing the Newton-Raphson method [21,22] for obtaining the parameter vectors α_{S_P} and $\alpha_\emptyset$. For comparing distributions, we utilize the Kullback-Leibler divergence metric *KL*. Thus, for Dirichlet distributions, comparing $Dir(\alpha)$ and $Dir(\beta)$, we obtain

$$KL(\alpha, \beta) = \log \frac{\Gamma(\alpha_0)}{\Gamma(\beta_0)} - \sum_{i=1}^{k} \left(\log \frac{\Gamma(\alpha_i)}{\Gamma(\beta_i)} - (\alpha_i - \beta_i)(\psi(\alpha_i) - \psi(\alpha_0)) \right) , \quad (1)$$

here Γ is the gamma and ψ the digamma function, $\alpha_0 = \sum_{i=1}^{k} \alpha_i, \beta_0 = \sum_{i=1}^{k} \beta_i$. Our novel quality function for comparing topic distributions for a specific subgroup S_P is then given by

$$q_D(P) = KL(\alpha_{S_P}, \alpha_\emptyset) , \quad (2)$$

with the distribution parameters α_{S_P} and $\alpha_\emptyset$ for subgroup/overall population.

3 Experiment

In this section we detail the critical aspects of an online experiment to collect descriptions and judgments about images of people. In the following sections we describe the set of images (i. e., the stimuli) and outline the experimental procedure for obtaining the textual descriptions and ratings of the images.

3.1 Procedure

The images consisted of 193 color photographs of the head and shoulders of people standing in front of an off-white background who were instructed to look directly into the camera lens and maintain a neutral facial expression. These individuals were recruited from the University of Adelaide campus and compensated AUD$10 for their participation.

For the rating task, 500 participants were recruited via Amazon Mechanical Turk and paid US$2 for approximately 12 minutes of work. Participants were shown five randomly selected images and for each image they were asked to determine a number of attributes and write a physical and non-physical description of the face (minimum four words and 10 characters). In this analysis we focus on discovering exceptional descriptions of the non-physical characteristics. The attributes of each description include three self-report attributes about the specific rater (age, gender, and country) as well as four subjectively rated attributes of the person in the image (age, gender, eye color, hair color, typicality, and attractiveness). Unfortunately, the reported ethnicity was incorrectly coded and not recorded, resulting in seven attributes for each description. The experiment resulted in a dataset consisting of 2491 descriptions of 193 faces.

3.2 Discovering subgroups

The application of the text-based subgroup discovery consisted of multiple stages:

1. The number of topics k and the probability distribution across words for each topic φ_k was determined. This was done by searching for the set of hyperparameters (α, β, and k) that produce sparse topics (few topics on average per document) that differ across documents. Thus, the objective function simultaneously minimized the number of topics per description and maximized the difference between documents[1]. In this phase all descriptions of the same image were treated as a single document to increase the stability of the inference process, particularly determining the topic distributions (φ).
2. These topic distributions were used to construct a probability distribution across topics for each individual description. The search processes in the first step resulted in a solution with nine topics and thus each of the 2491 descriptions was represented as a probability distribution across those topics.
3. Finally, all possible subgroups defined by the seven attributes were evaluated to identify deviating subgroups of descriptions. This was done exhaustively using the SD-Map algorithm [6], provided by the VIKAMINE system [4][2].

We identified deviating subgroups using different values of n for identifying the top-n subgroups, while we discuss results for the top 20 subgroups below; other result sets were consistent. A *minimal improvement filter* [7] was applied to the set of all subgroups to limit the set of attributes defining exceptional subgroups. Specifically, a specialization P' of a pattern P is considered a more exceptional subgroup if P' improves on the quality function compared to P: So, e. g., we consider the specialization of the pattern *face_hair_color = black* to *face_hair_color = black AND face_gender = male*, if the quality of the latter pattern increases. A minimal subgroup size threshold of 1% was used in this analysis.

