

大気質未計測都市の大気質推定のためのマルチドメイン適応

原田 圭[†] 佐々木勇和[†] 鬼塚 真[†]

[†] 大阪大学大学院情報科学研究科 〒 565-0871 大阪府吹田市山田丘 1-5

E-mail: †{harada.kei,sasaki,onizuka}@ist.osaka-u.ac.jp

あらまし 大気汚染は世界各地の都市で重大な社会問題となっており、大気質観測ステーションを建設する都市が増加している。しかしながら、建設コストや土地価格を主な理由として観測ステーションを建設していない都市（大気質未計測都市）も未だ多く存在する。大気質未計測都市において大気質データを高精度で推定することは、これまで取得不可能であった大気汚染情報を人々に周知させるとともに今後の観測ステーション建設戦略にも有効であると考えられる。そのため、本稿では大気質未計測都市において大気質データを高精度に推定することを目的とし、他都市の観測ステーションで計測したデータを利用してマルチドメイン適応による効果的な学習を行う。マルチドメイン適応ではターゲットドメインに対して各ソースドメイン毎に影響度を学習することで、推定位置と推定に利用する各他都市の関係性を明示的に学習する。中国国内の 20 都市の実データを用いた推定実験において、提案手法は既存手法と比較して推定精度を向上させる。

キーワード 時空間データ分析, 深層学習, Mixture-of-Experts

1. はじめに

大気汚染は世界各地の都市で重大な社会問題となっている。それに伴い、大気汚染情報の周知や大気汚染改善へ向けての施策案出を目的として大気質観測ステーションを建設する都市が増加している。しかしながら、建設コストや土地価格を主な理由として観測ステーションを建設していない都市（大気質未計測都市）も未だ多く存在する。大気質未計測都市において大気質データを高精度に推定することは、これまで取得不可能であった大気汚染情報を人々に周知させるとともに今後の観測ステーションの建設計画にも有効である。本稿では大気質未計測都市において大気質データを高精度に推定することを目的とする。

大気質推定は、観測ステーションを建設している都市内の任意の位置の大気質データを推定することを目的として盛んに研究されている。都市の大気質状態は交通状態や気象情報、土地の利用方法など様々な要因に依存し場所毎に非線形的に変化するため、大気の拡散モデルなどによる古典的な物理シミュレーションでは空間的に細かな粒度で推定することができない [1]。一方、近年では都市の交通状態や気象情報、土地の利用方法といった都市の特徴データと大気質データの関係性を学習する深層学習が高い精度を達成している [2] [3]。

大気質未計測都市の大気質推定では、観測ステーションを持つ他都市で収集されたデータを推定に利用する。推定位置と観測ステーションの位置関係を大域的に捉えれば既存手法を適用することができる。しかしながら、大気質データの非線形的分布を考慮すると物理シミュレーションを適用した場合には細かな粒度で高い推定精度を達成できない。また、既存の深層学習手法を利用する場合には、利用する他都市のデータによって推定精度が変化する。そのため、より高い精度での推定を行うためには推定都市毎に最適な都市を選択する必要がある。例えば、

大気質は流動的であり空間的に近い位置では類似した大気質となる可能性が高いため、地理的に最も近い他都市のデータを選択することが考えられる。しかしながら、地理的に最も近い都市のデータを利用した場合に最も高い精度で推定が可能であることは保証されない。つまり一定の評価軸を用いて推定都市毎の最高推定精度を達成する他都市を選択することは困難であり。また、他都市の全組合せに対して推定精度を評価することは時間コストが非常に高い。そこで、複数の他都市のデータを利用することが考えられるが、既存の深層学習手法は推定対象都市と推定に利用する各都市の関係性を考慮しないため高い精度での推定を達成できない。

以上の課題に対して、複数の他都市のデータを効果的に利用し高い精度で大気質未計測都市の大気質データを推定する深層学習手法を提案する。提案手法では観測ステーションを持つ他都市で収集された大気質データおよび都市の交通状態や気象情報、土地の利用方法などのデータを特徴量として利用する。この際、気象情報などの時間的に変化する特徴量は時系列性を考慮する。また、推定位置から各観測ステーションまでの地理的距離と角度を利用し推定位置と各観測ステーションの空間的な関係性を考慮した推定を行う。さらに、複数の他都市のデータを効果的に利用するために推定モデルの訓練にマルチドメイン適応を利用する。マルチドメイン適応ではターゲットドメインに対して各ソースドメイン毎に影響度を学習することで、大気質未計測都市と推定に利用する各他都市の関係性を明示的に学習する。本稿では中国国内の 20 都市の実データを用いて大気質未計測都市の大気質推定実験を行う。実験結果から、提案手法は単純に複数の他都市のデータを利用した既存手法と比較して高い精度で大気質未計測都市の大気質データを推定可能であることを示す。

本稿の構成は次のとおりである。まず2章で本稿の問題定義について説明する。次に3章で提案手法を導入し、4章で実験について述べる。5章では本稿に関連する研究を説明し、最後に6章にて本稿のまとめを行う。

2. 問題定義

本章では大気質未計測都市の大気質推定の入力と出力を定式化し問題定義を行う。都市 c_k に存在する観測ステーションの集合を $\mathbf{S}_k = \{s_{k1}, \dots, s_{kn}\}$ と表す。ここで、都市 c_k の中心地 l_{c_k} を c_k の位置と呼ぶ。また、観測ステーション s_{ki} は位置 $l_{s_{ki}}$ に設置されている。PM_{2.5}, PM₁₀, NO_x, SO_x といった粒子やガス状物質の大気中における密度を大気質と呼び、特定の気質を仮定したとき、観測ステーションがタイムスタンプ毎に計測した大気質を大気質データと呼ぶ。各観測ステーションは、時間 $\mathbf{T} = \{t_1, t_2, \dots, t_T\}$ 内の各タイムスタンプで大気質データ $s_{ki}[t_j]$ を計測する。都市 c_k の観測ステーション集合 \mathbf{S}_k がタイムスタンプ t で計測した大気質データを $\mathbf{A}_k^t = \{s_{k1}[t], \dots, s_{kn}[t]\}$, および \mathbf{T} 内の大気質データを $\mathbf{A}_k = \{\mathbf{A}_k^1, \dots, \mathbf{A}_k^T\}$ と表す。

本稿では、大気質データを予測するための特徴量として、POIデータ、道路ネットワークデータ、および気象データを用いる。それぞれ以下に定義する。

定義 1 (POI データ) 特定の位置の座標、名称、およびカテゴリ (e.g. 工場) を持つデータを POI データと呼ぶ。すなわち、POI データ p は属性として座標 $p.location$, 名称 $p.name$, およびカテゴリ $p.category$ を持つ。都市 c_k 内の全ての POI データを \mathbf{P}_k と表す。

