Machine Learning in Long-term Mortality Forecasting

Presentor: Yang Qiao

Coauthors: Chou-Wen Wang Wenjun Zhu

outline

- 1 Introduction
- 2 Methodology
- 3 Data and empirical results
- 4 References

Introduction

- We propose a new method for predicting population mortality rate for the next 20 years.
- For the mortality rate at age x in year t, we use its neighbours at adjacent [x-x_1,x+x_2] ages with a lag of k years (1≤k≤s) as inputs to predict the mortality rate at year t+1 using XGBoost, Catboost, Lightgbm, and CNN algorithms.
- and then put the predicted rates back into the inputs again to predict the mortality rate in year t+2, and so on.
- In this way, we can predict the mortality rate at any future time.

Introduction

- By comparing with the actual data, we found that the MAPE is less than those of the three benchmark models: APC, CBD and LC.
- We also put all data from 32 countries and regions into the training dataset, and still use the above model (the difference is that the above model is predicted separately) to predict the mortality rate of all countries and regions at once(hereafter ALL model).
- It is found that the MAPE obtained by the ALL model is only half of the MAPE of the benchmark model LC.

Introduction

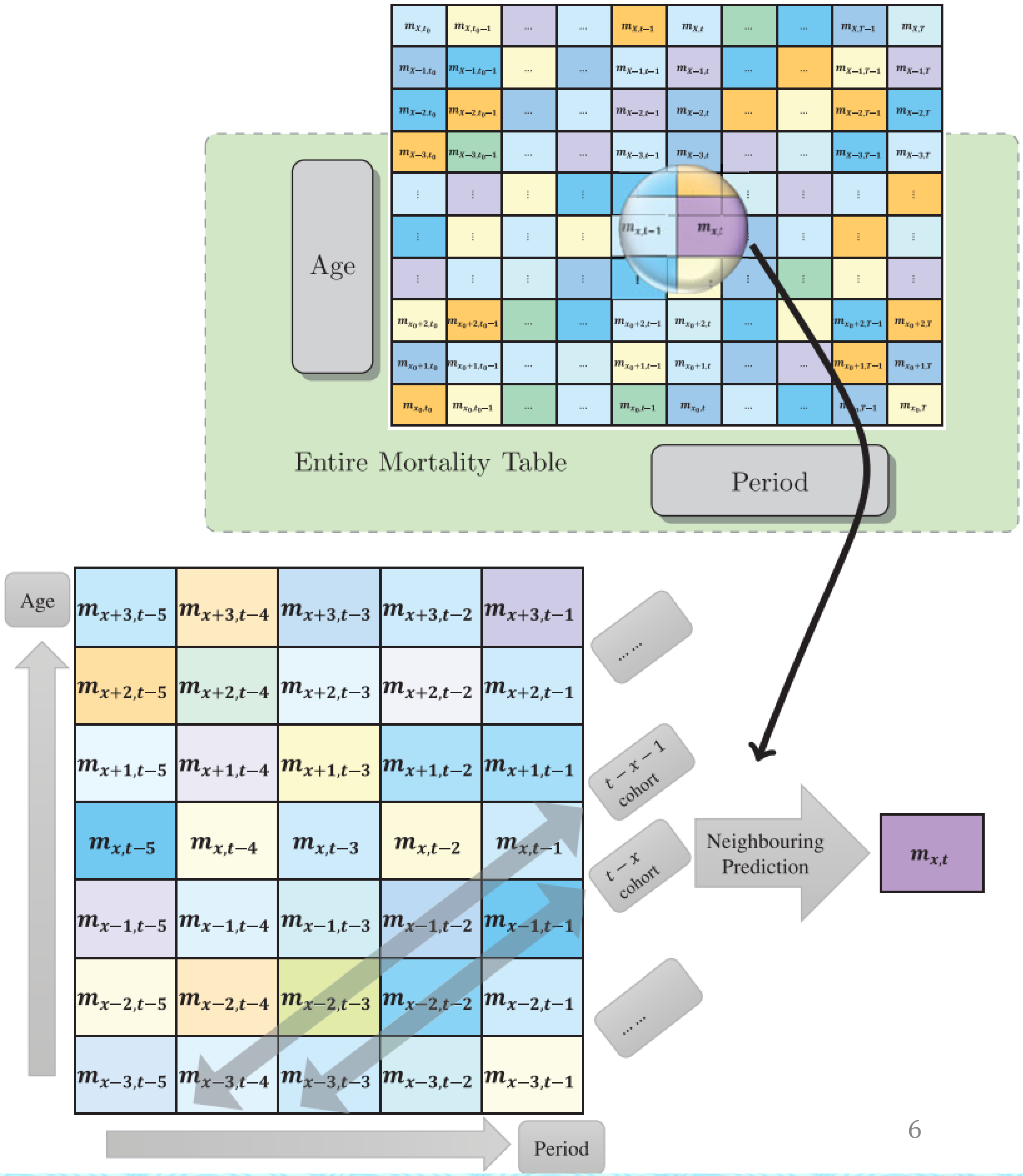
- Our paper is an extension of paper ***NEIGHBOURING PREDICTION FOR MORTALITY(Wang,Zhang & Zhu, 2021)***.
- the main differences are the following two points:
- 1 The model draws on the principle of ensembing, and each Mortality prediction uses 13 different neighbours as the average of the predicted values obtained from the inputs.
- 2 The model combines the ideas of ensembling and boosting together to produce a more reasonable and robust forecast.

Methodology

In the paper *NEIGHBOURING PREDICTION FOR MORTALITY, the authors propose the definition of neighbouring.*

For each mortality rate at age x in year t, denoted as $m_{x,t}$, they construct the 2-dimensional "image" of neighbourhood mortality data around $m_{x,t}$, that is, $\varepsilon_{m_{x,t}}(x_1, x_2, s)$, which includes mortality information for ages in $[x - x_1, x + x_2]$, lagging k years $(1 \le k \le s)$

