Diversity metrics and visualization; estimating AIs gender & ethnic biases

NamSor, 2019

This paper describes new metrics that we would like to add to NamSor API v2. These metrics will be built on top of Gender, Origin, Diaspora and US 'Race'/Ethnicity classification – but help analyse real life phenomenon such as : gender or ethnic discrimination; urban ethnic segregation; gender and ethnic biases in artificial intelligence models….

We propose to experiment how those metrics and visualizations tools perform in any real-life processes that resemble a 'funnel', such as : recruitment, credit allocation, medical trials, grant allocation, start-up financing etc.

Contents

Introduction

How many scientific papers have been published by female scientists compared to their male counterparts? This is the sort of question NamSor API helps answering.

Among the different methods that are commonly used for gender attribution, name inference presents several advantages: it can be applied retroactively to any database that contains personal names and it doesn't rely on any secondary source of data.

Table 1: Summary of methodological approaches to attribute gender

Method	Type	Advantages	Disadvantages	Examples
Primary data collection	Direct	+ Captures information directly at source + Self-declaration allows more gender diverse categories	– Cannot be applied retroactively – Implementation time – Difficult in multiple countries	Walsh & Nagaoka (2009)
Attribution based on secondary source on individuals' data	Indirect	+ Can be as reliable as primary data if based on unique identifiers. + Self-declaration in secondary source may also allow more gender diverse categories. + Can be applied retroactively if secondary source permits.	– Depends heavily on secondary source coverage. – May be difficult to collect secondary source in multiple countries and years.	Jung & Ejermo (2014)
Attribution based on name gender semantics	Indirect	+ Can be applied retroactively if language or customs permit. + Can be applied to countries sharing the same language conventions	– Depends heavily on quality and coverage of naming rules. – Difficult for languages without clear-cut rules. – Affected by migration and naming trends	Park & Yoon (2007), Tripathi & Faruqui (2011)
Attribution based on name-gender dictionary	Indirect	+ Can be applied retroactively + Can be applied to countries sharing the same naming conventions	– Depends heavily on the dictionary coverage. – Affected by migration and naming trends	Frietsch et al. (2009), Naldi & Parenti (2002a, 2002b), UKIPO (2016a, 2016b)

Figure 1 - Identifying the gender of PCT inventors (WIPO, 2016)

Inferring the gender of names has numerous uses in research, business, and public services. Among the different methods that are commonly used for gender attribution, name inference presents several advantages:

- It can be applied retroactively to any database that contains personal names;
- It doesn't rely on any secondary sources;
- It is fast, and;
- It is cost efficient.

NamSor's goal is to provide a reliable name gender prediction model with global coverage (for all countries, all languages, all alphabets, etc.).

To overcome the challenges of traditional methods (name gender semantics, name gender dictionaries) and specifically reduce the impact of migration trends on gender inference, NamSor can interpret the

full name (ex. ***Mary Smith***, ***John W. Smith, Smith; Mary,*** *王晓明*) or both first & last names (***Mary|Smith***; ***John|Smith,*** *晓明|王*).

NamSor also includes API endpoints to infer the origin or ethnicity of individuals based on their names, and these developments reinforce the quality of our gender estimation. A given name like ***Andrea*** may be male or female depending on the cultural context. The cultural context is inferred using the surname combined with the first name and there is also an optional geographic context (country ISO2 code). In some cultures, a surname ending may change depending on gender, so the API automatically recognizes if gender can be inferred from the first name (e.g., ***Carl***) or the last name (e.g., ***Sokolova***).

NamSor uses extensive machine learning techniques (supervised and non-supervised), and we collaborate also with linguists, geographers, anthropologists, and historians to increase our model's accuracy in various cultural contexts.

Our case for algorithmic transparency

NamSor has been used by many Universities and Research Organizations worldwide and we are periodically asked to be more transparent on the algorithms we use and our model's accuracy.

Since Artificial Intelligence has become hype, there have been several controversies about AI's gender and racial biases. It is, in fact, very difficult for an A.I. *not to learn* gender, ethnic, or racial biases.