定義 2 (道路ネットワークデータ) 特定の道路の始点、終点、およびカテゴリ (e.g. 高速道路) を持つデータを道路ネットワークデータと呼ぶ。すなわち、道路ネットワークデータ r は属性として始点 $r.start$, 終点 $r.end$, およびカテゴリ $r.category$ を持つ。都市 c_k 内の全ての道路ネットワークデータを \mathbf{R}_k と表す。

定義 3 (気象データ) 位置 l およびタイムスタンプ t での気象データを $\mathbf{m}_l^t = \{m_1, \dots, m_M\}_l^t$ で表す。ここで、 m は降雨量や風向といった属性を表す。 \mathbf{m}_l^t は都市内の無数の点で定義される。タイムスタンプ t での都市 c_k 内の全ての気象データを \mathbf{M}_k^t と表し、時間 \mathbf{T} 内における都市 c_k 内の全ての気象データを $\mathbf{M}_k = \{\mathbf{M}_k^t\}_{t=1}^T$ と表す。

これらの特徴量は大気質に大きな影響を及ぼす。例えば、POI データにおいては、工場や発電所が多いエリアでは大気質は劣悪となり森林公園の付近では良好となるような傾向があると考えられる。また、自動車が排出するガスは大気汚染の重要な要因であるため、あるエリアの道路ネットワーク構造はそのエリアの大気質状態に影響すると考えられる。さらに、大気質は降雨量や風向きといった気象情報の影響によって大きく変化する。そのため、POI データ、道路ネットワークデータ、および気象

データは大気質推定の特徴量として様々な研究で用いられている [1] [3]。

観測ステーションを建設している都市をソース都市、観測ステーションを建設していない都市 (i.e. 大気質未計測都市) をターゲット都市と呼ぶ。ソース都市の集合 $\mathbf{C} = \{c_1, \dots, c_{|C|}\}$ が観測する大気質データおよび POI などの都市の特徴量を用いてターゲット都市 \tilde{c} 内の任意の点 \tilde{l} の大気質データを推定することを目的とする。以下に本研究の問題定義を示す。

問題定義 複数のソース都市 $\mathbf{C} = \{c_1, \dots, c_{|C|}\}$ に存在するデータ $\{\mathbf{A}_1, \mathbf{P}_1, \mathbf{R}_1, \mathbf{M}_1, \dots, \mathbf{A}_{|C|}, \mathbf{P}_{|C|}, \mathbf{R}_{|C|}, \mathbf{M}_{|C|}\}$, ターゲット都市 \tilde{c} に存在するデータ $\{\mathbf{P}_{\tilde{c}}, \mathbf{R}_{\tilde{c}}, \mathbf{M}_{\tilde{c}}\}$, 時間 \mathbf{T} , および \tilde{c} 内における任意の位置 \tilde{l} が与えられたとき、時間 \mathbf{T} 内の各タイムスタンプにおける \tilde{l} の大気質データを推定する。

3. 提案手法

本章では複数のソース都市から収集された POI データなどの都市の特徴量と大気質データ、およびターゲット都市で収集された都市の特徴量を利用してターゲット都市の任意の位置の大気質データを推定する大気質推定手法を導入する。まず、3.1 節にて提案手法の設計方針について説明する。次に、3.2 節にて提案モデルに入力する特徴量の形状について説明する。続いて、3.3 節にて提案モデルの詳細について述べる。最後に、3.4 節でモデルの訓練方法について説明する。

3.1 設計方針

大気質未計測都市の大気質推定では、観測ステーションを持つ他の都市で収集されたデータを推定に利用する。より高い精度で大気質データを推定するためには推定精度が高くなるような都市を選択する必要がある。例えば、大気質は流動的であり空間的に近い位置では類似した大気質となる可能性が高いため、地理的に最も近い都市のデータを利用することで高い推定精度を達成できる可能性がある。しかしながら、大気質は環境や人々の活動といった都市の特徴にも影響を受けるため、最も近い都市のデータを利用することが大気質未計測都市の大気質推定に最適である保証はない。すなわち、客観的な指標を用いてターゲット都市に対して最高精度で推定を行うことができるソース都市を選択することは困難である。そのため提案手法では複数のソース都市のデータを利用して推定を行う。

大気質推定では、近年、深層学習を利用した手法が物理シミュレーションなどと比較して高い精度を達成している。よって提案モデルにおいても深層学習を利用する。深層学習を利用した大気質推定では都市内の POI データ、道路ネットワークデータ、および気象データといった都市の特徴データと大気質データの関係性を学習する。すなわち推定モデルの訓練時に観測ステーション周辺の特徴データを特徴量、観測ステーションで計測された大気質データを正解値とする。また、周囲の観測ステーションの影響を考慮することで推定精度が向上することが知られている [3]。よって、周囲の観測ステーション周辺の特徴データ、観測ステーションで計測された大気質データ、および推定位置との空間的位置関係の特徴量として考慮する。また複数のソース

都市で収集されたデータを推定に利用する場合、全てのソース都市から得られる特徴量を同等に入力し推定に利用するとターゲット都市に対する各ソース都市の関係性が考慮されず高い精度での推定を達成できない。そこで、ターゲット都市と各ソース都市間の空間的位置関係の特徴量として与えることで、複数のソース都市のデータを効果的に推定モデルの訓練に利用する。

提案モデルでは特徴量として推定位置周辺の特徴データ、観測ステーション周辺の特徴データ、観測ステーションで計測された大気質データ、推定位置と観測ステーション間の空間的位置関係、ターゲット都市とソース都市間の空間的位置関係を入力し、推定位置における大気質データを出力する。大気質データは時間的に周期性や連続性を持って変化するため提案モデルでは時間的に変化する特徴量の時系列性を考慮した推定を行う。また、観測ステーション毎に推定位置に対する影響度を付与することで周囲の観測ステーションの空間的位置関係を考慮する。さらに、ターゲット都市における大気質推定に複数のソース都市の特徴量を効果的に利用するためにマルチドメイン適応技術の一つである Mixture-of-Experts (MoE) フレームワーク [4] を利用する。MoE フレームワークでは各ソースドメイン毎に出力した推定値に影響度を積算し集約することで、ターゲット都市と各ソース都市の関係性を明示的に学習しながらターゲットドメインのラベルデータを推定する。MoE フレームワークでは、各ソースドメイン毎に出力した推定値に積算する影響度を計算するためにターゲットドメインと各ソースドメイン間の関係性を利用する。影響度の学習にはターゲット都市と各ソース都市間の空間的位置関係を用いる。

3.2 特徴量

本節では提案モデルに入力する際の特徴量の形状を説明する。提案手法では、(1) ターゲット都市内の推定位置周辺で抽出される特徴量、(2) 各観測ステーションで抽出される特徴量、および (3) ターゲット都市とソース都市の関係から抽出する特徴量を提案モデルに入力し推定を行う。ここで、各特徴量は位置および時刻毎に提案モデルに入力される。

