Extended Abstract: Comparing Artificial Neural Network and Cohort-Component Models for Population Forecasts

Viktoria Riiman* Reed Milewicz◇ Amalee Wilson◇ Peter Pirkelbauer◇

* Center for Business and Economic Research, University of Alabama.
◇ Computer and Information Sciences Department, University of Alabama at Birmingham.

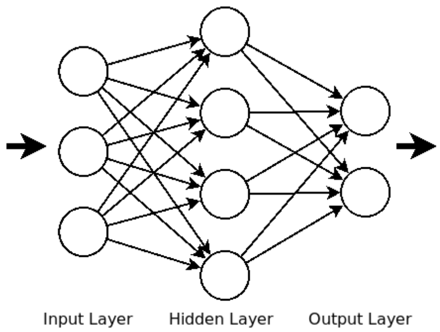
Abstract

Artificial neural network (ANN) models are rarely used to forecast population in spite of their growing prominence in other fields. We compare the forecasts generated by ANN models with population projections from traditional cohort-component method for counties in Alabama. We use forecasts for all 67 counties that offer diversity in terms of population and socioeconomic characteristics. When comparing predicted values with total population counts from the 2010 decennial census, the cohort-component method used by the Center for Business and Economic Research at the University of Alabama in 2001 produced more accurate results than a basic ANN model. Only when we proxy a forecaster's experience and personal judgment with potential economic forecasts, preliminary results from ANN models improve. The results indicate the significance of forecaster's experience and judgment for cohort-component model and difficulty, but not impossibility of substituting these insights with available data.

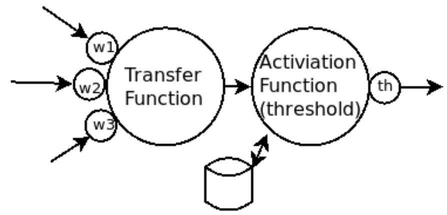
Introduction

Although artificial neural networks (ANN) dominate forecasting in numerous domains, not many attempts at using them for population projections have been made so far (Tang et al. (2006); Folorunso et al. (2010); Nordbotten (1996)). This is despite the fact that the potential for using ANN models in projecting population was noted more than a decade ago (Smith et al. (2001)). Prior use of ANN models have utilized feed-forward networks with back propagation (Folorunso et al. (2010)) or fuzzy networks (Tang et al. (2006)). These models were shown to perform better than ratio correlation regression model for projecting population (Tang et al. (2006)) and better than forecasts that plug projected fertility, mortality, and migration data into a cohort-component model (Folorunso et al. (2010)). Specifically, Tang et al. use data on birth, death, and school enrollment, while Folorunso et al. used fertility, mortality, and migration data produced by the National Population Commission, entering the data into a cohort-component model equation.

In this paper, we base our prediction on a long short term memory (LSTM) network. LSTM models become increasingly popular due to their benefits with back propagation and gradient descent methods. We compare prediction capabilities from our ANN model with the population projections developed at the Center for Business and Economic Research (CBER) at the University of Alabama in 2001. CBER



(a) A neural network



(b) A neuron in a LTSM network

Figure 1: Overview of Neural Network

forecaster projected population using cohort-component method. We assess the accuracy of both methods by comparing them to actual population counts from the 2010 decennial census.

To attain a high-quality cohort-component model, forecasters refine the basic methods based on their experience. ANN models do not have that capability yet, hence we experiment by proxying the lack of cognitive ability with actual economic data. After experimenting with different types of models and training methods, preliminary results show that this model predicts population more accurately.

Artificial Neural Network Model

Artificial neural networks (ANN) is a machine learning approach that attempts to simulate cognitive functions. As depicted by Fig. 1, a simple ANN consists of layers of neurons. Each neuron receives input from neurons of the preceding layer and outputs to the neurons of the following layer. A neuron's transfer function computes a weighted combination of its input connection, and fires an output if it exceeds a certain threshold. Before we can utilize an ANN, we need to train it. We used supervised learning where we provide input data and expected output data. During the training phase, the ANN adjusts its neuron's weights and thresholds to produce the desired output.

Several variations of neural networks have been proposed. We utilize a long short term memory network (LSTM) (Hochreiter and Schmidhuber (1997); Gers et al. (2001)). LSTM enhances ANN models developed earlier with the addition of long short term memory. Thus, the transfer function can take into account the past history of its outputs. LSTM's long term memory component helps avoid quick vanishing of gradient errors that has been observed with other neural networks. We implemented our model using the Chainer neural network framework[1].

Cohort-Component Method

The cohort-component method is a traditional forecasting approach in demography (Smith et al. (2013)). It is also the most popular method among the members of the U.S. Census Bureau Federal-State Cooperative for Population Projections (FSCPP). According to the 2015 FSCPP survey, 75% of FSCPP members use cohort-component method based on historical demographic data (Hunsinger (2015)). The next two popular methods were used only by 27.5% and 22.5% of respondents, respectively: trend extrapolation of total population data and top-down methods such as constant-share, shift-share, and share of growth.