The right figure illustrates an example of neighbourhood mortality data of $\varepsilon_{m_{x,t}}$ (3, 3, 5), where $x_1 = x_2 = 3$, s = 5



a new model to construct neighbourhood

ages	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	......	2020
63	0.12	0.11	0.11	0.1	0.1	0.09					......	
64	0.13	0.12	0.12	0.11	0.11	0.1					......	
65	0.14	0.14	0.13	0.12	0.12	0.11	0.12	0.12	0.11	0.11	......	0.11
66	0.16	0.15	0.14	0.14	0.13	0.13	0.13	0.13	0.13	0.13	......	0.13
67	0.18	0.17	0.16	0.16	0.15	0.14	0.14	0.14	0.14	0.14	......	0.14
68	0.2	0.19	0.19	0.18	0.16	0.16	0.16	0.15	0.15	0.15	......	0.15
69	0.22	0.21	0.19	0.19	0.18	0.17	0.17	0.17	0.16	0.16	......	0.16
70	0.23	0.23	0.23	0.21	0.21	0.2	0.19	0.18	0.18	0.18	......	0.18
71	0.25	0.24	0.24	0.23	0.22	0.22	0.21	0.2	0.2	0.2	......	0.2
72	0.28	0.27	0.27	0.26	0.25	0.24	0.23	0.23	0.22	0.22	......	0.22
73	0.32	0.3	0.29	0.29	0.28	0.27	0.26	0.25	0.25	0.25	......	0.25
74	0.36	0.32	0.31	0.32	0.31	0.3	0.29	0.28	0.28	0.28	......	0.28
75	0.37	0.37	0.36	0.34	0.33	0.32	0.32	0.31	0.3	0.3	......	0.3
76	0.43	0.39	0.4	0.38	0.36	0.36	0.35	0.35	0.34	0.34	......	0.34
77	0.47	0.44	0.44	0.43	0.41	0.41	0.39	0.39	0.38	0.38	......	0.38
78	0.51	0.49	0.49	0.46	0.46	0.44	0.44	0.43	0.42	0.42	......	0.42
79	0.58	0.55	0.54	0.52	0.49	0.49	0.48	0.48	0.47	0.47	......	0.47
80	0.63	0.62	0.61	0.6	0.58	0.54	0.55	0.53	0.52	0.52	......	0.52
81	0.71	0.67	0.68	0.65	0.62	0.61	0.6	0.59	0.58	0.58	......	0.58
82	0.78	0.73	0.74	0.75	0.68	0.67	0.67	0.65	0.65	0.65	......	0.65
83	0.85	0.81	0.83	0.79	0.78	0.75	0.74	0.73	0.71	0.71	......	0.71
84	0.96	0.93	0.9	0.91	0.88	0.86	0.82	0.8	0.79	0.79	......	0.79
85	1.03	0.99	1.03	1	0.95	0.96	0.92	0.89	0.88	0.88	......	0.88
86	1.15	1.13	1.12	1.11	1.04	1.06	1.03	1	0.97	0.97	......	0.97
87	1.25	1.25	1.22	1.24	1.2	1.15	1.13	1.11	1.09	1.09	......	1.09
88	1.38	1.34	1.35	1.32	1.29	1.31	1.21	1.19	1.17	1.17	......	1.17
89	1.58	1.52	1.48	1.49	1.37	1.4	1.37	1.32	1.3	1.3	......	1.3
90	1.69	1.57	1.66	1.62	1.56	1.55					......	
91	1.79	1.76	1.85	1.8	1.71	1.71					......	

Different from predicting mortality rate of only one year, If method in the above paper is still used, the age data at both ends are constantly facing the problem of missing values, for example, when we predict mortality rate of age 65 in 2001, the neighbourhood should include ages from 59 to 64 in 2000, but these data do not actually exist.

Therefore, **we propose a new model to construct neighbourhood**: for age 65, we use all the ages on the right. For age 89, we use all the ages on the left.

model0: X = 65,66,67,68…78

m(x,t-6)	m(x,t-5)	m(x,t-4)	m(x,t-3)	m(x,t-2)	m(x,t-1)	m(x,t)
m(x+1,t-6)	m(x+1,t-5)	m(x+1,t-4)	m(x+1,t-3)	m(x+1,t-2)	m(x+1,t-1)	
m(x+2,t-6)	m(x+2,t-5)	m(x+2,t-4)	m(x+2,t-3)	m(x+2,t-2)	m(x+2,t-1)	
m(x+3,t-6)	m(x+3,t-5)	m(x+3,t-4)	m(x+3,t-3)	m(x+3,t-2)	m(x+3,t-1)	
m(x+4,t-6)	m(x+4,t-5)	m(x+4,t-4)	m(x+4,t-3)	m(x+4,t-2)	m(x+4,t-1)	
m(x+5,t-6)	m(x+5,t-5)	m(x+5,t-4)	m(x+5,t-3)	m(x+5,t-2)	m(x+5,t-1)	
m(x+6,t-6)	m(x+6,t-5)	m(x+6,t-4)	m(x+6,t-3)	m(x+6,t-2)	m(x+6,t-1)	
m(x+7,t-6)	m(x+7,t-5)	m(x+7,t-4)	m(x+7,t-3)	m(x+7,t-2)	m(x+7,t-1)	
m(x+8,t-6)	m(x+8,t-5)	m(x+8,t-4)	m(x+8,t-3)	m(x+8,t-2)	m(x+8,t-1)	
m(x+9,t-6)	m(x+9,t-5)	m(x+9,t-4)	m(x+9,t-3)	m(x+9,t-2)	m(x+9,t-1)	
m(x+10,t-6)	m(x+10,t-5)	m(x+10,t-4)	m(x+10,t-3)	m(x+10,t-2)	m(x+10,t-1)	
m(x+11,t-6)	m(x+11,t-5)	m(x+11,t-4)	m(x+11,t-3)	m(x+11,t-2)	m(x+11,t-1)	
m(x+12,t-6)	m(x+12,t-5)	m(x+12,t-4)	m(x+12,t-3)	m(x+12,t-2)	m(x+12,t-1)	

model1: X = 66,67,68…78

m(x-1,t-6)	m(x-1,t-5)	m(x-1,t-4)	m(x-1,t-3)	m(x-1,t-2)	m(x-1,t-1)	
m(x,t-6)	m(x,t-5)	m(x,t-4)	m(x,t-3)	m(x,t-2)	m(x,t-1)	m(x,t)
m(x+1,t-6)	m(x+1,t-5)	m(x+1,t-4)	m(x+1,t-3)	m(x+1,t-2)	m(x+1,t-1)	
m(x+2,t-6)	m(x+2,t-5)	m(x+2,t-4)	m(x+2,t-3)	m(x+2,t-2)	m(x+2,t-1)	
m(x+3,t-6)	m(x+3,t-5)	m(x+3,t-4)	m(x+3,t-3)	m(x+3,t-2)	m(x+3,t-1)	
m(x+4,t-6)	m(x+4,t-5)	m(x+4,t-4)	m(x+4,t-3)	m(x+4,t-2)	m(x+4,t-1)	
m(x+5,t-6)	m(x+5,t-5)	m(x+5,t-4)	m(x+5,t-3)	m(x+5,t-2)	m(x+5,t-1)	
m(x+6,t-6)	m(x+6,t-5)	m(x+6,t-4)	m(x+6,t-3)	m(x+6,t-2)	m(x+6,t-1)	
m(x+7,t-6)	m(x+7,t-5)	m(x+7,t-4)	m(x+7,t-3)	m(x+7,t-2)	m(x+7,t-1)	
m(x+8,t-6)	m(x+8,t-5)	m(x+8,t-4)	m(x+8,t-3)	m(x+8,t-2)	m(x+8,t-1)	
m(x+9,t-6)	m(x+9,t-5)	m(x+9,t-4)	m(x+9,t-3)	m(x+9,t-2)	m(x+9,t-1)	
m(x+10,t-6)	m(x+10,t-5)	m(x+10,t-4)	m(x+10,t-3)	m(x+10,t-2)	m(x+10,t-1)	
m(x+11,t-6)	m(x+11,t-5)	m(x+11,t-4)	m(x+11,t-3)	m(x+11,t-2)	m(x+11,t-1)	