POI データ、道路ネットワークデータ、および気象データは都市全域で収集されるデータである。提案モデルでは各データを推定位置および観測ステーション周辺の局所的な範囲で抽出し推定に用いる。位置 l および距離の閾値 d が与えられたとき、 $dist(l, l') \leq d$ となるような l' の無限集合を l 周辺の局所的な範囲とし、 $\mathcal{L}(l)$ と表記する。

POI データのカテゴリ集合を C^P とするとき、カテゴリ毎の $\mathcal{L}(l)$ 内に存在する POI データ数を位置 l の POI データ特徴量と呼び、 $\mathbf{X}_l^P = \{x_l^P(c)\}_{c \in C^P}$ と表す。ここで、 $x_l^P(c) = |\{p \in \mathbf{P} | p.location \text{ in } \mathcal{L}(l) \wedge p.category = c\}|$ である。また、 $x_l^P(c)$ は時間 T 内で同一とする。

道路ネットワークデータのカテゴリ集合を C^R とするとき、カテゴリ毎の $\mathcal{L}(l)$ に重なる道路ネットワークデータ数を位置 l の道路ネットワークデータ特徴量と呼び、 $\mathbf{X}_l^R = \{x_l^R(c)\}_{c \in C^R}$ と表す。 l' を $r.start$ から $r.end$ 間 ($r \in \mathbf{R}$) の任意の位置をとるとき、 $x_l^R(c) = |\{r \in \mathbf{R} | l' \text{ in } \mathcal{L}(l) \wedge r.category = c\}|$ である。また、 $x_l^R(c)$ は時間 T 内で同一とする。

タイムスタンプ t における位置 l 周辺の気象データ特徴量として l における気象データ m_l^t を利用する。ただし、気象データの属性の中にはカテゴリデータ (e.g. 天気) が存在するため、それらの属性には one-hot エンコーディングを適用する。タイムスタンプ t における位置 l の気象データ特徴量を \mathbf{X}_l^{Mt} と表す。

大気質推定では各観測ステーションの大気質データ、および推定位置に対する地理的關係性を推定に利用することで推定精度が向上する [3]。よって、観測ステーションで抽出される特徴量として POI データ、道路ネットワークデータ、気象データ、大気質データ、および推定位置からの地理的距離と角度を特徴量として利用する。観測ステーション s がタイムスタンプ t で計測する大気質データを大気質データ特徴量と呼び、 s の位置 l_s を用いて $\mathbf{X}_{l_s}^{At}$ と表す。また、推定位置 \tilde{l} を基準点として観測ステーション s のまでの距離と角度を観測ステーションの地理データ特徴量と呼び、 s の位置 l_s を用いて $\mathbf{X}_{l_s}^S$ と表す。 $\mathbf{X}_{l_s}^S$ は時間 T 内で同一とする。

提案手法では異なる複数の都市で収集されるデータを推定に利用する。ターゲット都市に対する各ソース都市の関係を明示的に学習するために、ターゲット都市とソース都市の地理的關係性を特徴量として用いる。ターゲット都市 \tilde{c} の位置 \tilde{l}_c を基準点としてソース都市 c の位置 l_c までの距離と角度をソース都市の地理データ特徴量と呼び、 c の位置 l_c を用いて $\mathbf{X}_{l_c}^C$ と表す。 $\mathbf{X}_{l_c}^C$ は時間 T 内で同一とする。

以上の特徴量を用いて、(1) ターゲット都市内の推定位置周辺で抽出される特徴量 \mathbf{X}^{local} 、(2) 各観測ステーションで抽出される特徴量 $\mathbf{X}_s^{station}$ 、および (3) ターゲット都市とソース都市の関係から抽出する特徴量 \mathbf{X}_c^{city} をそれぞれ以下のように表す。

$$\begin{aligned}\mathbf{X}^{local} &= \mathbf{X}_l^P \cup \mathbf{X}_l^R \cup \mathbf{X}_l^{Mt} \\ \mathbf{X}_s^{station} &= \mathbf{X}_{l_s}^P \cup \mathbf{X}_{l_s}^R \cup \mathbf{X}_{l_s}^{Mt} \cup \mathbf{X}_{l_s}^{At} \cup \mathbf{X}_{l_s}^S \\ \mathbf{X}_c^{city} &= \mathbf{X}_{l_c}^C\end{aligned}$$

ここで、 t は推定する大気質データのタイムスタンプ、 \tilde{l} は推定位置、 l_s は観測ステーション s の位置、 l_c はソース都市 c の位置を表す。

3.3 提案モデル

提案モデルは複数のソース都市から得られる観測ステーションで抽出される特徴量、およびターゲット都市とソース都市の関係から抽出する特徴量を入力として受け取る。すなわち提案モデルは特定のタイムスタンプ t における推定位置の大気質データ \tilde{y} を推定するために、入力として \mathbf{X}^{local} 、 $\cup_{c_k \in C} \{\cup_{s_{ki} \in S_k} \{\mathbf{X}_{s_{ki}}^{station}\}\}$ 、および $\cup_{c_k \in C} \{\mathbf{X}_{c_k}^{city}\}$ を受け取る。ここで、 C はソース都市の集合、 S_k はソース都市 c_k 内の観測ステーションの集合を表す。提案モデルはエンコード層、Station-based Attention (SA) 層、中間出力層、City-based Attention (CA) 層、および最終出力層から構成される。

提案モデルの概要図を図 1 に示す。エンコード層では入力される特徴量の内 \mathbf{X}^{local} および $\mathbf{X}_s^{station}$ をニューラルネットワークの中間表現に変換する。ここで、 \mathbf{X}^{local} および $\mathbf{X}_s^{station}$ の中で時間的に変化する特徴量データ (e.g. 気象データ特徴量) は LSTM を利用して時系列性を考慮する。他の特徴量は

