

# Deep Recurrent Neural Network による 環境モニタリングデータの予測

Predicting Environment Monitoring Data by Deep Recurrent Neural Networks

杉浦 孔明      Ong Bun Theang      是津 耕司  
Komei Sugiura      Ong Bun Theang      Koji Zettsu

情報通信研究機構

National Institute of Information and Communications Technology

Air pollutants have a considerable impact on human health. It is estimated that with better predictions of fine particulate matter (PM<sub>2.5</sub>), US\$9 billion can be saved over a ten-year period in the U.S. Therefore, it is crucial to keep developing models and systems that can accurately predict the concentration of major air pollutants. In this paper, our target is to predict PM<sub>2.5</sub> concentration in Japan using environmental monitoring data obtained from physical sensors with improved accuracy over the currently employed prediction models. To do so, we propose a Deep Recurrent Neural Network (DRNN) that is enhanced with a novel pre-training method using auto-encoder especially designed for time series prediction. Additionally, sensors selection is performed within DRNN without harming the accuracy of the predictions by taking advantage of the sparsity found in the network. The numerical experiments show that DRNN with our proposed pre-training method is superior than when using a canonical and a state-of-the-art auto-encoder training method when applied to time series prediction. Our experimental results show that our technique improves the accuracy of PM<sub>2.5</sub> concentration level predictions when compared against the PM<sub>2.5</sub> prediction system VENUS.

## 1. はじめに

大気汚染は我々の健康および生活に直結する社会問題である。米国海洋大気庁によれば、大気汚染を防止することで10年間で約9000億円の社会・医療コストの軽減できると試算されている。高度経済成長期には二酸化硫黄による大気汚染が問題となったが、技術および法律による対策により、1990年代以降は汚染が減少している。

一方、近年問題となっている大気汚染に微小粒子状物質PM<sub>2.5</sub>によるものがある。1995年にアメリカにおいて死亡リスクとの関連があるとの報告 [Pope III 95] がなされたことから注目されるようになった。[Chi 12]によれば、PM<sub>2.5</sub>による健康被害により、北京のみで約300億円(2012年)の経済損失があったと試算されている。PM<sub>2.5</sub>の健康への影響については、[Zheng 13, Dergham 11]などが詳しい。

PM<sub>2.5</sub>の対策が各国で始まっているものの、PM<sub>2.5</sub>の予測は簡単ではない。PM<sub>2.5</sub>予測手法の代表的手法は気象モデルを用いるものである [Wakamatsu 13]。一方、日本各地の観測局で得られた環境汚染物質のデータは公開され<sup>\*1</sup>、日々蓄積されていることから、大量のデータを用いて予測精度を向上できる可能性がある。このような背景から、我々はウェブ上のセンサデータや科学データなどのオープンデータを収集と利活用に取り組んでいる [Ong 14, Takeuchi 14]。

本研究ではPM<sub>2.5</sub>の予測問題に対し、Deep Learningに基づく手法を提案する。Deep Learning手法としてDeep Recurrent Neural Network (DRNN)を用いた。また、時系列予測に特化したPre-Training手法を開発した。提案手法はバッチおよびオンラインの予測が可能である。日本の52都市のPM<sub>2.5</sub>データを対象として提案手法と既存手法の比較を行い、予測精度の向上を確認した。入力として、PM<sub>2.5</sub>以外に気温、風速など一般的に気象モデルで用いられている特徴量と同等のものを用いている。実験で用いた2年分のデータは全て公開されているもの

である。

提案手法の独自性は以下である。

- Dynamic Pre-Training (DPT)を導入したDeep Recurrent Neural Network (DRNN)を提案した
- PM<sub>2.5</sub>の予測問題に対し、Deep Learning手法を初めて適用し、既存手法を上回る予測性能を得た

## 2. 関連研究

本節では関連研究と提案手法の類似性および差異について説明する。

一般的な時系列予測問題を扱ったものは非常に多く存在する(例えば [Cheng 06])。予測問題におけるDeep Neural Networkの構造を検討したものに [Crone 11]がある。[Kuremoto 14]では、2つのrestricted Boltzmann machineからなるDeep Belief Networkを用いた時系列予測手法が提案されている。提案手法とこれらの手法の違いは、時系列に特化したPre-Training手法を提案することと、環境モニタリングデータの予測にDeep Recurrent Neural Networkを適用したことである。