Model2: x = 67,68…79

m(x-2,t-6)	m(x-2,t-5)	m(x-2,t-4)	m(x-2,t-3)	m(x-2,t-2)	m(x-2,t-1)	
m(x-1,t-6)	m(x-1,t-5)	m(x-1,t-4)	m(x-1,t-3)	m(x-1,t-2)	m(x-1,t-1)	
m(x,t-6)	m(x,t-5)	m(x,t-4)	m(x,t-3)	m(x,t-2)	m(x,t-1)	m(x,t)
m(x+1,t-6)	m(x+1,t-5)	m(x+1,t-4)	m(x+1,t-3)	m(x+1,t-2)	m(x+1,t-1)	
m(x+2,t-6)	m(x+2,t-5)	m(x+2,t-4)	m(x+2,t-3)	m(x+2,t-2)	m(x+2,t-1)	
m(x+3,t-6)	m(x+3,t-5)	m(x+3,t-4)	m(x+3,t-3)	m(x+3,t-2)	m(x+3,t-1)	
m(x+4,t-6)	m(x+4,t-5)	m(x+4,t-4)	m(x+4,t-3)	m(x+4,t-2)	m(x+4,t-1)	
m(x+5,t-6)	m(x+5,t-5)	m(x+5,t-4)	m(x+5,t-3)	m(x+5,t-2)	m(x+5,t-1)	
m(x+6,t-6)	m(x+6,t-5)	m(x+6,t-4)	m(x+6,t-3)	m(x+6,t-2)	m(x+6,t-1)	
m(x+7,t-6)	m(x+7,t-5)	m(x+7,t-4)	m(x+7,t-3)	m(x+7,t-2)	m(x+7,t-1)	
m(x+8,t-6)	m(x+8,t-5)	m(x+8,t-4)	m(x+8,t-3)	m(x+8,t-2)	m(x+8,t-1)	
m(x+9,t-6)	m(x+9,t-5)	m(x+9,t-4)	m(x+9,t-3)	m(x+9,t-2)	m(x+9,t-1)	
m(x+10,t-6)	m(x+10,t-5)	m(x+10,t-4)	m(x+10,t-3)	m(x+10,t-2)	m(x+10,t-1)	

Model3: x = 68…80

m(x-3,t-6)	m(x-3,t-5)	m(x-3,t-4)	m(x-3,t-3)	m(x-3,t-2)	m(x-3,t-1)	
m(x-2,t-6)	m(x-2,t-5)	m(x-2,t-4)	m(x-2,t-3)	m(x-2,t-2)	m(x-2,t-1)	
m(x-1,t-6)	m(x-1,t-5)	m(x-1,t-4)	m(x-1,t-3)	m(x-1,t-2)	m(x-1,t-1)	
m(x,t-6)	m(x,t-5)	m(x,t-4)	m(x,t-3)	m(x,t-2)	m(x,t-1)	m(x,t)
m(x+1,t-6)	m(x+1,t-5)	m(x+1,t-4)	m(x+1,t-3)	m(x+1,t-2)	m(x+1,t-1)	
m(x+2,t-6)	m(x+2,t-5)	m(x+2,t-4)	m(x+2,t-3)	m(x+2,t-2)	m(x+2,t-1)	
m(x+3,t-6)	m(x+3,t-5)	m(x+3,t-4)	m(x+3,t-3)	m(x+3,t-2)	m(x+3,t-1)	
m(x+4,t-6)	m(x+4,t-5)	m(x+4,t-4)	m(x+4,t-3)	m(x+4,t-2)	m(x+4,t-1)	
m(x+5,t-6)	m(x+5,t-5)	m(x+5,t-4)	m(x+5,t-3)	m(x+5,t-2)	m(x+5,t-1)	
m(x+6,t-6)	m(x+6,t-5)	m(x+6,t-4)	m(x+6,t-3)	m(x+6,t-2)	m(x+6,t-1)	
m(x+7,t-6)	m(x+7,t-5)	m(x+7,t-4)	m(x+7,t-3)	m(x+7,t-2)	m(x+7,t-1)	
m(x+8,t-6)	m(x+8,t-5)	m(x+8,t-4)	m(x+8,t-3)	m(x+8,t-2)	m(x+8,t-1)	
m(x+9,t-6)	m(x+9,t-5)	m(x+9,t-4)	m(x+9,t-3)	m(x+9,t-2)	m(x+9,t-1)	

model11: x = 76…88

m(x-11,t-6)	m(x-11,t-5)	m(x-11,t-4)	m(x-11,t-3)	m(x-11,t-2)	m(x-11,t-1)	
m(x-10,t-6)	m(x-10,t-5)	m(x-10,t-4)	m(x-10,t-3)	m(x-10,t-2)	m(x-10,t-1)	
m(x-9,t-6)	m(x-9,t-5)	m(x-9,t-4)	m(x-9,t-3)	m(x-9,t-2)	m(x-9,t-1)	
m(x-8,t-6)	m(x-8,t-5)	m(x-8,t-4)	m(x-8,t-3)	m(x-8,t-2)	m(x-8,t-1)	
m(x-7,t-6)	m(x-7,t-5)	m(x-7,t-4)	m(x-7,t-3)	m(x-7,t-2)	m(x-7,t-1)	
m(x-6,t-6)	m(x-6,t-5)	m(x-6,t-4)	m(x-6,t-3)	m(x-6,t-2)	m(x-6,t-1)	
m(x-5,t-6)	m(x-5,t-5)	m(x-5,t-4)	m(x-5,t-3)	m(x-5,t-2)	m(x-5,t-1)	
m(x-4,t-6)	m(x-4,t-5)	m(x-4,t-4)	m(x-4,t-3)	m(x-4,t-2)	m(x-4,t-1)	
m(x-3,t-6)	m(x-3,t-5)	m(x-3,t-4)	m(x-3,t-3)	m(x-3,t-2)	m(x-3,t-1)	
m(x-2,t-6)	m(x-2,t-5)	m(x-2,t-4)	m(x-2,t-3)	m(x-2,t-2)	m(x-2,t-1)	
m(x-1,t-6)	m(x-1,t-5)	m(x-1,t-4)	m(x-1,t-3)	m(x-1,t-2)	m(x-1,t-1)	
m(x,t-6)	m(x,t-5)	m(x,t-4)	m(x,t-3)	m(x,t-2)	m(x,t-1)	m(x,t)
m(x+1,t-6)	m(x+1,t-5)	m(x+1,t-4)	m(x+1,t-3)	m(x+1,t-2)	m(x+1,t-1)	

