FindHer: a Filter to Find Women Experts

Gabriela Ferraro
CSIRO Data61, Canberra, Australia
Australian National University, Australia
gabriela.ferraro@data61.csiro.au

Zoe Piper
CSIRO, Canberra, Australia
zoe.piper@csiro.au

Rebecca Hinton
CSIRO, Canberra, Australia
rebecca.hinton@csiro.au

Abstract

Women are underrepresented in many spheres of our societies, including research. A common excuse for exclusively male line-ups is that suitable women could not be found. One way of promoting visibility of women in industry and academia is to explicitly provide solutions to find them. Expert Connect is a publicly searchable database of Australia's researchers that now includes FindHer, a filter to find women experts in any field of research. In this industry paper, we evaluate Natural Language Processing and Computer Vision technologies for gender determination within the aim of automating gender profile tagging for FindHer. We found current off-the-shelf tools are highly effective in detecting gender from names and photos. Nevertheless, a human-in-the-loop approach should be preferred to a fully automatic one, since ethical concerns might arise.

1 Introduction

Women representation remains critically low in a range of fields, including the research sector (Larivire et al., 2013; West et al., 2013; Mihaljevic-Brandt et al., 2016; Bonham and Stefan, 2017). A common excuse for exclusively male line-ups is that suitable women could not be found. This has led to a proliferation of initiatives encouraging women to list their details in various skills and expertise related directories. These directories are then promoted to conference organisers, company boards, investor groups, the media and beyond.

These directories do promote women and improve their chance of discovery; they also put the onus on women to create and update their profiles across multiple platforms. This can be time-consuming and repetitive. What is more, it may or may not result in extra opportunities since often the only people looking at these directories are people who already know they want to engage a woman. We can be smarter about how we manage the public data that already exists. We want to meet the challenge of someone saying that "a suitable woman could not be found". In short, we need to put gender-based information in places where people are already looking.

Expert Connect[1] is a publicly searchable database of Australian research expertise designed to boost industry-researcher collaboration. Since International Women's Day in 2019, Expert Connect can now be filtered to find women experts in any field of research using FindHer.[2] Currently, only 15 percent of the Expert Connect data is included in the filter (with over 4,500 women profiled). It's not a perfect process, but we are continually working on improvements with the number of women profiled continuing to grow.

Work like this raises ethical considerations. In Section 3.1, we discuss some of the issues we considered in the process of submitting our ethics approval for this work. We present experiments on automatic gender determination using off-the-shelf Natural Language Processing (NLP) and Computer Vision (CV) technologies. Our aim is to assess the feasibility of off-the-shelf technologies to support the classification of expert profiles according to gender. Our study shows that current technologies achieved high precision in gender classification using peoples names and profile's photos. However, automatically assigning the correct gender to people's profiles is far from perfect.

The rest of the paper is organised as follows: Section 2 includes a description of the methods we evaluate for gender determination. Section 3 is about the experiments carried-out and the results

[1] https://expertconnect.global/
[2] https://expertfindher.global/

found. In Section 4 we conclude this paper and present some ideas for future work.

2 Gender Determination Methods

2.1 Title lookup

Title lookup is a simple and effective deterministic method. Titles are usually one or more words prefixing peoples names such as Miss, Dr., President, among many others. Titles might signify gender, an official position, or a professional or academic qualification. It is not uncommon that people use titles in their public profiles, for example, in the Expert Connect platform. This is the most reliable source of gender since people assign themselves titles to with which they identified. The women titles we use are the following: *Mrs.*, *Ms.*, *Miss*, and *Sister*.

Note that other titles commonly found in expert profiles such as Dr., Professor President, are gender neutral and not useful for gender determination. This method works as follows: given a user profile, if the profile contained a title, then the gender associated with that title is assigned as the user's gender.

2.2 Name lookup

Name lookup is a widely used deterministic method that relies on directories or databases of female and male names. In our experiments, we use two lists of Australian scientist female names: *500 women scientists* and *Women in science Australia*, which contained 150 female names. Since Australia is a multicultural society the lists include diverse names, for example, **Shaghik**, **Jessica**, and **Samia**, just to mention a few. This method works as follows: given a user profile, the user name string is looked up in the names list, and if found, the gender associated with the name is assigned as the user's gender.