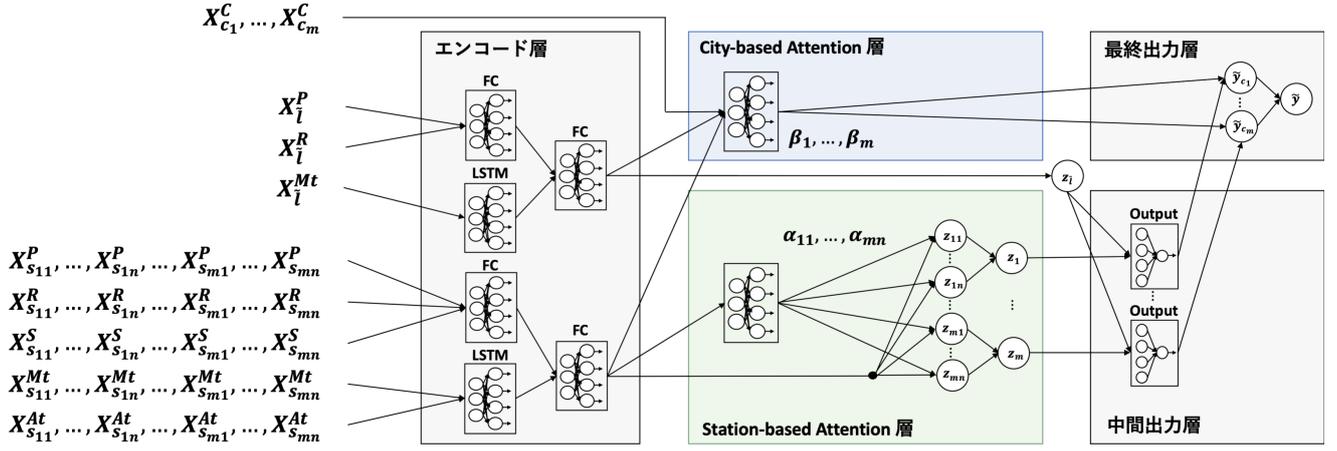


図1 提案モデル

全結合ニューラルネットワーク (FCNN) を利用して中間表現に変換する。LSTM および FCNN によって中間表現に変換された特徴量はさらに FCNN を利用して結合される。それぞれ推定位置の結合された中間表現および観測ステーションの結合された中間表現と呼ぶ。SA 層では、アテンション機構を用いて観測ステーションの結合された中間表現に対して推定位置との空間的位置関係を考慮した影響度を付与する。SA 層は推定位置および観測ステーションの結合された中間表現を入力とした FCNN で実装され、出力する影響度はモデルの訓練と共に学習される。また、SA 層はソース都市毎に実行される。中間出力層では、影響度を付与された各観測ステーションの結合された中間表現と推定位置の結合された中間表現を利用して推定位置の大気質データを出力する。中間出力層は出力次元数を 1 とする FCNN として実装され、ソース都市毎に推定値の出力を行う。CA 層ではアテンション機構を利用してソース都市毎の出力値にターゲット都市に対する影響度を付与する。ここで CA 層は推定位置および観測ステーションの結合された中間表現、加えて X_c^{city} を入力とした FCNN で実装され、出力する影響度はモデルの訓練と共に学習される。最終出力層では影響度を付与されたソース都市毎の出力を合計し、推定位置の最終的な大気質データを出力する。続いて各層の詳細な設計について説明する。

3.3.1 エンコード層

エンコード層では X^{local} および $X_s^{station}$ に対して異なるニューラルネットワークで中間表現に変換する。また、入力される特徴量の中で時間的に変化する特徴量および時間的に変換しない特徴量の変換にはそれぞれ LSTM および FCNN を用いる。LSTM および FCNN を用いて変換された特徴量はさらに FCNN を利用してそれぞれ結合する。 X^{local} 内の気象データ特徴量 X_i^{Mt} に対する中間表現 h_i^t への変換は以下の通りである。

$$\begin{aligned} \mathbf{i}_i^t &= \sigma(\mathbf{W}_{ix} \mathbf{X}_i^{Mt} + \mathbf{W}_{ih} \mathbf{h}_i^{t-1} + \mathbf{W}_{ic} \odot \mathbf{c}_i^{t-1} + \mathbf{b}_i) \\ \mathbf{f}_i^t &= \sigma(\mathbf{W}_{fx} \mathbf{X}_i^{Mt} + \mathbf{W}_{fh} \mathbf{h}_i^{t-1} + \mathbf{W}_{fc} \odot \mathbf{c}_i^{t-1} + \mathbf{b}_f) \\ \mathbf{c}_i^t &= \mathbf{f}_i^t \odot \mathbf{c}_i^{t-1} + \mathbf{i}_i^t \odot \tanh(\mathbf{W}_{cx} \mathbf{X}_i^{Mt} + \mathbf{W}_{ch} \mathbf{h}_i^{t-1} + \mathbf{b}_c) \\ \mathbf{o}_i^t &= \sigma(\mathbf{W}_{ox} \mathbf{X}_i^{Mt} + \mathbf{W}_{oh} \mathbf{h}_i^{t-1} + \mathbf{W}_{oc} \odot \mathbf{c}_i^t + \mathbf{b}_o) \\ \mathbf{h}_i^t &= \mathbf{o}_i^t \odot \tanh(\mathbf{c}_i^t) \end{aligned}$$

ここで、 \mathbf{W} は重み行列、 \mathbf{b} はバイアスペクトル、 \odot はアダマール積をそれぞれ表す。また、 $\mathbf{i}, \mathbf{f}, \mathbf{o}, \mathbf{c}$, および \mathbf{h} はそれぞれ、入力ゲート、忘却ゲート、出力ゲート、メモリセル、および隠れ層の最終状態を表す。推定対象となるタイムスタンプ t から τ ステップ前までの特徴量を時系列的に LSTM へ入力し、 t での LSTM の隠れ層の最終状態 h_i^t を中間表現とする。また、 $X_s^{station}$ に対する中間表現 h_s^t への変換は $X_i^{Mt} \oplus X_i^{At}$ を入力として行う。モデルのスケラビリティおよび汎化性を高めるため全ての観測ステーションで共通の LSTM を利用する。

続いて時間的に変換しない特徴量の変換について説明する。 X^{local} 内の X_i^P および X_i^R に対する中間表現 $z_i^{(n)}$ への変換は以下の通りである。

$$\mathbf{z}_i^{(n)} = \begin{cases} \phi(\mathbf{W}_i^{(n)} (\mathbf{X}_i^P \oplus \mathbf{X}_i^R) + \mathbf{b}_i^{(n)}), n = 1 \\ \phi(\mathbf{W}_i^{(n)} \mathbf{z}_i^{n-1} + \mathbf{b}_i^{(n)}), 1 < n \leq L \end{cases}$$

ここで L は FCNN における隠れ層の層数を表す。LSTM と同様にモデルのスケラビリティおよび汎化性を高めるため全ての観測ステーションで共通の FCNN を利用する。

続いて、LSTM と FCNN でそれぞれ変換された特徴量を入力として、FCNN により以下の通りに結合された中間表現 $\mathbf{z}_*^{(n')}$ を計算する。

$$\mathbf{z}_*^{(n')} = \begin{cases} \phi(\mathbf{W}_{*'}^{(n')} (\mathbf{z}_*^L \oplus \mathbf{h}_*^L) + \mathbf{b}_{*'}^{(n')}), n' = L + 1 \\ \phi(\mathbf{W}_{*'}^{(n')} \mathbf{z}_*^{n'-1} + \mathbf{b}_{*'}^{(n')}), n' \in [L + 2, L + L'] \end{cases}$$

ここで、 $*$ は \tilde{l} または s とする。また、 L' は FCNN における隠れ層の層数を表す。 X^{local} および $X_s^{station}$ に対しては異なる FCNN を利用し、全観測ステーション間では共通の FCNN を利用する。