model12:   x = 77…88,89

m(x-12,t-6)	m(x-12,t-5)	m(x-12,t-4)	m(x-12,t-3)	m(x-12,t-2)	m(x-12,t-1)	
m(x-11,t-6)	m(x-11,t-5)	m(x-11,t-4)	m(x-11,t-3)	m(x-11,t-2)	m(x-11,t-1)	
m(x-10,t-6)	m(x-10,t-5)	m(x-10,t-4)	m(x-10,t-3)	m(x-10,t-2)	m(x-10,t-1)	
m(x-9,t-6)	m(x-9,t-5)	m(x-9,t-4)	m(x-9,t-3)	m(x-9,t-2)	m(x-9,t-1)	
m(x-8,t-6)	m(x-8,t-5)	m(x-8,t-4)	m(x-8,t-3)	m(x-8,t-2)	m(x-8,t-1)	
m(x-7,t-6)	m(x-7,t-5)	m(x-7,t-4)	m(x-7,t-3)	m(x-7,t-2)	m(x-7,t-1)	
m(x-6,t-6)	m(x-6,t-5)	m(x-6,t-4)	m(x-6,t-3)	m(x-6,t-2)	m(x-6,t-1)	
m(x-5,t-6)	m(x-5,t-5)	m(x-5,t-4)	m(x-5,t-3)	m(x-5,t-2)	m(x-5,t-1)	
m(x-4,t-6)	m(x-4,t-5)	m(x-4,t-4)	m(x-4,t-3)	m(x-4,t-2)	m(x-4,t-1)	
m(x-3,t-6)	m(x-3,t-5)	m(x-3,t-4)	m(x-3,t-3)	m(x-3,t-2)	m(x-3,t-1)	
m(x-2,t-6)	m(x-2,t-5)	m(x-2,t-4)	m(x-2,t-3)	m(x-2,t-2)	m(x-2,t-1)	
m(x-1,t-6)	m(x-1,t-5)	m(x-1,t-4)	m(x-1,t-3)	m(x-1,t-2)	m(x-1,t-1)	
m(x,t-6)	m(x,t-5)	m(x,t-4)	m(x,t-3)	m(x,t-2)	m(x,t-1)	m(x,t)

In total, we construct **13** such models corresponsding to diffreent location of the mortality rate to be predicted.

Model0 for predicting age 68 at 2001

m(68,1995)	m(68,1996)	m(68,1997)	m(68,1998)	m(68,1999)	m(68,2000)	m(68,2001)
m(69,1995)	m(69,1996)	m(69,1997)	m(69,1998)	m(69,1999)	m(69,2000)	
m(70,1995)	m(70,1996)	m(70,1997)	m(70,1998)	m(70,1999)	m(70,2000)	
m(71,1995)	m(71,1996)	m(71,1997)	m(71,1998)	m(71,1999)	m(71,2000)	
m(72,1995)	m(72,1996)	m(72,1997)	m(72,1998)	m(72,1999)	m(72,2000)	
m(73,1995)	m(73,1996)	m(73,1997)	m(73,1998)	m(73,1999)	m(73,2000)	
m(74,1995)	m(74,1996)	m(74,1997)	m(74,1998)	m(74,1999)	m(74,2000)	
m(75,1995)	m(75,1996)	m(75,1997)	m(75,1998)	m(75,1999)	m(75,2000)	
m(76,1995)	m(76,1996)	m(76,1997)	m(76,1998)	m(76,1999)	m(76,2000)	
m(77,1995)	m(77,1996)	m(77,1997)	m(77,1998)	m(77,1999)	m(77,2000)	
m(78,1995)	m(78,1996)	m(78,1997)	m(78,1998)	m(78,1999)	m(78,2000)	
m(79,1995)	m(79,1996)	m(79,1997)	m(79,1998)	m(79,1999)	m(79,2000)	
m(80,1995)	m(80,1996)	m(80,1997)	m(80,1998)	m(80,1999)	m(80,2000)	

Model1 for predicting age 68 at 2001

m(67,1995)	m(67,1996)	m(67,1997)	m(67,1998)	m(67,1999)	m(67,2000)	
m(68,1995)	m(68,1996)	m(68,1997)	m(68,1998)	m(68,1999)	m(68,2000)	m(68, 2001)
m(69,1995)	m(69,1996)	m(69,1997)	m(69,1998)	m(69,1999)	m(69,2000)	
m(70,1995)	m(70,1996)	m(70,1997)	m(70,1998)	m(70,1999)	m(70,2000)	
m(71,1995)	m(71,1996)	m(71,1997)	m(71,1998)	m(71,1999)	m(71,2000)	
m(72,1995)	m(72,1996)	m(72,1997)	m(72,1998)	m(72,1999)	m(72,2000)	
m(73,1995)	m(73,1996)	m(73,1997)	m(73,1998)	m(73,1999)	m(73,2000)	
m(74,1995)	m(74,1996)	m(74,1997)	m(74,1998)	m(74,1999)	m(74,2000)	
m(75,1995)	m(75,1996)	m(75,1997)	m(75,1998)	m(75,1999)	m(75,2000)	
m(76,1995)	m(76,1996)	m(76,1997)	m(76,1998)	m(76,1999)	m(76,2000)	
m(77,1995)	m(77,1996)	m(77,1997)	m(77,1998)	m(77,1999)	m(77,2000)	
m(78,1995)	m(78,1996)	m(78,1997)	m(78,1998)	m(78,1999)	m(78,2000)	
m(79,1995)	m(79,1996)	m(79,1997)	m(79,1998)	m(79,1999)	m(79,2000)	