2.3 Genderize

Genderize is an existing third-party webservice for infering the gender of a first name. The service can be accessed through a free API and has a limit of 1000 queries per day. Genderize utilised big datasets of information from user profiles across major social networks across 79 countries and 89 languages. The response includes a confidence value and a count, which represents the number of data entries used to calculate the response. This method works as follows: given a user profile, the user name string (not including the surname) is sent to the API[3], and returns the probability estimate of its gender. The API also accept two optional parameters, *location-id* and *language-id*, which not used in our experiments.

2.4 Chicksexer

Chicksexer is a Python package designed to perform gender classification. It is based on a machine learning classifier that uses a character level multilayer LSTM network (Hochreiter and Schmidhuber, 1997). The model is trained using names with gender annotation from Dbpedia Person Data,[4] Popular baby names in the US,[5] and names datasets curated by Milos Bejda.[6] The output prediction includes the probability assigned to each gender class. This method works as follows: given a user profile, the user name string (not including the surname) is sent to the predict-gender function, which returns a probability estimate of its gender.

2.5 Facifier

Facifier is an emotion and gender detector developed in Python and OpenCV,[7] It is a machine learning based classifier that uses HaarCascade (Viola and Jones, 2001) to detect human faces in photos and a gender classifier. The gender classifier is trained with the KDEF[8] dataset consisting of 4900 images, and 2000 images from the IMDB dataset.[9]

2.6 CNN-Gender

CNN-Gender (Levi and Hassner, 2015) uses Convolutional Neuronal Networks for gender determination using images. We used a TensorFlow reimplementation.[10] The model is trained with the Adience dataset,[11] which contains 26,580 images.

	Female	Males
Names	568	263
Images	3934	3147

Table 1: Evaluation data for names and images

	Title lookup	**%**
Ms.	1	0.03
Mrs.	8	0.25
Miss	0	0
Sister	0	0
Total females	9	0.28
Mr.	19	0.60
Dr.	322	10
Prof.	66	2
	Name lookup	**%**
Females	64	11

Table 2: Title and Name lookup and experiment results. Title percentage is calculated over the total number of users in Expert Connect: 3157. Name percentage is calculated over the total number of females in the Names dataset: 568.

3 Experiments and Results

We use two datasets to evaluate the methods described in Section 2, Names and Images, respectively. Details about the datasets are shown in Table 1. The Names dataset was compiled by querying the Expert Connect database. Duplicate names and second names were removed, thus the dataset consists of single-term names. The Images dataset was compiled by querying the Expert Connect database and manually classifying user profile photos as female or male. Note that the datasets were only used for testing and never for training the methods.

Results for lookup methods are shown in Table 2. As expected, lookup methods have a poor coverage. Only 0.28 percent of the users indicate their gender using female titles. This is not surprising since female titles like Miss, Mrs. and Ms. indicate marital status and might be considered obsolete in modern societies.

For anecdotal purposes we also investigated the use of other titles. We found that only 0.6 percent of the users chose to set their title as Mr. (male title). Academic titles such as Dr. and Prof. seem to be preferred when choosing a title for professional profiles, showing that 10 percent of the users are identified as Drs. Nevertheless, only a small percentage of the profiles in Expert Connect include a title.

The Names lookup approach covers 11 percent of the total number of females in the Names dataset, demonstrating that catalogues are usually incomplete, and therefore, an unreliable source for finding female researchers.

	Precision	Recall	F1
Chicksexer	0.982	0.876	0.925
Genderize	0.925	0.901	0.912
Facifier	0.421	0.738	0.534
CNN gender	0.933	0.861	0.895

Table 3: Off-the-shelve methods for gender determination experiment results

Results for off-the-shelf methods are presented in Table 3. Both named-based methods achieved high precision and recall. Chicksexer performs slightly better than Genderize. Another advantage of Chicksexer is that is it trainable, and its source code is available under an open source license. To understand why named-based methods can make incorrect predictions, we investigated some false negative instances given by Chicksexer. We randomly chose 25 names, and found half of them are non-western, e.g., Vikneswary, Fincina Anumitra, Ya-juan, among others. Some names are unisex, e.g., Sasha, Ali; and some users used short versions of their names, which made them gender neutral or unisex, e.g., Cat, Steph, Nicky, Charlie, and Billie. Some false negatives are likely to be reduced by including culturally diverse examples for training. However, predictions for gender neutral names are likely to remain confused when using only names for gender determination, either for humans or machines.