The Center for Business and Economic Research at the University of Alabama, established in 1930, has

[1] `http://chainer.org/`

a long history of developing population projections using cohort-component method. The Center forecasters use bridged-race population estimates from the Centers for Disease Control and Prevention (CDC), specified in five-year age groups from 0-4 through 80-84 (CBER (2001)). For computational purposes, the 0-4 age group is split into under 1 and 1-4 components, while individuals 85 and over are grouped in a single category. Breaking each age group down by race and gender yields 76 age/race/gender cohorts (using two race groups: white population and black and others population).

The basic equation of the projections is:

$$P_t = P_{t-1} + B_{t-1,t} - D_{t-1,t} + M_{t-1,t}$$

where, P_t refers to population at time t; P_{t-1} population at time $t-1$; $B_{t-1,t}$ the number of births in the interval from time $t-1$ to time t; $D_{t-1,t}$ the deaths in the interval from time $t-1$ to time t; and $M_{t-1,t}$ is the net migration in the interval from time $t-1$ to time t, which equals in-migration minus out-migration.

The projection process is carried out in five-year increments, run independently for each geography. The calculation of birth, death, and migration components forms the basis for these projections. Thus, forecaster's experience and personal judgment in making assumptions for these components are important for the accuracy of projections.

Data

For ANN models we use data available by county: mid-year intercensal population estimates developed by the U.S. Census Bureau from the Bureau of Economic Analysis (BEA) and economic data from BEA such as proprietors' employment, wage and salary employment, real per capita income, and real average earnings. We also use births and deaths data from the Alabama Center for Health Statistics and net migration data that are estimated by subtracting births and deaths from total population. Population data are available from 1969 and births and deaths data from 1990, although we are in the process of obtaining earlier data for latter variables. In addition, we use dummies for economic development from Alabama Workforce Development Councils that divide the state territory into 10 geographically compact regions.

We also have more detailed data by county from decennial Censuses going back to 1820, total population and death data by race, age, and gender, birth data by race, and in- and out-migration data from IRS. We are planning to use these data to examine their effect on the accuracy of ANN models.

Since projections using cohort-component method for 2010 were based on data up to 2000, in order to have equivalent projections for comparison, we start by using the data up to 2000 for ANN models as well. We then add economic data to substitute for forecaster's knowledge on upcoming events affecting population such as economic and housing developments.

Expected Results

Since previous papers comparing ANN with other methods used for population projections showed the results favorable for ANN, we expected a similar outcome when comparing those with projections from utilizing cohort-component method in Alabama. However, when running preliminary ANN models we found that cohort-component method yield more accurate results displayed in Table 1. Thus, forecaster's experience and personal judgment are very important for the accuracy of results. When forecasting population in 2001, she took into account planned economic developments, potential formation of new school districts, changes to prison populations, possible army personnel movements, and all other potentially useful information to make very accurate forecasts.

With the continuing development of ANN models, we are expecting to receive improved forecasts, but we need to find ways to substitute for information available to experts developing population projections.

Using more data, such as economic forecasts, could be one such option. Overall, using ANN models for projecting only some components such as migration instead of total population, could be worthwhile to explore in future. This could make ANN models another alternative tool in the toolbox of experienced demographer.

Table 1: Estimated Errors from Comparing Projections with 2010 Census

Method	RMSE	MAE	MAPE
Cohort-Component Method	5,251	3,216	6.5%
ANN-LSTM (basic)	13,544	5,837	10.6%
ANN-LSTM + EF (w/ economic forecasts)	7,643	3,538	6.1%

Notes on preliminary results: The ANN-LSTM + EF model shows a higher RMSE and lower MAPE when compared to the cohort-component model. The reason for the difference is that the ANN-LSTM + EF model produces outliers for large counties. The refinement of the model to address this issue is work in progress.

References

CBER, 2001. Alabama Population Projections 2010-2040. Methodology. Center for Business and Economic Research, The University of Alabama, Culverhouse College of Commerce.

Folorunso, O., Akinwale, A., Asiribo, O., Adeyemo, T., 2010. Population prediction using artificial neural network. African Journal of Mathematics and Computer Science Research 3(8), 155–162.

Gers, F., Eck, D., Schmidhuber, J., 2001. Applying lstm to time series predictable through time-window approaches. In: Proceedings of the International Conference on Artificial Neural Networks, ICANN ’01, pp. 669–676, Springer-Verlag, London, UK, UK.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9(8), 1735–1780.

Hunsinger, E., 2015. FSCPP Members 2015 Questionnaire. Alaska Department of Labor and Workforce Development, State Demographer’s Office.

Nordbotten, S., 1996. Neural network imputation applied to the Norwegian 1990 population census data. Journal of Official Statistics 12(4), 385–401.

Smith, S. K., Tayman, J., Swanson, D., 2001. State and Local Population Projections. Methodology and Analysis. Kluwer Academic/Plenum Publishers, NY.

Smith, S. K., Tayman, J., Swanson, D., 2013. A Practitioner’s Guide to State and Local Population Projections. Springer Media, Dordrecht.

Tang, Z., Leung, C. W., Bagchi, K., 2006. Improving population estimation with neural network models. In: Proceedings of the Third International Conference on Advances in Neural Networks - Volume Part III, ISNN’06, pp. 1181–1186, Springer-Verlag, Berlin, Heidelberg.