これまで種々のDeep Learning手法が提案されているが、大気汚染物質の予測にDeep Learningが適用された例はほとんどない。PM<sub>2.5</sub>予測手法の代表的手法は気象モデルを用いるものである [Wakamatsu 13]。通常のRecurrent Neural Network(RNN)を用いて屋内の大気汚染物質の予測を行ったものとしては [Kim 09]が挙げられる。

Deep Learningにおいて学習データの提示法を検討した研究としては、Curriculum Learningと呼ばれるアプローチがある [Bengio 09]。[Bengio 09]では、学習データをランダムに提示するのではなく、提示順を変更することでDNNの性能を向上できることが示されている。[Bengio 09]では画像処理や言語モデルが議論の中心であるが、他のタスクについても有効であることが示唆される。

連絡先: 杉浦孔明, 京都府相楽郡精華町光台 3-5

\*1 例えば <http://soramame.taiki.go.jp/>

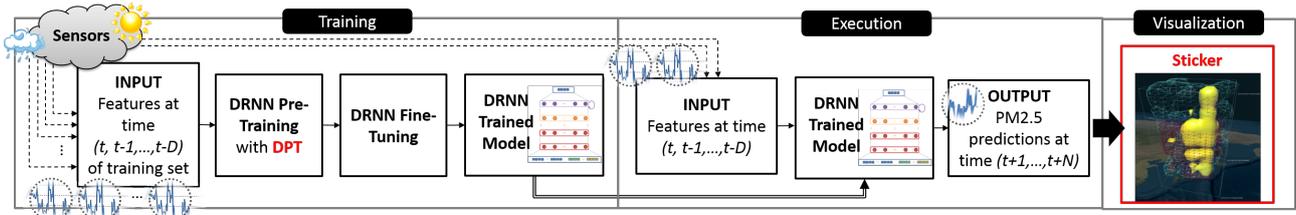


図 1: 提案手法の処理の流れ

オートエンコーダ (AE) を Deep Learning の Pre-Training フェーズで用いる手法が提案されたことで, AE の関連研究が広く行われるようになってきた. 代表的なものとしては, Regularized AE [Ranzato 07] や, Sparse AE [Ranzato 06] が挙げられる. 提案手法の一部である Dynamic Pre-Training(DPT) は, 入力を変更するという点でこれらの手法と異なる.

[Vincent 08] および [Rifai 11] では, Denoising AE(DAE) および Contractive AE が提案されている. 提案手法の DPT は, 入力の再現誤差を変換する意味において DAE および Contractive AE と共通する部分がある. DPT とこれらの手法の違いは, DPT では誤差関数にペナルティ項を必要としないことと, DPT は時系列に特化した変換を用いる点である.

### 3. Dynamic Pre-Training によるオートエンコーダの学習

$L$  時間前までのデータを用いて  $N$  時間後の  $PM_{2.5}$  を予測することを考える. 一般的に時系列予測問題では,  $N$  が大きくなるにつれて予測誤差が増加すると考えることが合理的である. そこで, この特徴を利用して学習サンプルを徐々に変化させ, タスクの複雑さを少しずつ増加させることを考える.

本研究で提案する Dynamic Pre-Training (DPT) は, Pre-Training におけるオートエンコーダの学習を対象とする. いま, 長さ  $D$  の時系列  $\mathbf{x} = \{x_1, \dots, x_D\}$  が得られたとする. 表記の都合上, 特徴量は 1 次元であるものとする. ただし, 実際には多次元の特徴量を扱う.

DPT では, 入力時系列  $\mathbf{x}$  を順序を保ったまま  $\eta$  個の部分時系列に分割する. 分割された  $j(=1, \dots, \eta)$  番目の部分時系列  $\mathbf{z}_j$  は以下で与えられる.

$$\mathbf{z}_j = \{x_k | k = m(j-1) + i; i = 1, \dots, m\} \quad (1)$$

ここに,  $j = 1, \dots, \eta$  であり,  $m(=D/\eta)$  は部分集合の要素数である.