Model2 for predicting age 68 at 2001

m(66,1995)	m(66,1996)	m(66,1997)	m(66,1998)	m(66,1999)	m(66,2000)	
m(67,1995)	m(67,1996)	m(67,1997)	m(67,1998)	m(67,1999)	m(67,2000)	
m(68,1995)	m(68,1996)	m(68,1997)	m(68,1998)	m(68,1999)	m(68,2000)	m(68, 2001)
m(69,1995)	m(69,1996)	m(69,1997)	m(69,1998)	m(69,1999)	m(69,2000)	
m(70,1995)	m(70,1996)	m(70,1997)	m(70,1998)	m(70,1999)	m(70,2000)	
m(71,1995)	m(71,1996)	m(71,1997)	m(71,1998)	m(71,1999)	m(71,2000)	
m(72,1995)	m(72,1996)	m(72,1997)	m(72,1998)	m(72,1999)	m(72,2000)	
m(73,1995)	m(73,1996)	m(73,1997)	m(73,1998)	m(73,1999)	m(73,2000)	
m(74,1995)	m(74,1996)	m(74,1997)	m(74,1998)	m(74,1999)	m(74,2000)	
m(75,1995)	m(75,1996)	m(75,1997)	m(75,1998)	m(75,1999)	m(75,2000)	
m(76,1995)	m(76,1996)	m(76,1997)	m(76,1998)	m(76,1999)	m(76,2000)	
m(77,1995)	m(77,1996)	m(77,1997)	m(77,1998)	m(77,1999)	m(77,2000)	
m(78,1995)	m(78,1996)	m(78,1997)	m(78,1998)	m(78,1999)	m(78,2000)	

Model3 for predicting age 68 at 2001

m(65,1995)	m(65,1996)	m(65,1997)	m(65,1998)	m(65,1999)	m(65,2000)	
m(66,1995)	m(66,1996)	m(66,1997)	m(66,1998)	m(66,1999)	m(66,2000)	
m(67,1995)	m(67,1996)	m(67,1997)	m(67,1998)	m(67,1999)	m(67,2000)	
m(68,1995)	m(68,1996)	m(68,1997)	m(68,1998)	m(68,1999)	m(68,2000)	m(68, 2001)
m(69,1995)	m(69,1996)	m(69,1997)	m(69,1998)	m(69,1999)	m(69,2000)	
m(70,1995)	m(70,1996)	m(70,1997)	m(70,1998)	m(70,1999)	m(70,2000)	
m(71,1995)	m(71,1996)	m(71,1997)	m(71,1998)	m(71,1999)	m(71,2000)	
m(72,1995)	m(72,1996)	m(72,1997)	m(72,1998)	m(72,1999)	m(72,2000)	
m(73,1995)	m(73,1996)	m(73,1997)	m(73,1998)	m(73,1999)	m(73,2000)	
m(74,1995)	m(74,1996)	m(74,1997)	m(74,1998)	m(74,1999)	m(74,2000)	
m(75,1995)	m(75,1996)	m(75,1997)	m(75,1998)	m(75,1999)	m(75,2000)	
m(76,1995)	m(76,1996)	m(76,1997)	m(76,1998)	m(76,1999)	m(76,2000)	
m(77,1995)	m(77,1996)	m(77,1997)	m(77,1998)	m(77,1999)	m(77,2000)	

Model4 can't be used for predicting age 68 at 2001 because m(64,2000) actually does not exist!!!

m(64,1995)	m(64,1996)	m(64,1997)	m(64,1998)	m(64,1999)	??????	
m(65,1995)	m(65,1996)	m(65,1997)	m(65,1998)	m(65,1999)	m(65,2000)	
m(66,1995)	m(66,1996)	m(66,1997)	m(66,1998)	m(66,1999)	m(66,2000)	
m(67,1995)	m(67,1996)	m(67,1997)	m(67,1998)	m(67,1999)	m(67,2000)	
m(68,1995)	m(68,1996)	m(68,1997)	m(68,1998)	m(68,1999)	m(68,2000)	m(68, 2001)
m(69,1995)	m(69,1996)	m(69,1997)	m(69,1998)	m(69,1999)	m(69,2000)	
m(70,1995)	m(70,1996)	m(70,1997)	m(70,1998)	m(70,1999)	m(70,2000)	
m(71,1995)	m(71,1996)	m(71,1997)	m(71,1998)	m(71,1999)	m(71,2000)	
m(72,1995)	m(72,1996)	m(72,1997)	m(72,1998)	m(72,1999)	m(72,2000)	
m(73,1995)	m(73,1996)	m(73,1997)	m(73,1998)	m(73,1999)	m(73,2000)	
m(74,1995)	m(74,1996)	m(74,1997)	m(74,1998)	m(74,1999)	m(74,2000)	
m(75,1995)	m(75,1996)	m(75,1997)	m(75,1998)	m(75,1999)	m(75,2000)	
m(76,1995)	m(76,1996)	m(76,1997)	m(76,1998)	m(76,1999)	m(76,2000)	

In order to avoid using future data or null values, **only model0 is used for age 65** (if model1 is used, age 64 on the left will be used, and age 64 is not in age 65-89, so it is a null value),

Age 66 uses the average of model0 and model1.

......

age 77 is the average of all 13 models.

......

Age 88 uses only model11 and model12.

For age 89, only the age on the left is used, which is model12.

The orange bottom of the table below shows the models used for each age.

age	model 0	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10	model 11	model 12	predict
65	0.12													0.12
66	0.13	0.13												0.13
67	0.14	0.14	0.14											0.14
68	0.16	0.16	0.16	0.16										0.16
69	0.18	0.18	0.18	0.18	0.18									0.18
70	0.19	0.19	0.19	0.19	0.19	0.19								0.19
71	0.21	0.21	0.21	0.21	0.21	0.21	0.21							0.21
72	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23						0.23
73	0.26	0.25	0.25	0.26	0.26	0.25	0.25	0.25	0.25					0.25
74	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28				0.28
75	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31			0.31
76	0.33	0.34	0.34	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33		0.33
77	0.36	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37
78		0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
79			0.44	0.44	0.45	0.45	0.45	0.44	0.44	0.45	0.44	0.45	0.44	0.45
80				0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
81					0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54
82						0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
83							0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
84								0.74	0.74	0.74	0.74	0.73	0.73	0.74
85									0.82	0.82	0.82	0.82	0.82	0.82
86										0.92	0.92	0.92	0.92	0.92
87											1.02	1.02	1.02	1.02
88												1.13	1.13	1.13
89													1.26	1.26

The last column is the final prediction which is the average of corresponidng models used for each age.