3.3.2 Station-based Attention (SA) 層

SA 層では各観測ステーションの結合された中間表現に対して影響度を付与する。影響度を付与された中間表現の合計値を各都市内の全観測ステーションの中間表現 \mathbf{z}_k とする。また、影響度の計算はソース都市毎に行う。 \mathbf{z}_k は以下の式で計算される。

$$\mathbf{z}_k = \sum_{s_{ki} \in S_k} a_{ki} \mathbf{z}_{ki}^{(L+L')}$$

ここで、各観測ステーション s_{ki} の影響度 α_{ki} は推定位置および各観測ステーションの結合された中間表現を用いて以下の式で計算される。

$$\alpha'_{ki} = \mathbf{w}_\alpha^T \phi(\mathbf{W}_\alpha(\mathbf{z}_i^{(L+L')} \oplus \mathbf{z}_{ki}^{(L+L')}) + \mathbf{b}_\alpha) + b_\alpha$$

$$\alpha_{ki} = \frac{\exp(\alpha'_{ki})}{\sum_{s_{ki} \in \mathbf{S}_k} \exp(\alpha'_{ki})}$$

3.3.3 中間出力層

中間出力層では \mathbf{z}_k および推定位置の結合された中間表現を利用し各ソース都市毎に大気質データの推定値 \tilde{y}_{c_k} を出力する。なお、中間出力層は推定に利用するソース都市毎に異なるニューラルネットワークを利用し、同一のソース都市に対してはモデルの訓練および推定を通じて同一のニューラルネットワークを利用して出力を行う。

$$\tilde{y}_{c_k} = \mathbf{w}_o^T \phi(\mathbf{W}_o(\mathbf{z}_i^{(L+L')} \oplus \mathbf{z}_k)) + \mathbf{b}_o) + b_o$$

3.3.4 City-based Attention (CA) 層

CA 層では中間出力層で出力された各ソース都市毎の大気質データの推定値にターゲット都市と各ソース都市の地理的距離と角度を考慮した影響度を付与する。各ソース都市毎に付与する影響度 β_{c_k} は以下の式で計算される。

$$\mathbf{z}'_k{}^{(L+L')} = \mathbf{z}_{k1}^{(L+L')} \oplus \dots \oplus \mathbf{z}_{kn}^{(L+L')}$$

$$\beta'_{c_k} = \mathbf{w}_\beta^T \phi(\mathbf{W}_\beta(\mathbf{z}'_i{}^{(L+L')} \oplus \mathbf{z}'_k{}^{(L+L')}) \oplus \mathbf{X}_{c_k}^C) + \mathbf{b}_\beta) + b_\beta$$

$$\beta_{c_k} = \frac{\exp(\beta'_{c_k})}{\sum_{c_k \in \mathbf{C}} \exp(\beta'_{c_k})}$$

3.3.5 最終出力層

最終出力層では CA 層で影響度を付与された各ソース都市毎の推定値を合計し最終的な大気質データの推定値を計算する。推定値 \tilde{y} は以下の式で計算される。

$$\tilde{y} = \sum_{c_k \in \mathbf{C}} \beta_{c_k} y_{c_k}$$

3.4 訓練方法

大気質未計測都市の大気質推定ではターゲット都市に大気質データが存在しないことを仮定するため、推定モデルの訓練ではターゲットドメインの正解値を利用しない教師なし学習を実施する。提案手法ではマルチドメイン適応に対して効果的な訓練方法である meta-training アプローチ [4] を採用する。|C| 個のソースドメインが与えられたとき、ある一つのソースドメインをターゲットドメインみなし残りの |C| - 1 個のソースドメインと合わせて一つの訓練ペアとみなす。訓練時にターゲットドメインとみなすソースドメインおよび残りのソースドメインをそれぞれ meta-target および meta-source と表記する。meta-target は時間 T 内の \mathbf{X}^{local} の集合 \mathbb{X}^{local} を含み、各 meta-source は時間 T 内の $\mathbf{X}_s^{station}$ および \mathbf{X}_c^{city} の集合 $\mathbb{X}_s^{station}$ 、および \mathbb{X}_c^{city} を含む。各訓練ステップでは meta-target 内から \mathbf{X}^{local} 、各 meta-source 内から $\mathbf{X}_s^{station}$ および \mathbf{X}_c^{city} を抽出し推定モデルに入力する。推定モデルが出力した値と選択した meta-target 内の \mathbf{X}^{local} に対応する正解値を利用して誤差を計算し、誤差を推定モデルに逆伝搬することで推定モデルを

訓練する。なお、与えられた |C| 個のソースドメインに対して全ての meta-target と meta-source の組合せを推定モデルの訓練に利用する。また、モデルの評価時にはターゲットドメイン内のデータを \mathbf{X}^{local} とするため、全ての |C| - 1 個のソースドメインの組合せに対して評価を行う。

続いて、推定モデルの訓練時に最小化する目的関数である誤差関数について説明する。提案手法では複数の誤差関数を組み合わせるマルチタスク学習を行う。MoE フレームワークでは逆伝搬する誤差の計算として最終的な出力値と正解値との誤差、各ソースドメインのデータから出力した推定値と正解値との誤差、および全ソースドメインとターゲットドメイン間の分布間不一致を最小とする敵対的学習のための誤差を組み合わせる。最終的な出力 \tilde{y} と正解値 y の誤差の計算には平均二乗誤差を採用し、 \mathcal{L}_{pred} と表記する。

$$\mathcal{L}_{pred} = \frac{1}{N} \sum_{i=1}^N (\tilde{y} - y)^2$$

なお、 N は一度の訓練ステップで推定モデルに入力するデータ数であるバッチサイズを表す。また、各ソース都市毎の出力 \tilde{y}_{c_k} と正解値 y の誤差の計算にも MSE 誤差を採用する。各ソース都市の出力毎に平均二乗誤差を計算しその平均値を推定モデルへ逆伝搬する。以降ではこれを \mathcal{L}_{mean} と表記する。

$$\mathcal{L}_{mean} = \frac{1}{|\mathbf{C}|} \sum_{c_k \in \mathbf{C}} \left(\frac{1}{N} \sum_{i=1}^N (\tilde{y}_{c_k} - y)^2 \right)$$

最後に、全ソース都市とターゲット都市間の特徴量分布不一致を最小化する敵対的学習のために [4] で利用されている MMD (Maximum Mean Discrepancy) 誤差を利用する。以降ではこれを \mathcal{L}_{adv} と表記する。

$$\mathcal{L}_{adv} = MMD^2(\mathbf{z}''_1 \cup \dots \cup \mathbf{z}''_{|\mathbf{C}|}, \tilde{\mathbf{z}}'')$$

$$\mathbf{z}''_k = \mathbf{z}_{k1}^{(L+L')} \cup \dots \cup \mathbf{z}_{kn}^{(L+L')}$$

$$\tilde{\mathbf{z}}'' = \tilde{\mathbf{z}}_1^{(L+L')} \cup \dots \cup \tilde{\mathbf{z}}_n^{(L+L')}$$

$$MMD(\mathcal{S}, \mathcal{T}) = \left\| \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \phi(s) - \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \phi(t) \right\|_{\mathcal{H}}$$