各部分時系列は, 反復回数  $e$  に応じて変化する重要度  $w_j(e) \in [0, 1]$  が割り当てられる. 重要度は  $[0, 1]$  に含まれる実数であるものとする. 重要度を模式的に示したものが図 2 左図である.  $w_j(e)$  は以下のように更新される.

$$w_j(e) = \begin{cases} 1 & \text{if } j < c \\ e/\gamma - j + 1 & \text{if } j = c \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

ここに,  $c = \text{ceiling}(e/\gamma)$ ,  $\gamma = H/\eta$ ,  $H$  は反復回数の最大値である. 上記の  $w_j(e)$  を用いて各部分時系列を重み付けして結合し, 実際の学習に用いるサンプル  $\tilde{\mathbf{x}}(e)$  を作成する.  $\tilde{\mathbf{x}}(e)$  は以下で定義される.

$$\tilde{\mathbf{x}}(e) = \{w_1(e)\mathbf{z}_1, \dots, w_\eta(e)\mathbf{z}_\eta\}. \quad (3)$$

ここに,  $e = 1, \dots, H$  である.

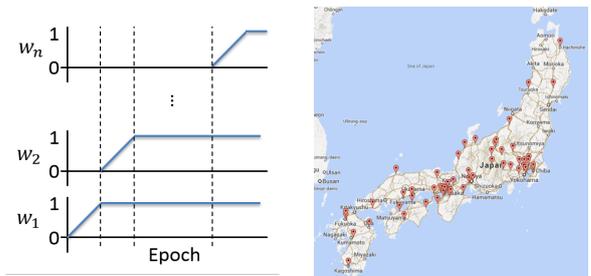


図 2: 左: 反復回数  $e$ (Epoch) と重要度  $w_j$  の関係, 右: 本研究で用いた環境モニタリングデータの観測局の位置

$\mathbf{x}$  から  $\tilde{\mathbf{x}}$  に変換された入力を用いて, Stochastic Gradient Descent により以下の損失関数  $L_{DPT}$  を最小化する学習パラメータを探索する.

$$\begin{aligned} L_{DPT} &= L(\tilde{\mathbf{x}}, g(f(\tilde{\mathbf{x}}))) \\ &= \frac{1}{2DM} \sum_{i=1}^D \sum_{j=1}^M (\tilde{x}_{ij} - g(f(\tilde{x}_{ij})))^2 \end{aligned}$$

ここに,  $f, g$  はそれぞれエンコーダ, デコーダである.

### 4. 実験設定

実験の客観性の観点からは, 比較評価に用いるデータは公開され誰でもアクセスできることが望ましい. そこで本研究では, 国立環境研究所により公開されているデータを用いる<sup>\*2</sup>. 入力特徴量として, 以下の特徴量を用いる.

- $PM_{2.5}$ (対象都市),  $PM_{2.5}$ (近隣都市), 濃度, 風速, 風向, 気温, 日照量, 湿度, 降雨量

入力特徴量の選択には, VENUS で用いられている入力量を参考に選定した. また,  $PM_{2.5}$ (対象都市) は, 予測対象の都市の過去  $D$  時間分のデータであり,  $PM_{2.5}$ (近隣都市) は対象都市近隣の  $K$  個の都市における  $PM_{2.5}$  の濃度である. 用いたデータの例を図 3 に示す.

52 箇所の対象都市は日本各地に分散するように, 県庁所在地を中心に選択した(図 2 右図参照). これらの都市における観測値は公開されており, その中から約 2 年分のデータを用いた. このうち, 学習データ (training set), 検証データ (validation set), テストデータ (test set) の割合を 60%, 20%, 20% とした. 手法のハイパーパラメータは Cross-Validation により自動的に決定した. パラメータの設定値を表 1 にまとめる. データの事前処理として標準的に用いられる外れ値の除去および正規化を行った.

\*2 [http://www.nies.go.jp/igreen/tj\\_down.html](http://www.nies.go.jp/igreen/tj_down.html)

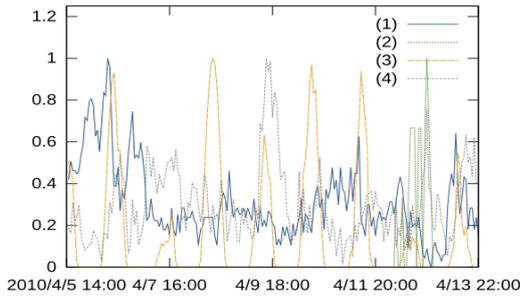


図 3: PM<sub>2.5</sub> を含むセンサ値時系列の例。(1)PM<sub>2.5</sub>, (2)降雨量, (3)日照量, (4)風速. それぞれ区間の最大値で正規化を行っているため縦軸の単位を省略した.