[1] The difference between documents was calculated as the sum across all pairs of descriptions of the cosine similarity of the topic probability distributions. The number of topics per document was calculated as the sum across all descriptions of the conditional entropy of the topic probability distribution.

[2] http://www.vikamine.org

Table 1. Exceptional subgroup attribute frequencies. The attributes and values of a description that are indicative of the top 20 exceptional subgroups. The count column indicates the number of subgroups that are distinguished by this attribute. R and I in the column headers indicate Rater and Image attributes, respectively.

R. Country	R. Gender	I. Gender	I. Eye Color	I. Hair Color	I. Ratings
USA (3) India (2)	Female (3) Male (7)	Female (8) Male (3)	Black (12) Brown (1) Green (1)	Black (6) Blond (1)	Typicality (4) Attract. (5)

4 Results and Discussion

Below, we first present an analysis of the types of attributes more likely to define deviating subgroups and their relationships. Next, we aggregate across exceptional subgroups and evaluate the frequency of specific images in exceptional subgroups and according raters. We also briefly summarize results with an alternative quality function, and conclude with a discussion.

4.1 The attributes of distinct subgroups

Interesting patterns emerge across the seven attributes that identify the 20 subgroups most dissimilar to the overall set of descriptions (Table 1). Though the gender of the rater and the face in the image were two of the most common attributes to define exceptional subgroups, only four subgroups were defined by both the gender of the rater and face. This suggests the interaction between rater and face gender was no more likely to identify an exceptional subgroup than either factor independently. Furthermore, the eye color of the image was a defining attribute of 14 of 20 exceptional subgroups and the attribute value "black" was by far the most frequently occurring value. This is particularly surprising given that eye color attribute was described as "black" for only 13% of descriptions, which was second frequent after "brown" and more frequent than "blue" and "green." One possible explanation of this pattern is that perceived eye color is highly correlated with perceived ethnicity or race for these descriptions, a possibility we discuss further in the discussion section.

4.2 Distributions of images and raters in distinct subgroups

Figure 1 shows the markedly different distribution of images (left panel) and raters (right panel) that occur in distinct subgroups. The majority of images occur in at least one subgroup, but only nine images (0.4.%) had at least 40% of the descriptions of them included in a subgroup and only one image had more than 50%. The majority of raters, in contrast, do not have a description in any deviating subgroup. This is natural given that raters only generate five descriptions but each image has, on average, 12.9 descriptions. Despite the high number of raters with zero descriptions in the exceptional subgroups, a small but significant group of raters (5.2%) had more than 50% of their descriptions be identified as belonging to at least one deviating subgroup.

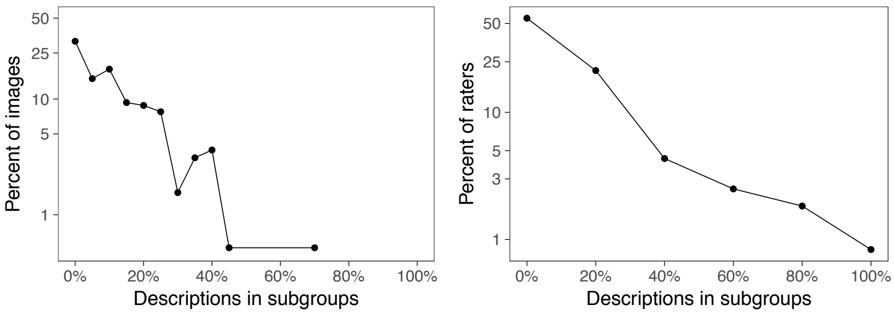

Fig. 1. The proportion of descriptions of specific images (left) and raters (right) that occur in at least one exceptional subgroup. The y-axis of both plots is in log units.