NamSor is designed to maximize gender, ethnic, or racial biases for applications where such information is desired (data mining, gender equality studies, human geography, analysis of urban segregation, etc.), and we see a business opportunity for NamSor to assess other A.I.'s gender, ethnic, or racial biases for applications where such biases are neither desired nor allowed (human resources / recruitment, credit rating, …). We believe being transparent on our algorithms can be useful to build trust in our capability to assess other A.I.'s gender, ethnic, or racial biases.

Consequently, in 2019, we have launched a new version of NamSor API v2 with the objective of being more transparent on the algorithms.

As a company, we also separately maintain two distinct groups of algorithms:
- NamSor v1 Core, our historical software (2012-2018) which still has some interesting features for non-supervised machine learning (clustering name by ethnic / linguistic groups);
- NamSor ML, a deep learning framework for processing international names using word embedding (FastText or Word2Vec pre-trained models).

Naïve Bayes Algorithm for Name Classification

NamSor API v2 uses Naïve Bayes Classifiers, a class of algorithms which is excellent at classification. Currently, we provide classification to the following taxonomies :

Gender API

The Gender API

Infer the likely gender of a name, optionally given a local context (ISO2 country code). It is documented here,
https://v2.namsor.com/NamSorAPIv2/apidoc.html#/personal/genderGeo

Origin API

Infer the likely country of origin of a personal name. It assumes names as they are in the country of origin. It is documented here,
https://v2.namsor.com/NamSorAPIv2/apidoc.html#/personal/origin

Country API

Infer the likely country of residence of a personal full name, or one surname. Assumes names as they are in the country of residence OR the country of origin. It is documented here,
https://v2.namsor.com/NamSorAPIv2/apidoc.html#/personal/country

Diaspora API

Infer the likely ethnicity/diaspora of a personal name, given a country of residence ISO2 code (ex. US, CA, AU, NZ etc.) It is documented here,
https://v2.namsor.com/NamSorAPIv2/apidoc.html#/personal/diaspora

US 'Race'/Ethnicity API

Infer a US resident's likely race/ethnicity according to US Census taxonomy, using (optional) ZIP5 code info. Output is W_NL (white, non latino), HL (hispano latino), A (asian, non latino), B_NL (black, non latino). It is documented here,
https://v2.namsor.com/NamSorAPIv2/apidoc.html#/personal/usRaceEthnicity

What is diversity ? Diversity metrics and visualization

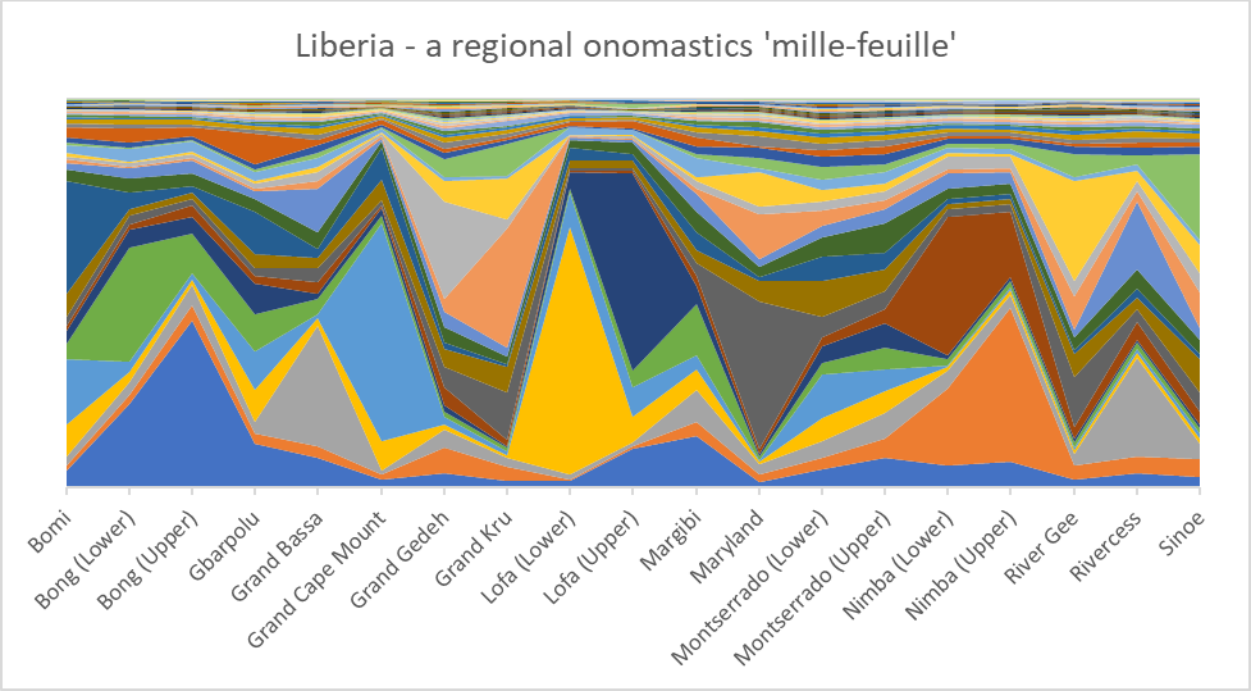
There are different definitions of diversity and many angles to look at it. For example, in the United-States, diversity is mostly analysed from the angle of gender and 'race'/ethnicity (White, Black, Asian, Hispano-Latino etc.) Those two taxonomies will not be so relevant for a country in Europe, Latin America, Asia, Africa. Each will have very specific and different angles.