ここで、 $\mathbf{z}_{k1}^{(L+L')}$ はソース都市 c_k における一つの観測ステーションに対するエンコード層の出力結果、 \mathbf{z}''_k はソース都市 c_k 内での集合和を表す。また、 $\tilde{\mathbf{z}}''$ はターゲット都市内における各観測ステーションに対するエンコード層の出力結果の集合和とする。MMD スコアの計算における \mathcal{H} は再生核ヒルベルト空間 (Reproducing Kernel Hilbert Space: RKHS) を表し ϕ は RKHS 空間への写像関数を表す。提案手法では [5] に従い、MMD スコアの計算をカーネル法にて行う。カーネル法では MMD スコアを以下の通りに近似計算する。

$$MMD(\mathcal{S}, \mathcal{T}) = \frac{1}{|\mathcal{S}|(|\mathcal{S}| - 1)} \sum_{a \neq b} \mathcal{K}(s_a, s_b)$$

$$+ \frac{1}{|\mathcal{T}|(|\mathcal{T}| - 1)} \sum_{c \neq d} \mathcal{K}(t_c, t_d)$$

$$- \frac{2}{|\mathcal{S}||\mathcal{T}|} \sum_{a=1}^{|\mathcal{S}|} \sum_{c=1}^{|\mathcal{T}|} \mathcal{K}(s_a, t_c)$$

表 1 各都市内の観測ステーション数

都市	位置	観測ステーション数	PM _{2.5} データ範囲	平均	分散
北京	北部	36	[2.00, 389.00]	71.84	54.72
天津	北部	27	[1.00, 645.00]	66.50	40.31
石家荘市	北部	24	[1.00, 816.00]	94.09	60.58
唐山市	北部	18	[3.00, 617.00]	89.86	54.45
秦皇島市	北部	9	[1.00, 525.00]	60.93	49.64
保定市	北部	27	[1.00, 705.00]	93.91	48.17
張家口市	北部	19	[1.00, 628.00]	36.38	31.57
承德市	北部	14	[1.00, 462.00]	50.72	38.55
滄州市	北部	16	[1.00, 359.00]	72.30	35.66
廊坊市	北部	12	[1.00, 496.00]	84.60	55.57
衡水市	北部	11	[1.00, 1273.00]	91.32	40.72
シ博市	北部	12	[8.00, 275.00]	75.63	26.50
深セン	南部	11	[1.00, 122.00]	18.52	14.04
広州	南部	42	[1.00, 282.00]	34.97	20.44
香港	南部	15	[1.00, 118.00]	15.92	12.16
東莞市	南部	5	[6.00, 120.00]	30.06	15.52
仏山市	南部	8	[2.00, 161.00]	28.49	20.81
惠州市	南部	7	[1.00, 147.00]	23.25	14.87
江門市	南部	7	[1.00, 400.00]	23.48	18.16
汕頭市	南部	6	[1.00, 105.00]	21.79	14.73

ここで \mathcal{K} はカーネル関数を表し、提案手法ではガウシアンカーネルを利用する。ガウシアンカーネルは以下の式で計算される。

$$\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

ただし、 $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ はユークリッド空間上の距離の 2 乗、 σ は標準偏差でありハイパーパラメータとする。また、推定モデルの過学習を防ぐために [4] で利用されている正則化手法を採用する。meta-training では各訓練ステップにおいて meta-target に対する全 meta-source の係数分布 β が計算される。本稿で採用する正則化手法では、 β のエントロピーを計算しその合計値を利用して誤差関数の最小化に対する正則化を行う。

$$\mathcal{R} = \sum_{c_k \in \mathbf{C}} \beta_{c_k} \cdot \log \beta_{c_k} \quad (1)$$

最小化する誤差関数はハイパーパラメータ λ , γ および η を用いて以下のように定義する。

$$\mathcal{L} = \lambda \cdot \mathcal{L}_{pred} + (1 - \lambda) \cdot \mathcal{L}_{mean} + \gamma \cdot \mathcal{L}_{adv} + \eta \cdot \mathcal{R}$$

4. 実験

4.1 データセット

実験では中国国内の 20 都市で 2014/6/1–2014/9/30 の 4ヶ月の間に 1 時間単位で収集された大気質データ、POI データ、道路ネットワークデータ、および気象データを用いて実験を行う。データセット内の都市は中国北東部に位置する都市および中国南部に位置する都市の 2 クラスタに分類できる。続いて、各データ属性の詳細について述べる。

大気質データ: 本実験では推定する大気質として PM_{2.5} を用いる。表 4.1 に各都市が属するクラスタ、各都市に設置されている観測ステーション数、および都市毎の PM_{2.5} データの平均値と分散値をまとめる。データは^(注1)からダウンロード可能である。

(注1) : www.microsoft.com/en-us/research/project/urban-computing/

ある。

POI データ: POI データは各観測ステーションから半径 $d = 1[\text{km}]$ 以内に存在する POI データを Foursquare^(注2) から収集し、{ 芸術 & エンタメ, 学校, イベント, 飲食, ナイトライフ, アウトドア & レクリエーション, 住居, ショップ & サービス, 旅行 & 交通機関, その他 } の 10 カテゴリに分類する。なお、カテゴリ分類は Foursquare の公式カテゴリ^(注3)を参考としている。

道路ネットワークデータ: 道路ネットワークデータは各観測ステーションから半径 $d = 1[\text{km}]$ 以内に存在する道路ネットワークデータを OpenStreetMap^(注4) からダウンロードし、{ 高速道路, 幹線道路, その他 } の 3 カテゴリに分類する。

気象データ: 気象データは各時刻における各観測ステーションの位置に対応する { 天気, 気温, 気圧, 湿度, 風速, 風向 } の 6 属性のデータを利用する。データは^(注1)からダウンロード可能である。

4.2 ターゲット都市の設定および評価手法

次に、実験の評価手法について説明する。本実験ではデータセット内の一つの都市をターゲット都市と仮定し、その他の都市をソース都市と仮定することで推定モデルの精度を検証する。また、ターゲット都市内の推定位置として観測ステーションが存在する位置を仮定し、観測ステーションが計測した大気質データを正解値とする。なお、ターゲットおよび各ソース都市の観測ステーション数の偏りを防ぐために、実験時には各都市からランダムに v 個の観測ステーションを選択し実験データとする。ターゲット都市として中国北部に位置する北京、天津、中国南部に位置する深セン、広州を設定する。また、各ターゲット都市に対してターゲット都市を除く 19 都市をソース都市と設定する。 $|\mathbf{C}|$ 個のソース都市に対して提案モデルは $|\mathbf{C}| - 1$ 個のソース都市を用いて推定を行うため、実験では全ての $|\mathbf{C}| - 1$ 個の組み合わせに対して評価を行い、平均値を計算する。