表 1: 実験で用いたパラメータ設定

学習データサイズ	17545 (約 2 年分)
観測周期	1 時間
学習データ	60%
検証データ	20%
テストデータ	20%
最大予測時間 ( $N$ )	12
過去データの長さ ( $D$ )	48
Pre-Training 時の学習率	0.01
Fine-Tuning 時の学習率	0.001
近隣都市数 ( $K$ )	3
最大反復数 ( $H$ )	200
部分時系列数 ( $\eta$ )	25

## 5. 実験結果

### 5.1 Dynamic Pre-Training の評価

まず, 提案手法の一要素である DPT の評価を行う. DPT とベースライン手法の比較結果を表 2 に示す. ベースライン手法として, 通常のオートエンコーダ (AE) と denoising auto encoder (DAE) [Vincent 08] を用いた. 評価の尺度として, 以下で定義される二乗平均平方根 (RMSE) を用いる.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

ここに,  $y_i$  および  $\hat{y}_i$  は, PM<sub>2.5</sub> の真値および予測値を表す. 予測の良好さについては種々の尺度があり得るが, 本論文では RMSE が小さいことを予測精度が高いとみなす.

本実験では, fine-tuning 後の予測精度を比較評価した. 各手法 (AE, DAE, DPT) を用いて pre-training を行った後, 得られた重みを用いて fine-tuning を行った. pre-training のみで入力の再現性を比較することは技術的には可能であるが, 最終的な性能への寄与が不明確であるという欠陥がある. そこで, 表 2 では, Fine-tuning 後の RMSE を比較評価した. 表の結果は, 日本の代表都市 52 箇所の予測値 (12 時間後) の平均を比較したものである. 初期値をランダムに設定して実験を 10 回繰り返して, 平均値を比較した.

DAE はパラメータとして corruption rate ( $v$  とおく) を持つ. 結果の客観性の観点からは  $v$  を手動で設定しないことが望ましい. 本研究では, グリッドサーチにより複数の  $v$  の中から最も良い性能を示した値を選択した. その結果,  $v = 0.2$  となり, これは DAE の文献で推奨される値と同水準であった.

表 2 より, DPT はベースラインと比較して RMSE が小さいことがわかる. また, DAE は AE よりも RMSE が大きい. 画像の分類タスクなどでは DAE の方が良好な結果を示すことが

報告されているが, 本研究で扱うデータにはミスマッチであることが示唆される. 以上より, AE および DAE と比較して, 提案手法である DPT の予測精度が高いことが示された.

Method	RMSE [ $\mu\text{g}/\text{m}^3$ ]
AE	7.11
DAE	7.33
DPT (proposed)	<b>6.92</b>

報告されているが, 本研究で扱うデータにはミスマッチであることが示唆される. 以上より, AE および DAE と比較して, 提案手法である DPT の予測精度が高いことが示された.

### 5.2 性能への影響の検討: ネットワークの構造

本節では, ネットワークの構造が性能に与える影響について検討する. ネットワークの構造の可能性を全て探索することは現実的でないため, 各層のノード数を固定し, 層数を変化させて予測誤差を比較する. 本実験では, 中間層において各層のノード数を 300 とした. 前節と同様に, 予測誤差の尺度として RMSE を用いる.

図 4 に, 層数に対して RMSE をプロットしたものを示す. 図において, 横軸は入出力層を含む層数を示す. 例えば, 図において層数が 4 の場合, 中間層が 2 層あることになる. 各折れ線は, (1)DPT を適用しない Deep Feedforward Neural Network (DFNN), (2)DPT を適用した DFNN (DPT-DFNN), (3)DPT を適用しない Deep Recurrent Neural Network (DRNN), (4)DPT を適用した DRNN (DPT-DRNN), を示す. 実験を 10 回繰り返して, 平均値を比較した.

(1)-(4) のいずれの場合も, ネットワークの層が 6 以上になると予測精度が悪化していくことがわかる. このことは, 本実験に用いたデータ量に対してパラメータ数が多すぎるため, 過学習に陥っていることを示唆している. また, (1) と (2), (3) と (4) を比較すると, DPT を適用した場合に予測精度が高いことがわかる. 紙面の都合上省略するが, 中間層のノード数を減少させた場合は図 4 と同様の傾向が得られ, ノード数が 300 の場合より全体的に性能が悪化した. 詳細については [Ong 14] を参照されたい. これより, 中間層のノード数を 300 とした 5 層の DPT-DRNN が, 比較対象の中で最も良い性能を示すことがわかった.