At NamSor, we believe there is no single answer but our goal is to support all the relevant taxonomies.

We've worked with geographers, ethnographers, linguists and sociolinguists to try and understand what stories personal names say – and how they can be interpreted in a local context.

Example : diversity of names in Liberia

This is a real example of visualization of the diversity of personal names in Liberia, West-Africa. Each colour corresponds to a different type of names that can be recognized from non-supervised ML machine learning model (name clustering). The name clusters correspond to different ethnic groups, who live in different regions/states of Liberia.



This visualization is an artistic view of diversity. In a less artistic format, we could represent the data as a 'stacked bar chart' or even a set of pie charts with the population breakdown for each state.

How it can be used: once a ML machine learning model is trained, it can be applied on other contexts. For example, to analyse diversity in political representation, to create maps of urban segregation in the capital city …

Proposition 1 – evaluate diversity index based on entropy (Diversity API)

We propose to evaluate different diversity index based on entropy and create a formal, well documented, diversity index that will be built on top of NamSor taxonomies (such as : gender, origin, diaspora, country, US 'race'/ethnicity as well as any additional known contextual attributes).

This will be published as a new API that takes a diverse population as input, some contextual information and return the diversity index values according to a number of taxonomies and dimensions.

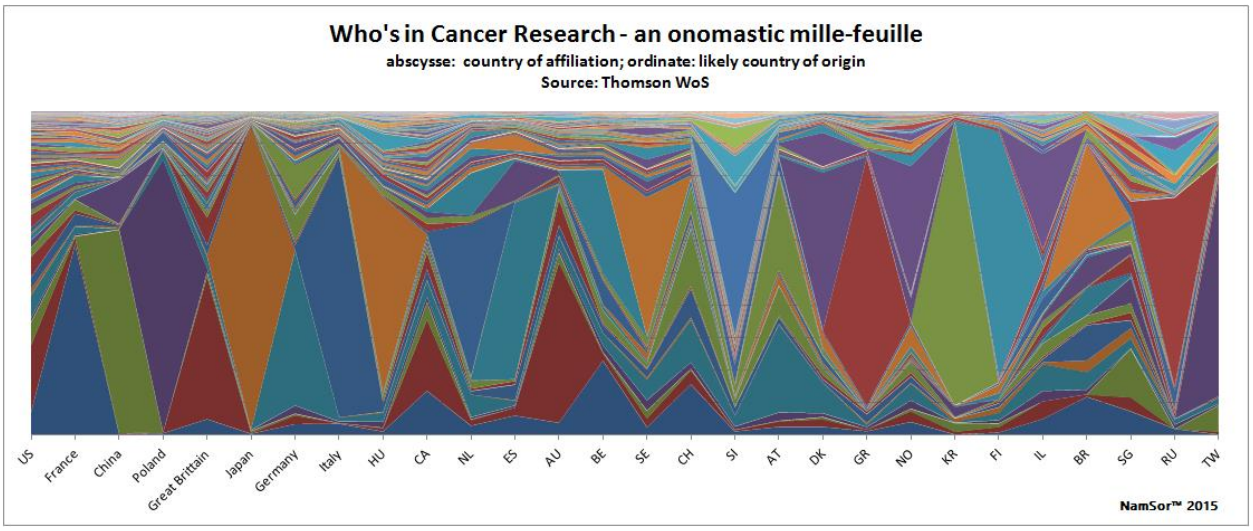
Applied to the context of collective intelligence, it will be possible to assess the diversity of a team; correlate diversity of a team and quality of a team outputs; estimates the level of segregation or discrimination in collaboration.