評価指標として二乗平均平方根誤差 (Root Mean Squared Error: RMSE) を採用する。推定モデルの出力および正解値をそれぞれ \hat{y} および y とするとき、RMSE は以下の式で計算される。

$$RMSE = \sqrt{\frac{\sum_{i=1}^{|\mathbf{T}|} (\hat{y}_i - y_i)^2}{|\mathbf{T}|}}$$

実験では、ターゲット都市からランダムに選択した v 個の観測ステーションの位置を推定対象として、全推定位置における RMSE の平均値を評価スコアとする。

4.3 比較手法

比較手法として、空間的な補完を利用した推定手法および深層学習を利用した推定手法を採用する。空間的な補完を利用した推定手法としては、 K Nearest Neighbor (KNN), Linear Interpolation (LI) を採用する。深層学習を利用した推定手法としては、Feedforward Neural Networks (FNN) および

(注2) : developer.foursquare.com/

(注3) : developer.foursquare.com/docs/resources/categories/

(注4) : www.openstreetmap.org/

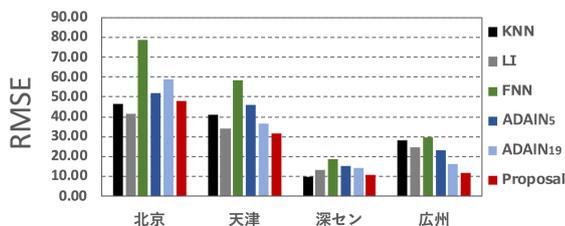


図 2 提案手法と比較手法の推定精度の比較

ADAIN を採用する。以降では各手法の詳細について述べる。

K Nearest Neighbor (KNN) : 推定位置から距離の近い K 個の観測ステーションで計測された大気質データの平均値を推定値とする。実験では $K = 3$ とする。

Linear Interpolation (LI) : 各観測ステーションで観測された大気質データに対して推定位置からの距離の逆数の比で重み付けした値の合計値を推定値とする。

Feedforward Neural Networks (FNN) 特徴量 X^{local} および $X^{station}$ を多層ニューラルネットワークに入力しその出力を推定値とする。実験では FNN の構造として 200 次元の 3 層ニューラルネットと出力層を利用する。なお、時間的に変化する特徴量の時系列性は考慮せず推定するタイムスタンプでの特徴量のみを入力する。

ADAIN: 特徴量 X^{local} および $X^{station}$ を [3] で提案されたモデルに入力しその出力を推定値とする。実験ではターゲット都市に地理的に近い 5 都市をソース都市とする ADAIN₅, およびターゲット都市を除く 19 都市をソース都市とする ADAIN₁₉ を比較手法とする。

4.4 実験結果

本項では評価実験の結果について述べる。図 2 に提案手法と比較手法の推定精度の比較結果を示す。各都市の結果において、棒グラフは左から順に KNN, LI, FNN, ADAIN₅, ADAIN₁₉, および提案手法を表す。実験結果より天津, 広州において提案手法が最高精度を達成しており提案手法の有効性を確認できる。また、提案手法は 4 都市の実験を通して深層学習を用いた比較手法の推定精度を超えている。

また、4 都市の実験を通して FNN および ADAIN は KNN および LI と比較して推定精度が低い。これは深層学習において推定対象とするドメインと異なるドメインのデータを利用して訓練したモデルは精度が低下することに由来する。提案手法は訓練とテストに利用するドメインの相違から発生する推定精度の低下を防ぐことができています。

5. 関連研究

5.1 大気質推定

本節では大気質推定について関連研究をまとめる。これまで研究されてきた大気質推定は大きく二つの手法に分類できる。一つは観測データと経験的仮定に基づき大気汚染物質の物理的拡散プロセスをシミュレーションすることによって未計測エリアの大気質データを推定する手法である [6] [7]。もう一つは都

市から取得できる特徴を利用して大気質データの推定を行う機械学習を利用した手法である。機械学習手法は近年多くの研究が行われており、シミュレーション手法と比較して高い精度を達成する [1]。[8] [9] [10] は線形回帰, 行列分解などの統計モデルを利用して大気質データの推定を行う手法である。これらのモデルは推定するエリアにおける大気汚染物質分布の時空間的な関係性を考慮することなく、推定エリアの局所的な特徴にのみ依存している。[1] はニューラルネットワークを用いた手法であり、Spatial Classifier と Temporal Classifier の二つの分類器を Co-training させることで大気汚染物質分布の時空間的な関係性を考慮する。提案された手法は Web アプリケーション [11] (注5) として公開されている。[2] では強化学習手法の一つである Deep Q-Learning を利用した大気質推定アーキテクチャが提案されており、北京大学が持つ大気汚染センシングシステムに対する最適化が行われている。[3] は学習データの時系列性と観測ステーションの空間的な関係性を考慮した手法である。学習データの内時間的に変化する特徴量は LSTM を利用してニューラルネットワークの中間表現に変換し、時間的に変化しない特徴量は全結合ニューラルネットワークを利用して中間表現に変換する。また、推定エリアからの地理的距離と角度を利用して各観測ステーションの学習データに適切な重みを付与するアテンション機構を利用して空間的な関係性を学習する。[12] は大気質推定をグラフデータの行列補完問題として定式化し取り組んだ研究である。提案された手法では推定エリアと観測ステーションの空間的な関係性をグラフデータとしてモデル化し行列補完手法を利用して大気質データの推定を行う。

本稿で提案する手法は現在 SoTA である [3] を参考にする。学習データをニューラルネットワークの中間表現へと変換する際、時間的に変化する特徴量は LSTM を適用し時間的に変化しない特徴量は全結合ニューラルネットワークを適用する。また、アテンション機構を利用し推定エリアに対して各観測ステーションデータの適切な重みを学習する。

5.2 マルチドメイン適応

ドメイン適応は、あるドメインの学習データを異なるドメインの推定/予測/分類タスク（以降では単にタスクと表記する）に適応させる技術である。学習データのドメインをソースドメイン、タスクを行うドメインをターゲットドメインと呼ぶ。中でも複数のソースドメインデータをターゲットドメインのタスクに適応させる技術をマルチドメイン適応と呼ぶ。また、ドメイン適応はターゲットドメインにラベルデータが存在するか否かで二つに分類される。ターゲットドメインにラベルデータが存在する場合、ソースドメインデータはターゲットドメインのタスクの精度を向上させるために補助的に利用される。具体的にはソースドメインでのタスクの出力を考慮しながら、ターゲットドメインにおけるタスクの出力の決定に利用する。[13] [14] [15] では全てのソースドメインがターゲットドメインに等しく重要であると仮定し補助利用する。また、[16] [17] [18] ではラベルデータを利用してドメイン間の類似度スコアを計算

