A Novel Encoder-Decoder Model based on Read-first LSTM for Air Pollutant Prediction

Bo Zhang,Guojian Zou#, Dongming Qin, Yunjie Lu, Yupeng Jin, Hui Wang\*

*Abstract*—Accurate air pollutant prediction allows effective environment management to reduce the impact of pollution and prevent pollution incidents. Existing studies of air pollutant prediction are mostly interdisciplinary involving environmental science and computer science where the problem is formulated as time series prediction. A prevalent recent approach to time series prediction is the Encoder-Decoder model, which is based on recurrent neural networks (RNN) such as long short-term memory (LSTM), and great potential has been demonstrated. An LSTM network relies on various gate units, but in most existing studies the correlation between gate units is ignored. This correlation is important for establishing the relationship of the random variables in a time series as the stronger is this correlation, the stronger is the relationship between the random variables. In this paper we propose an improved LSTM, named Read-first LSTM or RLSTM for short, which is a more powerful temporal feature extractor than RNN, LSTM and Gated Recurrent Unit (GRU). RLSTM has some useful properties: (1) enables better store and remember capabilities in longer time series and (2) overcomes the problem of dependency between gate units. Since RLSTM is good at long term feature extraction, it is expected to perform well in time series prediction. Therefore, we use RLSTM as the Encoder and LSTM as the Decoder to build an Encoder-Decoder model (EDSModel) for pollutant prediction in this paper. Our experimental results show, for 1 to 24 hours prediction, the proposed prediction model performed well with a root mean square error of 30.218. The effectiveness and superiority of RLSTM and the prediction model have been demonstrated.

Keywords—encoder-decoder model, recurrent neural networks, long short term memory, air pollutant prediction, deep learning, numerical analysis

1. *Introduction*

Air pollution has become an increasingly serious problem and has caused widespread concerns around the world [1]. The prediction of air pollutant concentration, or simply air pollutant prediction, plays a significant role in air pollution prevention and environment management [2], therefore it has received great attention recently in the research community and has been recognized as a key challenge in environment management research.

Traditionally, air pollutant prediction has been cast as a time series modelling problem using relevant historical data including meteorological factors (e.g. humidity and temperature) and other pollutant factors (e.g. PM10 and SO2). Many studies have shown that there are complex interactions between these factors in the formation of air pollution [3-7]. Therefore, the characteristics of such complex interactions must be extracted and used in air pollutant prediction.