Use of non-traditional data sources to nowcast migration trends through Artificial Intelligence technologies.

Diletta Goglia[1*], Laura Pollacci[1], Alina Sîrbu[1]

[1]Department of Computer Science, University of Pisa, Pisa, Italy
[*]Corresponding author: d.goglia@studenti.unipi.it

Keywords: International migration · migration flows · migration stocks · bilateral migration · Facebook social connectedness index · migration nowcasting · migration prediction · migration drivers

EXTENDED ABSTRACT

In recent years the pursuit of original drivers and methods is becoming an increasing requirement to migration studies, considering the new technologies used to characterise and understand the human migration phenomenon.

Many researchers have proposed to employ non-traditional data sources to study migration trends, including so-called social Big Data such as online social networks. This unconventional approach is intended to find an alternative methodology to ultimately answer open questions about the human mobility framework (i.e. nowcasting flows and stocks, studying integration of multiple sources and knowledge, and investigating migration drivers). The new data have the advantage of timeliness and large geographical coverage, but also disadvantages in terms of selection bias and amount of resources required to process. Therefore, models extracted from these data need to be carefully validated, typically with traditional data sources. In this context of meaningful data combination, many types of data exist, still very scattered and heterogeneous, making integration far from straightforward.

Our work focuses on the use of alternative types of data and proposes two different models, a new multi-feature dataset and a new indicator that could significantly contribute to the study of migration and to forecast emerging trends through the use of Artificial Intelligence technologies.

We present a dataset to be exploited in migration studies as a concrete example of this new integration-oriented approach: the Multi-aspect Integrated Migration Indicators (MIMI) dataset. It includes both official data about bidirectional human migration (traditional country-to-country flow and stock data) with multidisciplinary variables and original indicators, including economic, demographic, cultural and geographic indicators, together with the Facebook Social Connectedness Index (SCI).

Thanks to this variety of knowledge, experts from several research fields (demographers, sociologists, economists) could exploit MIMI to investigate the trends in the various indicators, and the relationship among them. Moreover, it could be possible to develop complex models based on this dataset, able to assess human migration by evaluating related interdisciplinary drivers, as well as models able to nowcast and predict traditional migration indicators in accordance with non-traditional variables, such as the strength of social connectivity. Here, the SCI could have an important role. It guarantees an anonymized collection of information on users and their friendships, measuring the relative probability that two individuals across two countries are friends with each other on Facebook: therefore it could be employed as a proxy of social connections across borders, to be studied as a possible driver of migration.

The inclusion of this indicator in the dataset enables it to be exploited as a non-traditional way to describe, understand and nowcast international migration.

To this end, we built a new measure, the Bidirectional Migration Probability (BMP) index, which takes into account both the inflows and outflows shared by two countries, and measures the relative probability of a person to be a migrant from country i to j and vice versa. BMP indicator allows to portray and predict bilateral migration trends relying on the intensity of social networking, since it shows significant correlations with SCI.

The first model we present is an ordinary least squares statistical model (OLS) that performs a linear regression to nowcast migration trends. Specifically, the model fits a subset of variables, including the Facebook SCI, and evaluates their relevance for the purpose of nowcasting the BMP index.

Together with the migration drivers resulting from the fit of the OLS model, we will also include knowledge about past migration flows to build a second model, which consists in a Machine Learning Artificial Neural Network able to forecast migration and so to predict future cross-border human mobility trends.

All in all, the motivations of our contribution lie in the need of new perspectives, methods and analyses that can no longer prescind from taking into account a variety of new factors. The heterogeneous and multidimensional sets of data released with MIMI and exploited in the two models with the aid of the BMP indicator offer an all-encompassing overview of the characteristics of human migration, enabling a better understanding and an original potential exploration of the relationship between migration and non-traditional sources of data.