4.3 Comparing the Dirichlet and Hotelling quality functions

In all previous analyses, we utilized the proposed quality function $q_D(P)$. These results were also compared with the standard Hotelling quality function [3], for comparing multivariate means: The two quality functions were not a significantly correlated ($R = 0.12$). Furthermore, the Hotelling quality function did not produce as coherent subgroup attributes as the novel quality function $q_D(P)$. This divergence highlights the importance of using a quality function ($q_D(P)$) that directly corresponds to the multinomial probability distribution that comprises the probability distribution representation of a description across topics.

4.4 Discussion

Our results suggest that descriptions of people that significantly deviate from the population of descriptions are relatively frequent. Furthermore, these exceptional descriptions are not exclusively driven by particularly exceptional images or particularly exceptional raters. Instead, the vast majority of descriptions that are identified as exceptional are descriptions from raters for whom most descriptions are not exceptional, and of images whose descriptions are mostly not exceptional.

The attributes that define the maximally deviating subgroups point to the types of features of raters and images that are likely to produce exceptional descriptions. Male raters and female images are attributes that are likely to define deviating subgroups, though these two attributes appear to independently contribute and not in combination. Additionally, in this population of images, black hair and black eyes are the attributes of images most likely to identify exceptional subgroups. Further work is necessary to understand if these attributes produce exceptional descriptions when embedded in a different sample of images, or if these attributes are predictive of latent attributes, such as ethnicity, that were not included as attributes for the subgroup discovery process.

The issue of detecting heterogeneity of descriptions is particularly important for descriptions of people because people systematically differ in how they encode, search for, and remember faces of other races and genders [15, 19], which may systematically bias the descriptions people generate of others. However, these results do not show strong evidence of subgroups of descriptions that are identified based on gender, age, or race, perhaps decreasing the fear that intelligent systems based on descriptions of people will inherit strong implicit biases from the raters [12, 23]. We do find certain attributes of images, particularly black eye color and black hair color, which are much more likely to produce exceptional descriptions than other attributes. This suggests that the topics and words people use when describing the non-physical characteristics of other people may vary widely. The degree to which the importance of these attributes are an artifact of the particular faces we studied is an open question, but it highlights the importance of not ignoring the heterogeneity of textual descriptions generated by people. These issues are increasingly important as intelligent systems, trained with labels and descriptions generated by people, become ubiquitous. These systems rely on human annotated descriptions that are clearly not homogeneous. When combined with methods like LDA for extracting lower dimensional semantic representations, exceptional model mining and subgroup discovery techniques can provide a necessary tool to help identify potential biases in these descriptions. Additionally, these tools can possibly suggest specific images, subgroups, and attributes where additional data would help alleviate the bias in the systems that rely on them.

5 Conclusions

This paper presents a novel method of combining topic modeling and subgroup discovery to identify interesting image descriptions. We present a novel definition of interestingness that compares the subgroup and general population using the Kullback-Leibler divergence between the Dirichlet distributions that characterizes the probability distribution of topics. This method is applied to the problem of subgroup discovery among descriptions of pictures of people, a domain that has broad implications for applied domains [14] while carrying a real risk of biased descriptions [19]. Our analysis method detects meaningful subgroups of image descriptions that diverge from the general set of descriptions and characterizes them based on both properties of the raters as well as the images. These subgroups suggests new norms for data collection methods and statistical models for web-based applications that are sensitive to the heterogeneous nature of descriptions of people. For future work, we aim to extend the analysis and data collection in order to investigate (dis-)similarities in more datasets. Furthermore, the inclusion of contextual domain knowledge is an interesting issue to consider.

Acknowledgments

Funding for data collection was provided by a University of Adelaide Interdisciplinary Research Grant to C. Semmler, A. Hendrickson, R Heyer, A. Dick and A. van den Hengel. Furthermore, this work has been partially supported by the German Research Foundation (DFG) project "MODUS" under grant AT 88/4-1.