### 5.3 モデルベース手法との比較

本節では, 提案手法と代表的な PM<sub>2.5</sub> の予測システムの比較評価を行う. ベースラインシステムとして, 国立環境研究所による VENUS を用いる. VENUS は内部的に SPRINTARS [Takemura 00] と呼ばれるモデルを用いている. SPRINTARS は, 大気汚染を引き起こすエアロゾルと気候の関係をシミュレートするモデルである.

前節までは比較尺度として RMSE を用いてきた. 一方, 公開されている VENUS の予測結果は連続値でないため, 2 値分類問題として提案手法との比較評価を行う. 前述の 52 都市に対し, ウェブサイトで公開されている VENUS の予測結果を取得し, 1 日平均値の環境基準である  $35\mu\text{g}/\text{m}^3$  を基準として 2 値化した.  $35\mu\text{g}/\text{m}^3$  という基準は, 環境基本法第 16 条第 1 項において, 人の健康の適切な保護を図るために維持されることが望ましい水準として定められている. DPT-DRNN については, 出力を上記の基準で 2 値化した.

学習データに含まれていないテストデータとして, 2013 年 12 月から 2014 年 2 月のデータを用いた. 12 時間後の PM<sub>2.5</sub> の予測値について, VENUS と DPT-DRNN の予測精度を比較

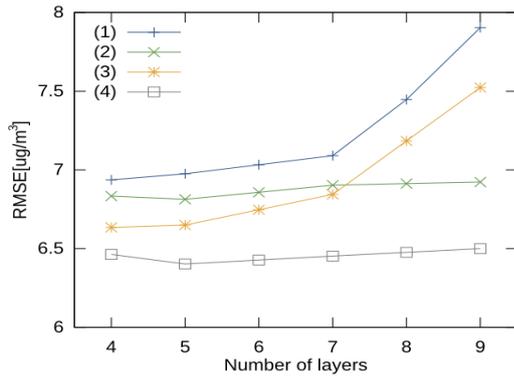


図 4: 層数に対する誤差の比較。(1)DPT を適用しない Deep Feedforward Neural Network(DFNN), (2)DPT を適用した DFNN(DPT-DFNN), (3)DPT を適用しない Deep Recurrent Neural Network(DRNN), (4)DPT を適用した DRNN(DPT-DRNN)。

した。DPT-DRNN の構造およびパラメータは前節の実験の結果を利用した。評価尺度として、適合率 (Precision; P), 再現率 R (Recall; R), F 値 ( $F = \frac{2PR}{P+R}$ ) を用いた。

表 3 に比較評価の結果を示す。表の値は 52 都市の平均を示したものである。表より、適合率は DPT-DRNN が優れ、再現率は VENUS が優れることがわかる。ただし、適合率と再現率の調和平均である F 値について、DPT-DRNN が優れるという結果を得た。

以上の結果は、これまで難しかった都市における  $PM_{2.5}$  の予測について、提案手法により良好な結果が得られることを示している。都市の  $PM_{2.5}$  は社会活動を含む複雑な関係に影響され得るが、提案手法を用いて都市ごとに周辺の都市の状況を取り込んだモデルを別々に学習させることで、既存システムを超える予測精度が達成可能であることを示唆している。

表 3: 気象モデルに基づく既存手法との比較評価。ベースラインシステムの予測結果は連続値として公開されていないため、2 値分類問題として評価した。

Method	Precision	Recall	F 値
VENUS [Takemura 00] (baseline)	0.523	<b>0.653</b>	0.567
DPT-DRNN (proposed)	<b>0.634</b>	0.606	<b>0.615</b>

## 6. おわりに

大気汚染の予測精度を向上させることで社会・医療コストを軽減できる可能性があるが、現状の予測精度は十分ではない。本論文では、微小粒子状物質  $PM_{2.5}$  の予測問題に対し DPT-DRNN を提案し、ベースライン手法との比較評価を行った。日本 52 都市・2 年分のデータを学習させた結果、ベースライン手法と比較して予測精度を向上できることを確認した。さらに、気象モデルに基づく既存手法と比較して、都市部における予測精度が高いことを示した。今後の課題としては、都市部に特化した予測モデルの改良が挙げられる。

## 参考文献

[Bengio 09] Bengio, Y., Louradour, J., Collobert, R., and Weston, J.: Curriculum Learning, in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 41–48 (2009)