(注5) : <http://urbanair.msra.cn>

し、各ソースドメインの重要度を決定する。さらに、[19]では各ソースドメインの重要度を決定するアテンション機構をタスクと同時に学習する。一方で、ターゲットドメインにラベルデータが存在しない場合、ソースドメインでのタスクの出力がターゲットドメインでの出力となる。[20][21]ではターゲットドメインの特徴量分布と類似するように各ソースドメインに重み付け、または構造化し、ソースドメインのデータを利用してタスクの出力を計算する。また、[4]では各ソースドメインの重要度を決定するアテンション機構の訓練をモデルの訓練と同時に行う Mixture-of-Experts (MoE) を提案している。このときターゲットドメインのラベルデータは利用せず、全ソースドメインとターゲットドメイン間の特徴量の分布間不一致を最小にする敵対的学習を行う。提案手法は MoE を参考にしており、全ソース都市とターゲット都市の特徴量の分布間不一致を最小にする敵対的学習を行っている。

6. おわりに

本稿では、大気質未計測都市の大気質推定に効果的な深層学習手法を提案した。提案手法は複数の他都市のデータを利用して大気質未計測都市の大気質データを推定する。また、複数の他都市のデータを効果的に利用するために、推定モデルの訓練にマルチドメイン適応の一つである MoE フレームワークを利用する。MoE フレームワークはターゲットドメインに対する各ソースドメイン毎の影響度を学習する手法であり、提案手法ではターゲット都市と各ソース都市間の地理的距離と角度を利用して影響度を学習する。中国国内の 20 都市の実データを用いた評価実験の結果より、提案手法は半数のターゲット都市に対して既存手法を超える推定精度達成した。また、全ターゲット都市に対して既存の深層学習手法を超える推定精度を達成した。

今後の課題は地理的に大きく離れた都市をソース都市として提案手法の推定精度を検証することである。周囲に観測ステーションを持つ都市が存在しないターゲット都市や、海を超えた場所に位置するターゲット都市を仮定して提案手法の有効性を検証する。

謝 辞

本研究は科学研究費 (16H01722) の支援により行われた。ここに記して謝意を表す。

文 献

- [1] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. U-Air: When urban air quality inference meets big data. In *Proceedings of the ACM SIGKDD*, pp. 1436–1444, 2013.
- [2] Zhiwen Hu, Zixuan Bai, Kaigui Bian, Tao Wang, and Lingyang Song. Real-time fine-grained air quality sensing networks in smart city: Design, implementation and optimization. *arXiv preprint arXiv:1810.08514*, 2018.
- [3] Weiyu Cheng, Yanyan Shen, Yanmin Zhu, and Linpeng Huang. A neural attention model for urban air quality inference: Learning the weights of monitoring stations. In *Proceedings of the AAAI*, 2018.
- [4] Jiang Guo and Regina Barzilay. Multi-source domain adaptation with mixture of experts. In *Proceedings of the ACL EMNLP*, pp. 4694–4703, 2018.
- [5] Konstantinos Bousmalis, George Trigeorgis, Nathan Silber-

- man, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Proceedings of the NIPS*, pp. 343–351, 2016.
- [6] N. Kh Arystanbekova. Application of gaussian plume models for air pollution simulation at instantaneous emissions. *The ELSEVIER Mathematics and Computers in Simulation*, Vol. 67, No. 4-5, pp. 451–458, 2004.
- [7] Minjoong J Kim, Rokjin J Park, and Jae Jin Kim. Urban air quality modeling with full O₃-NO_x-VOC chemistry: Implications for O₃ and PM air quality in a street canyon. *The ELSEVIER Atmospheric Environment*, Vol. 47, No. 1, pp. 330–340, 2012.
- [8] Rouzbeh Shad, Mohammad Saadi Mesgari, and Arefeh Shad. Predicting air pollution using fuzzy genetic linear membership kriging in GIS. *The ELSEVIER Computers, Environment and Urban Systems*, Vol. 33, No. 6, pp. 471–481, 2009.
- [9] David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin, Martin Fierz, and Lothar Thiele. Pushing the spatio-temporal resolution limit of urban air pollution maps. In *Proceedings of the IEEE PerCom*, pp. 69–77, 2014.
- [10] Yanan Xu and Yanmin Zhu. When remote sensing data meet ubiquitous urban data: Fine-grained air quality inference. In *Proceedings of the IEEE Big Data*, pp. 1252–1261, 2016.
- [11] Yu Zheng, Xuxu Chen, Qiwei Jin, Yubiao Chen, Xiangyun Qu, Xin Liu, Eric Chang, Wei-Ying Ma, Yong Rui, and Weiwei Sun. A cloud-based knowledge discovery system for monitoring fine-grained air quality. *Microsoft Tech Report*, 2014.
- [12] Tien Huu Do, Duc Minh Nguyen, Evaggelia Tsiligianni, Angel Lopez Aguirre, Valerio Panzica La Manna, Frank Pasveer, Wilfried Philips, and Nikos Deligiannis. Matrix completion with variational graph autoencoders: Application in hyperlocal air quality inference. *arXiv preprint arXiv:1811.01662*, 2018.
- [13] Shoushan Li and Chengqing Zong. Multi-domain sentiment classification. In *Proceedings of the ACL*, pp. 257–260, 2008.
- [14] Ping Luo, Fuzhen Zhuang, Hui Xiong, Yuhong Xiong, and Qing He. Transfer learning from multiple source domains via consensus regularization. In *Proceedings of the ACM CIKM*, pp. 103–112, 2008.
- [15] Koby Crammer, Michael Kearns, and Jennifer Wortman. Learning from multiple sources. *Machine Learning Research*, Vol. 9, No. Aug, pp. 1757–1774, 2008.
- [16] Jun Yang, Rong Yan, and Alexander G Hauptmann. Cross-domain video concept detection using adaptive svms. In *Proceedings of the ACM MM*, pp. 188–197, 2007.
- [17] Lixin Duan, Ivor W Tsang, and Dong Xu. Domain adaptation from multiple sources via auxiliary classifiers. In *Proceedings of the ICML*, pp. 289–296, 2009.
- [18] Mo Yu, Xiaoxiao Guo, Jinfeng Yi, Shiyu Chang, Saloni Potdar, Yu Cheng, Gerald Tesaro, Haoyu Wang, and Bowen Zhou. Diverse few-shot text classification with multiple metrics. In *Proceedings of the NAACL*, pp. 1206–1215, 2018.
- [19] Young-Bum Kim, Karl Stratos, and Dongchan Kim. Domain attention with an ensemble of experts. In *Proceedings of the ACL*, pp. 643–653, 2017.
- [20] Yishay Mansour, Mehryar Mohri, and Afshin Roshtamizadeh. Domain adaptation with multiple sources. In *Proceedings of the NIPS*, pp. 1041–1048, 2009.
- [21] Himanshu S Bhatt, Manjira Sinha, and Shourya Roy. Cross-domain text classification with multiple domains and disparate label sets. In *Proceedings of the ACL*, pp. 1641–1650, 2016.