Two advantages of our model

ages	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	......	2020
63	0.12	0.11	0.11	0.1	0.1	0.09					......	
64	0.13	0.12	0.12	0.11	0.11	0.1					......	
65	0.14	0.14	0.13	0.12	0.12	0.11	0.12	0.12	0.11	0.11	......	0.11
66	0.16	0.15	0.14	0.14	0.13	0.13	0.13	0.13	0.13	0.13	......	0.13
67	0.18	0.17	0.16	0.16	0.15	0.14	0.14	0.14	0.14	0.14	......	0.14
68	0.2	0.19	0.19	0.18	0.16	0.16	0.16	0.15	0.15	0.15	......	0.15
69	0.22	0.21	0.19	0.19	0.18	0.17	0.17	0.17	0.16	0.16	......	0.16
70	0.23	0.23	0.23	0.21	0.21	0.2	0.19	0.18	0.18	0.18	......	0.18
71	0.25	0.24	0.24	0.23	0.22	0.22	0.21	0.2	0.2	0.2	......	0.2
72	0.28	0.27	0.27	0.26	0.25	0.24	0.23	0.23	0.22	0.22	......	0.22
73	0.32	0.3	0.29	0.29	0.28	0.27	0.26	0.25	0.25	0.25	......	0.25
74	0.36	0.32	0.31	0.32	0.31	0.3	0.29	0.28	0.28	0.28	......	0.28
75	0.37	0.37	0.36	0.34	0.33	0.32	0.32	0.31	0.3	0.3	......	0.3
76	0.43	0.39	0.4	0.38	0.36	0.36	0.35	0.35	0.34	0.34	......	0.34
77	0.47	0.44	0.44	0.43	0.41	0.41	0.39	0.39	0.38	0.38	......	0.38
78	0.51	0.49	0.49	0.46	0.46	0.44	0.44	0.43	0.42	0.42	......	0.42
79	0.58	0.55	0.54	0.52	0.49	0.49	0.48	0.48	0.47	0.47	......	0.47
80	0.63	0.62	0.61	0.6	0.58	0.54	0.55	0.53	0.52	0.52	......	0.52
81	0.71	0.67	0.68	0.65	0.62	0.61	0.6	0.59	0.58	0.58	......	0.58
82	0.78	0.73	0.74	0.75	0.68	0.67	0.67	0.65	0.65	0.65	......	0.65
83	0.85	0.81	0.83	0.79	0.78	0.75	0.74	0.73	0.71	0.71	......	0.71
84	0.96	0.93	0.9	0.91	0.88	0.86	0.82	0.8	0.79	0.79	......	0.79
85	1.03	0.99	1.03	1	0.95	0.96	0.92	0.89	0.88	0.88	......	0.88
86	1.15	1.13	1.12	1.11	1.04	1.06	1.03	1	0.97	0.97	......	0.97
87	1.25	1.25	1.22	1.24	1.2	1.15	1.13	1.11	1.09	1.09	......	1.09
88	1.38	1.34	1.35	1.32	1.29	1.31	1.21	1.19	1.17	1.17	......	1.17
89	1.58	1.52	1.48	1.49	1.37	1.4	1.37	1.32	1.3	1.3	......	1.3
90	1.69	1.57	1.66	1.62	1.56	1.55					......	
91	1.79	1.76	1.85	1.8	1.71	1.71					......	

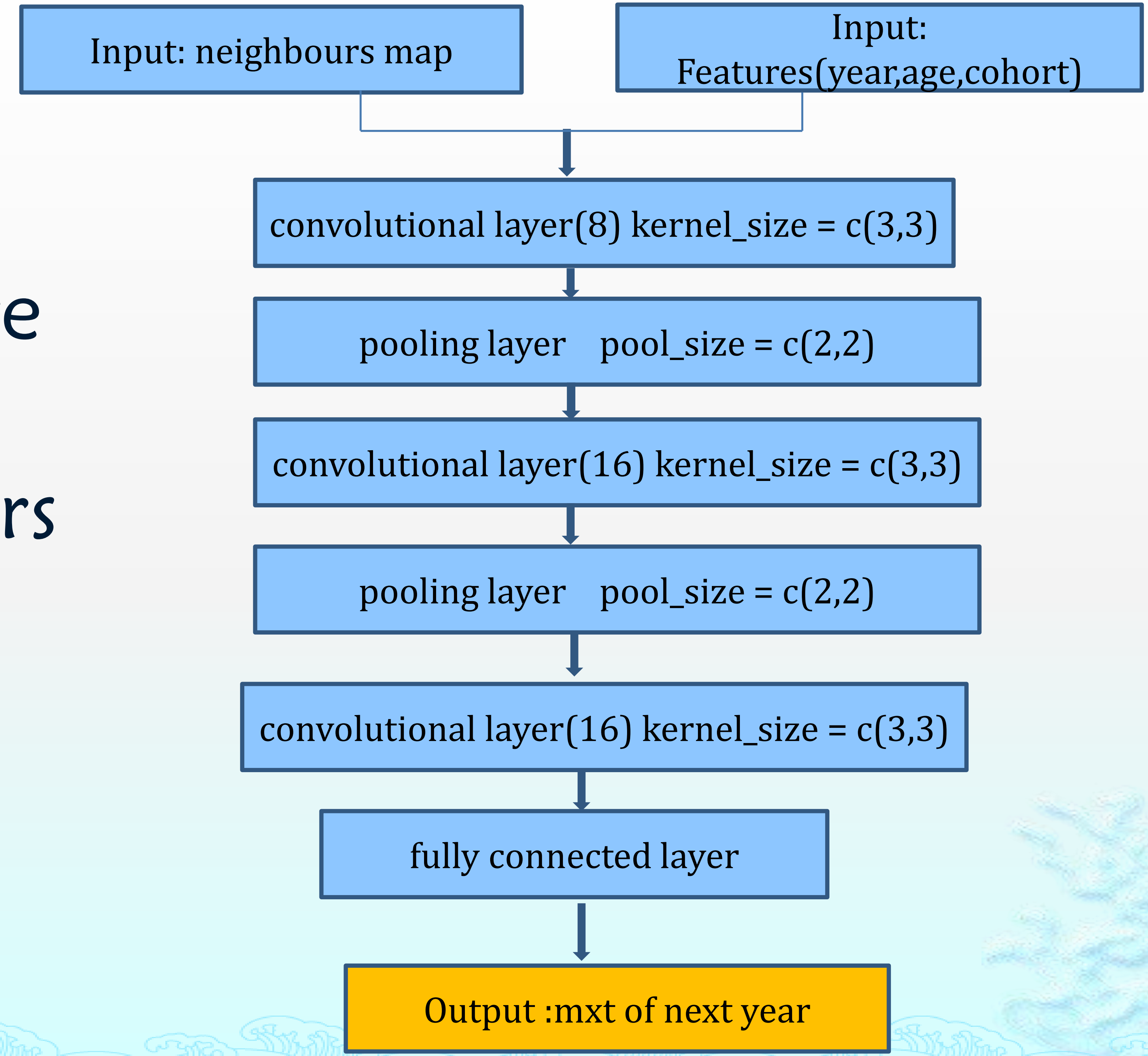
we adopt the above ensembling approach in order to solve the missing-value problem.

On the other hand, the final prediction is the average of the predictions calculated from multiple neighborhoods, which effectively smoothes out the noise in the data and results in a lower predicted MAPE.