References

1. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. In: Proc. VLDB. pp. 487–499. Morgan Kaufmann (1994)
2. Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Lawrence Zitnick, C., Parikh, D.: Vqa: Visual Question Answering. In: Proc. IEEE ICCV. pp. 2425–2433 (2015)
3. Atzmueller, M.: Subgroup Discovery. WIREs DMKD **5**(1), 35–49 (2015)
4. Atzmueller, M., Lemmerich, F.: VIKAMINE - Open-Source Subgroup Discovery, Pattern Mining, and Analytics. In: Proc. ECML/PKDD (2012)
5. Atzmueller, M., Lemmerich, F.: Exploratory Pattern Mining on Social Media using Geo-References and Social Tagging Information. IJWS **2**(1/2), 80–112 (2013)
6. Atzmueller, M., Puppe, F.: SD-Map - A Fast Algorithm for Exhaustive Subgroup Discovery. In: Proc. PKDD. pp. 6–17. Springer, Heidelberg, Germany (2006)
7. Bayardo, R., Agrawal, R., Gunopulos, D.: Constraint-Based Rule Mining in Large, Dense Databases. Data Mining and Knowledge Discovery **4**, 217–240 (2000)
8. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. JMLR **3** (2003)
9. Borji, A., Cheng, M.M., Jiang, H., Li, J.: Salient Object Detection: A Benchmark. IEEE Transactions on Image Processing **24**(12), 5706–5722 (2015)
10. Chrupała, G., Gelderloos, L., Alishahi, A.: Representations of Language in a Model of Visually Grounded Speech Signal. In: Proc. ACL. pp. 613–622 (2017)
11. Duivesteijn, W., Feelders, A.J., Knobbe, A.: Exceptional Model Mining. Data Mining and Knowledge Discovery **30**(1), 47–98 (2016)
12. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R.S.: Fairness Through Awareness. CoRR **abs/1104.3913** (2011)
13. Ganter, B., Wille, R.: Formal Concept Analysis. Wissenschaftliche Zeitschrift-Technischen Universitat Dresden **45**, 8–13 (1996)
14. Gatt, A., Tanti, M., Muscat, A., Paggio, P., Farrugia, R., Borg, C., Camilleri, K., Rosner, M., van der Plas, L.: Face2Text: Collecting an Annotated Image Description Corpus for the Generation of Rich Face Descriptions. In: Proc. LREC (2018)
15. Herlitz, A., Lovén, J.: Sex Differences and the Own-Gender Bias in Face Recognition: A Meta-Analytic Review. Visual Cognition **21**(9-10), 1306–1336 (2013)
16. Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.J., Shamma, D.A., et al.: Visual Genome: Connecting Language and Vision using Crowdsourced Dense Image Annotations. International Journal of Computer Vision **123**(1), 32–73 (2017)
17. Lemmerich, F., Atzmueller, M., Puppe, F.: Fast Exhaustive Subgroup Discovery with Numerical Target Concepts. Data Mining and Knowledge Discovery **30**, 711–762 (2016). https://doi.org/10.1007/s10618-015-0436-8
18. Lemmerich, F., Becker, M., Atzmueller, M.: Generic Pattern Trees for Exhaustive Exceptional Model Mining. In: Proc. ECML/PKDD. Springer (2012)
19. Levin, D.T.: Race as a Visual Feature: Using Visual Search and Perceptual Discrimination Tasks to Understand Face Categories and the Cross-Race Recognition Deficit. Journal of Experimental Psychology: General **129**(4), 559–574hypo (2000)
20. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft Coco: Common Objects in Context. In: Proc. ECCV. pp. 740–755. Springer (2014)
21. Minka, T.: Estimating a DirichletDdistribution. Technical report, MIT (2000)
22. Sklar, M.: Fast MLE Computation for the Dirichlet Multinomial. arXiv:1405.0099
23. Torralba, A., Efros, A.A.: Unbiased Look at Dataset Bias. In: Proc. EEE Conference on Computer Vision and Pattern Recognition. pp. 1521–1528. IEEE (2011)