The image-based methods show a considerable difference in performance between them. Facifier achieved modest results when applied to the im-

[3] https://pypi.org/project/Genderize/
[4] https://wiki.dbpedia.org/downloads-2016-10
[5] https://www.ssa.gov/oact/babynames/
[6] https://mbejda.github.io/
[7] https://opencv.org/
[8] http://kdef.se/
[9] https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/
[10] https://github.com/dpressel/rude-carnie
[11] http://www.openu.ac.il/home/hassner/Adience/data.html

ages from the Expert Connect database, with 0.421 Precision, 0.738 Recall, and an F1-score of 0.534. An error analysis on its output indicate that Facifier sometimes struggles to capture the face in images, and therefore is not able to predict the gender class. Note also that the datasets used to train Facifier are considerably smaller than the ones used in CNN-gender, hence a fair comparison between Facifier and CNN-gender is not possible. CNN-gender achieved high Precision: 0.933, high Recall: 0.861, and an F1-score of 0.895.

To better understand CNN-gender errors, we examined some false positive and false negative instances. We randomly select 25 false positive instances, and could not find a clear pattern among them. In 4 images males have long hair, in 4 images they are using glasses, and in 3 images the person appears small at a corner of the image. Similarly, we randomly select 25 false negative instances. This time we found clear patterns between them. In 20 images females are using glasses. In in all of them females have short hair or a pony tail, which make them look as if they have short hair. Although these traits are not strictly male ones, the datasets are probably biased, as there are many training instances of females with long hair, and males with short hair. For human judges it was very easy to determine the gender of the false positive and false negative instances, however the automatic classifier struggled to correctly predict them.

3.1 Ethical Concerns and Limitations

The study presented in this paper and the FindHer filter has ethical approval. The ExpertConnect Platform clearly let people know that they have a profile irrespective to gender. To ensure transparency, how the FindHer filter works is available to the public.

The authors are aware that names and images cannot be used to unambiguously determine the gender of a user, and that a user might not identify with the prototypical gender they look like, nor with their given name. All the methods studied in this paper see gender in a binary way: a name or an image can be either female or male. This poses a clear limitation since in modern societies gender is seen as a spectrum, rather than in a binary way.

As shown in this study, automatic gender binary determination methods are far from perfect and it is still in beta. Ethical considerations might arise if the wrong gender is assigned to a user. Therefore, the FindHer team use automatic methods for classifying profiles according to their gender and confidence scores, which are later manually assessed before updating the gender of users' profiles.

4 Conclusion and Future Work

Gender inequality persist around the world. Considerable effort and resources are currently invested to mitigate this issue and to promote gender equality. FindHer is an example of such efforts, as it allows anyone to explicitly find woman experts in Australia via a Web platform. In order to automate gender determination of expert profiles, we have studied the performance of off-the-shelf language and computer vision technologies, which use given names and profile photos, respectively. Our experiments show the assessed methods are successful, and performance is likely to be higher if name-based and image-based methods are combined. There are many other sources of names such as name repositories, administration records and country specific birth list. Performance can also be improved by re-training name-based methods using culturally diverse sets of names, so that the tools will reflect the cultural diversity of Australian society.

5 Acknowledgements

The authors would like to thank the ALTW reviewers for their thoughtful feedback. Special thanks to Stephen Wan for providing helpful feedback and for proofreading this paper; and to Trevor Fairhurst for his valuable contributions at the early stages of this project.

References

Kevin Bonham and Melanie I. Stefan. 2017. Women are underrepresented in computational biology: An analysis of the scholarly literature in biology, computer science and computational biology. *PLOS Computational Biology*, 13:e1005134.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

Vincent Larivire, Chaoqun Ni, Yves Gingras, Blaise Cronin, and Cassidy Sugimoto. 2013. Bibliometrics: Global gender disparities in science. *Nature*, 504:211–3.

Gil Levi and Tal Hassner. 2015. Age and gender classification using convolutional neural networks. *2015*

IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 34–42.

Helena Mihaljevic-Brandt, Lucia Santamaria, and Marco Tullney. 2016. The effect of gender in the publication patterns in mathematics. *PLOS ONE*, 11:e0165367.

Paul Viola and Michael Jones. 2001. Rapid object detection using a boosted cascade of simple features. volume 1, pages I–511.

Jevin West, Jennifer Jacquet, Molly King, Shelley Correll, and Carl Bergstrom. 2013. The role of gender in scholarly authorship. *PloS one*, 8:e66212.