[Cheng 06] Cheng, H., Tan, P.-N., Gao, J., and Scripps, J.: Multistep-Ahead Time Series Prediction, in Ng, W.-K., Kitsuregawa, M., Li, J., and Chang, K. eds., *Advances in Knowledge Discovery and Data Mining*, Vol. 3918 of *Lecture Notes in Computer Science*, pp. 765–774, Springer Berlin Heidelberg (2006)

[Chi 12] China: Study on premature deaths reveals health impact of  $PM_{2.5}$ , <http://www.minesandcommunities.org/article.php?a=12062/> (2012)

[Crone 11] Crone, S. F., Hibon, M., and Nikolopoulos, K.: Advances in forecasting with neural networks? Empirical evidence from the {NN3} competition on time series prediction, *International Journal of Forecasting*, Vol. 27, No. 3, pp. 635–660 (2011)

[Dergham 11] Dergham, M., Billet, S., Verdin, A., Courcot, D., Cazier, F., Shirali, P., and Garcon, G.: *Advanced Materials Research*, Vol. 324, chapter Chapter III: Applications, pp. 489–492 (2011)

[Kim 09] Kim, M., Kim, Y., Sung, S., and Yoo, C.: Data-driven prediction model of indoor air quality by the preprocessed recurrent neural networks, in *Proceedings of the International Joint Conference on Instrumentation, Control and Information Technology*, pp. 1688–1692 (2009)

[Kuremoto 14] Kuremoto, T., Kimura, S., Kobayashi, K., and Obayashi, M.: Time series forecasting using a deep belief network with restricted Boltzmann machines, *Neurocomputing*, Vol. 137, pp. 47–56 (2014)

[Ong 14] Ong, B. T., Sugiura, K., and Zetsu, K.: Dynamic pre-training of Deep Recurrent Neural Networks for predicting environmental monitoring data, in *Proc. IEEE Big Data*, pp. 760–765 (2014)

[Pope III 95] Pope III, C. A., Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E., and Heath Jr, C. W.: Particulate air pollution as a predictor of mortality in a prospective study of US adults, *American journal of respiratory and critical care medicine*, Vol. 151, No. 3 pt.1, pp. 669–674 (1995)

[Ranzato 06] Ranzato, M., Poultney, C. S., Chopra, S., and LeCun, Y.: Efficient Learning of Sparse Representations with an Energy-Based Model, in Scholkopf, B., Platt, J., and Hoffman, T. eds., *Advances in Neural Information Processing Systems 19*, pp. 1137–1144, MIT Press (2006)

[Ranzato 07] Ranzato, M., Boureau, Ian Y., and LeCun, Y.: Sparse Feature Learning for Deep Belief Networks, in Platt, J., Koller, D., Singer, Y., and Roweis, S. eds., *Advances in Neural Information Processing Systems 20*, pp. 1185–1192, MIT Press, Cambridge, MA (2007)

[Rifai 11] Rifai, S., Vincent, P., Muller, X., Glorot, X., and Bengio, Y.: Contractive Auto-Encoders: Explicit Invariance During Feature Extraction, in *Proceedings of the Twenty-eight International Conference on Machine Learning* (2011)

[Takemura 00] Takemura, T., Okamoto, H., Maruyama, Y., Numaguti, A., Higurashi, A., and Nakajima, T.: Global three-dimensional simulation of aerosol optical thickness distribution of various origins, *Journal of Geophysical Research: Atmospheres*, Vol. 105, No. 14, pp. 17853–17873 (2000)

[Takeuchi 14] Takeuchi, S., Akahoshi, Y., Ong, B. T., Sugiura, K., and Zetsu, K.: Spatio-temporal Pseudo Relevance Feedback for Large-Scale and Heterogeneous Scientific Repositories, in *Proc. IEEE International Congress on Big Data*, pp. 669–676 (2014)

[Vincent 08] Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A.: Extracting and Composing Robust Features with Denoising Autoencoders, in *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, pp. 1096–1103, New York, NY, USA (2008), ACM

[Wakamatsu 13] Wakamatsu, S., Morikawa, T., and Ito, A.: Air Pollution Trends in Japan between 1970 and 2012 and Impact of Urban Air Pollution Countermeasures, *Asian Journal of Atmospheric Environment*, Vol. 7, No. 4, pp. 177–190 (2013)

[Zheng 13] Zheng, Y., Liu, F., and Hsieh, H.-P.: U-Air: When Urban Air Quality Inference Meets Big Data, in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13*, pp. 1436–1444, New York, NY, USA (2013), ACM