Methodology

- After constructing the neighborhood data, we use two types of machine learning methods for prediction:
  - neural network (convolutional neural network methods).
  - Boosting

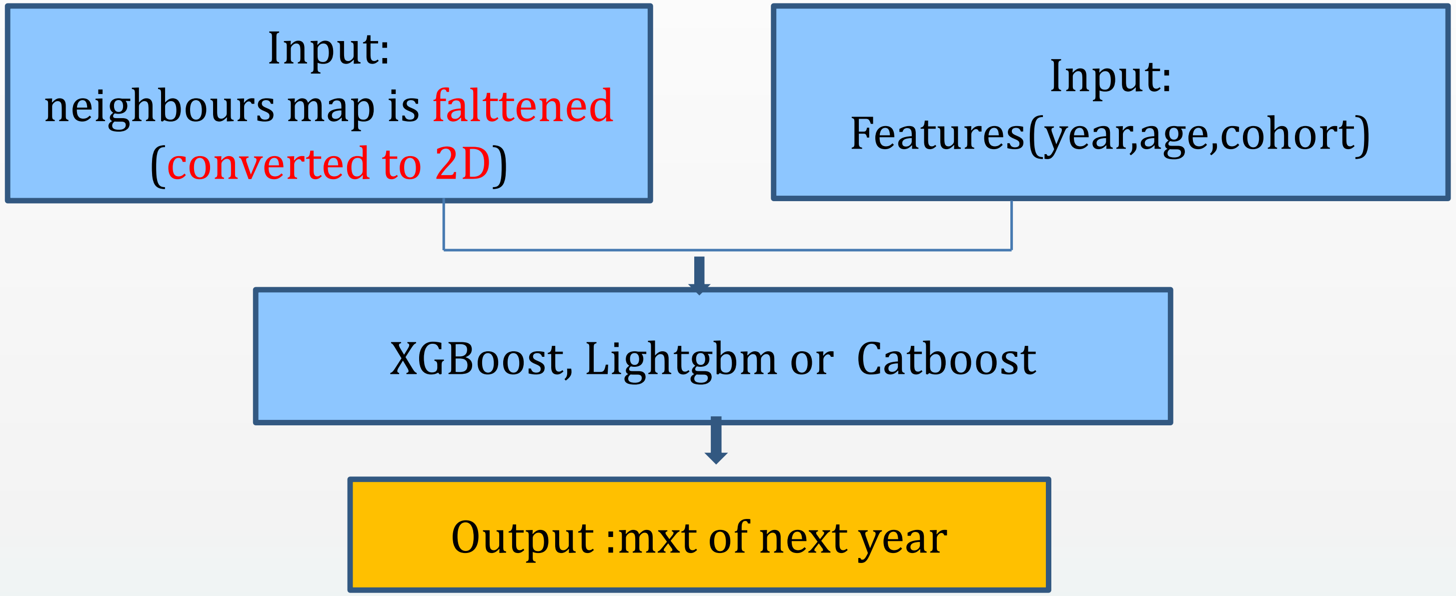
Architecture and Parameters of Model CNN



Boosting models: XGBoost, Lightgbm and Catboost

- Based on GBDT, XGBoost adds a second-order Taylor expansion term and a penalty term to the objective function, which makes it a more efficient and high-precision framework .
- Lightgbm usually has better performance and faster speed by adopting leaf-wise spliting nodes technics and histogram algorithm.
- Catboost's name comes from words category and boosting showing this model is good at handling categorical variables.

Architecture and Parameters of Boosting models



	max_depth	learning_rate	max_round (iteration number)
Xgboost	6	0.01	3000
Lightgbm	6	0.01	3000
Catboost	6	0.01	3000

Data

- Data source:
  32 countries and regions with complete data before 1960 from Human Mortality Database

- Training period:
  From beginning to 1999

- Test period:
  from 2000 to 2020.

Abbrev	Sample Period	Sample Size	Abbrev	Sample Period	Sample Size
AUS	1921-2019	2325	ISL	1901-2018	2800
AUT	1947-2019	1675	ITA	1872-2019	3550
BEL	1841-2020	4075	JPN	1947-2020	1700
BGR	1947-2017	1625	LTU	1959-2020	1400
BLR	1959-2018	1350	LUX	1960-2020	1375
CAN	1921-2019	2325	LVA	1959-2019	1375
CHE	1876-2020	3475	NLD	1850-2019	4100
CZE	1950-2019	1600	NOR	1846-2020	4225
DNK	1835-2021	4525	NZL	1948-2013	1500
ESP	1908-2020	2675	POL	1958-2019	1400
EST	1959-2019	1375	PRT	1940-2020	1875
FIN	1878-2020	3425	RUS	1959-2014	1250
FRA	1816-2019	4950	SVK	1950-2019	1600
GBR	1922-2018	2275	SWE	1751-2020	6600
HUN	1950-2020	1625	UKR	1959-2013	1225
IRL	1950-2017	1550	USA	1933-2019	2025

Table 2. **Summary of Prediction (%) Results for Benchmark Models**

1 MAPE of LC,CBD and APC is higher than those of CNN and CNN(ALL)

2 in the table, CNN(ALL) has the lowest MAPE for almost each year

Year	LC	CBD	APC	CNN	CNN (All)
2000	7.12	5.56	6.29	4.36	4.22
2001	8.41	6.19	7.45	5.2	4.99
2002	8.55	6.08	7.51	5.67	5.73
2003	8.88	6.6	8.09	6.54	6.97
2004	12.21	8.91	10.68	8.09	7.39
2005	12.56	9.08	10.8	8.53	8.55
2006	14.46	10.92	12.42	9.89	9.19
2007	14.89	11.42	12.92	10.55	9.56
2008	16.31	12.61	14.43	11.66	10.64
2009	17.51	13.85	15.39	12.77	10.96
2010	18.32	14.65	16.05	13.51	12.05
2011	20.36	16.75	17.57	15.32	12.62
2012	20.32	16.91	17.89	15.68	13.24
2013	21.38	17.96	18.91	16.53	13.74
2014	23.06	19.65	19.38	18.17	13.6
2015	21.43	18.13	18.54	17.14	13.5
2016	22.56	19.32	19.53	18.18	14.11
2017	22.39	19.04	19.1	18.6	14.01
2018	22.46	18.94	18.89	18.51	14.41
2019	23.14	20.03	18.74	18.56	14.41
2020	18.04	15.47	17.55	17.05	13.68
Avg	16.87	13.72	14.67	12.88	10.84