Proposition 2 – artistically visualize diversity as an 'mille-feuille' (MilleFeuille API)

We propose to make it easy to produce arty, colourful and meaningful data visualizations of diversity in a population.

Applied to the context of collective intelligence, it will be possible to visualize the diversity of a team; it will be possible also to visualize globally all teams and how they collectively address specific issues.

In May 2019, Nesta has hosted a workshop recently to explore the use of collective intelligence design and data to re-imagine the bowel cancer diagnosis journey. This is an example of visualizing diversity among Cancer Researchers produced in 2015 in collaboration with INSERM, based on scientific database Thomson Web Of Science :



Estimating gender, racial and ethnic biases in Artificial Intelligence models

Since Artificial Intelligence has become hype, there have been several controversies about AI's gender and racial biases. It is, in fact, very difficult for an A.I. *not to learn* gender, ethnic, or racial biases.

Proposition 3 – estimate gender biases in AI (GenderBias API)

We propose to create a formal method to estimate gender biases in any funnel-based process that involves an A.I. such as : credit allocation; recruitment or promotion in Human Resources; grant allocation; content curation etc.

Applied to the context of collective intelligence, it will make sure contributing women and men have equal chances of seeing their ideas curated, shared, discussed, equitably assessed.

Proposition 4 – estimate racial & ethnic biases in AI (RacialEthnicBias API)

We propose to create a formal method to estimate racial and ethnic biases in any funnel-based process that involves an A.I. such as : credit allocation; recruitment or promotion in Human Resources; grant allocation; content curation etc.

Applied to the context of collective intelligence, it will make sure contributing women and men of any 'race' or ethnicity have equal chances of seeing their ideas curated, shared, discussed, equitably assessed.

Team & Expert Group (proposed)

We consider the following team

- Elian, data scientist and founder NamSor.com
- Antoine, researcher based in Germany, specialized in quantifying discrimination;
- Anne-Laure, researcher based in the UK, specialized in building Gender Indexes;
- Karima, researcher based in France, specialized in women & minority representation in Corporate Governance;
- Aakaash, researcher based in Chicago, specialized in Diversity in Science.

plus a freelance statistician / mathematician.

Conclusion

NamSor, since 2012, has built tools and methods to analysing diversity from different angles, based on personal names. We have developed unique taxonomies (such as: gender, origin, diaspora, country, US 'race'/ethnicity).

As part of this project, we propose to improve the documentation for those taxonomies, as well as develop new aggregated metrics and visualization tools that will help

1) Estimate diversity with formal well documented metrics;
2) Visualize diversity in an arty, colourful and meaningful picture;
3) Estimate gender, ethnic or 'racial' biases in any real-world funnel-like process that involves an artificial intelligence (A.I.)

We propose to design aggregate metrics, beautiful data visualizations of 'diversity' that can work on global scale and accommodate the complexity of the World itself. Not push a single view of what 'diversity' means in a local context, such as the Silicon Valley, or the UK or France.

We will provide Application Programming Interfaces (APIs) that are simple to integrate. We will encourage our users to publish the resulting data as open data for other external researchers to use.