Table 3. **Summary of Prediction (%) Results for Different Ensemble Models**

1 Catboost here is the best model,even better than CNN(ALL) in terms of average MAPE

2 Three boosting models are better than CNN

Year	XGboost	Catboost	Lightgbm	CNN	CNN (All)
2000	3.77	3.47	3.82	4.36	4.22
2001	4.7	4.24	4.66	5.2	4.99
2002	4.58	4.39	4.69	5.67	5.73
2003	5.34	4.96	5.39	6.54	6.97
2004	7.44	6.6	7.46	8.09	7.39
2005	7.37	6.47	7.43	8.53	8.55
2006	9.14	8	9.17	9.89	9.19
2007	9.5	8.39	9.54	10.55	9.56
2008	10.43	8.97	10.56	11.66	10.64
2009	11.52	9.67	11.59	12.77	10.96
2010	12.13	10.2	12.25	13.51	12.05
2011	13.81	11.15	13.78	15.32	12.62
2012	13.94	11.63	14.05	15.68	13.24
2013	14.73	12.41	14.9	16.53	13.74
2014	16.07	13.45	16.11	18.17	13.6
2015	14.67	13.44	14.75	17.14	13.5
2016	15.4	14.01	15.51	18.18	14.11
2017	15.35	14.74	15.45	18.6	14.01
2018	15.02	14.75	15.31	18.51	14.41
2019	15.26	14.96	15.34	18.56	14.41
2020	13.99	17.27	14.11	17.05	13.68
Avg	11.15	10.15	11.23	12.88	10.84

Table 4. Summary of Prediction (%) Results for Ensemble Models with All Data Included

1. **Three All models are better than Catboost model in terms of Average MAPE**

2. **Average MAPE of All models is nearly half of benchmark model LC**

Year	XGboost (ALL)	Catboost (ALL)	Lightgbm (ALL)	Catboost	LC
2000	3.56	3.6	3.66	3.47	7.12
2001	4.16	4.25	4.27	4.24	8.41
2002	4.15	4.26	4.27	4.39	8.55
2003	4.71	4.93	4.72	4.96	8.88
2004	6.53	6.1	6.54	6.6	12.21
2005	6.55	6.24	6.64	6.47	12.56
2006	8.2	7.19	8.21	8	14.46
2007	8.49	7.48	8.43	8.39	14.89
2008	8.61	7.92	8.65	8.97	16.31
2009	9.06	8.35	9.18	9.67	17.51
2010	9.5	8.65	9.74	10.2	18.32
2011	10.35	9.68	10.66	11.15	20.36
2012	10.26	9.7	10.39	11.63	20.32
2013	10.58	10.14	11	12.41	21.38
2014	11.38	10.5	12.18	13.45	23.06
2015	9.86	10.31	10.71	13.44	21.43
2016	10.26	10.5	11.37	14.01	22.56
2017	10.66	11.12	11.32	14.74	22.39
2018	9.8	11.38	10.84	14.75	22.46
2019	8.73	10.72	10.17	14.96	23.14
2020	11.69	18.5	12.17	17.27	18.04
Avg	8.43	8.64	8.82	10.15	16.87

Conclusion

- We construct a new framework to forecast long-term motality. Using the average prediction of different neighbourhoods for each age, our model solves the missing-value problem when constructing neighbourhood and smoothes out noise in the data, resulting in a much lower MAPE compared with benchmark models LC, CBD and APC.
- *Three boosting ALL models achieve superior forecasting performance than their individual counterparts, which is consistent with* the paper *NEIGHBOURING PREDICTION FOR MORTALITY.*
- *Boosting models are generally better than* neural networks in our paper.
- Machine learning models in this paper might be extended to other fields, which is left for future work.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learn- ing. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283.
- Blake, D., Cairns, A., Coughlan, G., Dowd, K., and MacMinn, R. (2013). The new life market. *Journal of Risk and Insurance*, 80(3):501–558.
- Blake, D., MacMinn, R., Tsai, J. C., and Wang, J. (2018). Longevity risk and capital markets: The 2017-18 update. *Pension Institute Discussion Paper PI-1908*.
- Bottou, L. and Bousquet, O. (2008). The tradeoffs of large scale learning. In *Advances in neural information processing systems, eds. M. Jordan, Y. LeCun and S. Solla*, pages 161–168.
- Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.
- Caruana, R., Lawrence, S., and Giles, C. L. (2001). Overfitting in neural nets: Backpropaga- tion, conjugate gradient, and early stopping. In *Advances in neural information processing systems*, pages 402–408.
- Chen, H., MacMinn, R., and Sun, T. (2015). Multi-population mortality model: A factor copula approach. *Insurance: Mathematics and Economics*, 63:135–146.
- Chollet, F. et al. (2018). Keras: The Python deep learning library. *Astrophysics Source Code Library*.
- Dong, Y., Huang, F., Yu, H., and Haberman, S. (2020). Multi-population mortality fore- casting using tensor decomposition. *Scandinavian Actuarial Journal, forthcoming*.
- Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., and Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, 15(2):334–356.

References

- Hainaut, D. (2018). A neural-network analyzer for mortality forecast. *ASTIN Bulletin: The Journal of the IAA*, 48(2):481–508.
- Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83– 85.
- Jarner, S. F. and Kryger, E. M. (2011). Modelling adult mortality in small populations: The SAINT model. *ASTIN Bulletin*, 41(2):377–418.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419):659–671.
- Li, J. S.-H., Chan, W.-S., and Zhou, R. (2017). Semicoherent multipopulation mortality modeling: The impact on longevity risk securitization. *Journal of Risk and Insurance*, 84(3):1025–1065.
- Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of population: An extension to the classical Lee-Carter approach. *Demography*, 42(3):575–594.
- Perla, F., Richman, R., Scognamiglio, S., and Wuthrich, M. V. (2020). Time-series forecast- ing of mortality rates using deep learning. *Available at SSRN*.
- Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3):556–570.

References

- Richman, R. (2018). AI in actuarial science. *Available at SSRN 3218082*.
- Richman, R. and Wu ¨ thrich, M. V. (2019). A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*, page forthcoming.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958.
- Wang, C.-W., Yang, S. S., and Huang, H.-C. (2015). Modeling multi-country mortality de- pendence and its application in pricing survivor index swaps—a dynamic copula approach. *Insurance: Mathematics and Economics*, 63:30–39.
- Wang, H.-C., Yue, C.-S. J., and Chong, C.-T. (2018). Mortality models and longevity risk for small populations. *Insurance: Mathematics and Economics*, 78:351–359.
- Wang, C. W., Zhang, J., & Zhu, W. (2021). Neighbouring prediction for mortality. *ASTIN Bulletin: The Journal of the IAA, 51*(3), 689-718.
- Zhou, R., Li, J. S.-H., and Tan, K. S. (2013). Pricing standardized mortality securitizations: A two-population model with transitory jump effects. *Journal of Risk and Insurance*, 80(3):733–774.
- Zhou, Y.-T. and Chellappa, R. (1988). Computation of optical flow using a neural network.
- In *IEEE International Conference on Neural Networks*, volume 1998, pages 71–78.
- Zhu, W., Tan, K. S., and Wang, C.-W. (2017). Modeling multicountry longevity risk with mortality dependence: A L´evy subordinated hierarchical Archimedean copulas approach. *Journal of Risk and Insurance*, 84(S1):477–493.

Thank you for listening !