

15. アジア地域における大気汚染物質の大規模排出源の位置推計に関する研究

安福 一樹^{1*}・倉田 学児・松岡 譲

¹ 京都大学工学研究科都市環境工学専攻(〒615-8540 京都市西京区京都大学桂 C クラスター)

* Email: hcd30303@hcc5.bai.ne.jp

火力発電所を初めとする大気汚染物質の大規模排出源(LPS)が存在する場所は、それが存在しない場所と比べて高濃度の汚染物質による深刻な健康影響が生じている。これを考慮した健康影響の分析のためには LPS のより正確な位置を把握し、その排出量を把握する必要がある。しかし、一般に世界全域をカバーして入手可能な既存のプラント一覧には正確な位置や住所が書かれていない事が多く、書かれていても間違いや表記ゆれなどが多く含まれている。そこで本研究では、情報理論における文字列同士の類似度の概念を用いて間違いや表記ゆれなどを吸収すると同時に、入手可能なあらゆる地名データベースと地方行政区地図を利用して、情報が不確かなプラントデータベースから LPS のより正確な位置を推計する手法を開発した。また、この手法を用いてアジア地域及びアメリカの火力発電所の位置を推計し、既存のプラント位置情報と検証を行い、排出量推計の精度を向上できる可能性を検証した。

この結果、火力発電所の正確な位置情報の存在する日本、アメリカ、中国の一部と検索結果とを検証した結果、20km 以内に発電容量あたり 8~9 割の発電所が一致した。これより既存のプラント一覧と地理空間情報を用いて LPS の位置を 20km 程度の精度で位置推計可能であることが分かった。

Key Words: Large Point Source, Geographic Information System, Levenshtein distance, Air pollution

1. 序論

近年、アジア地域では新興国が急速な発展を遂げており、これに伴いアジア地域ではエネルギー消費量が増大し、大気汚染物質の排出量が増大している。大気汚染は人間の健康に対する主要なリスクの一つであり、多くの都市で SO_x や NO_x などの暴露による健康被害が生じている。

これらの大気汚染物質排出源の中には火力発電所や製鉄所といった、活動に伴い大量に大気汚染物質を排出する大規模排出源(Large Point Source: LPS)が存在する。LPS の周辺では、それ以外の場所と比べて、高濃度の汚染物質による深刻な健康影響が考えられる。この健康影響の差を正確に分析するためには、LPS が存在する場所とその排出量を正確に把握する必要がある。

著者らはこれまで国際機関や業界団体が整備した各種プラント一覧、地名辞典(Gazetteer)や地方行政区地図を元に、アジア地域の LPS の位置推計とその排出量の推計を行ってきた。しかし、アジア地域の新興国や発展途上国も含めた世界全域をカバーして、一般に入手可能なプラント一覧には、正確な位置や住所が書かれていないという問題が存在する。また、プラント一覧

及び地名辞典には誤植、翻字ミス、勘違いなど不確かな情報が多く含まれており、位置推計の際に大きな問題となる。情報が不確かなプラント一覧から LPS の位置を正確に推計し、LPS による排出を大気汚染問題の解決に繋げるためには、これらの不確かな情報を考慮して、できるだけ近い位置を割り出した上、各国の国全体の排出量を元に各 LPS の排出量を推計できるかどうかを検討する必要がある。

本研究では、不確かな情報を考慮した上で、発展途上国を含むアジア地域のなるべく正確な LPS を推計する必要性から、情報理論における編集距離を用いた LPS 位置の推計検索ツールを開発した。ツールの手法はプラント名、所属州、所属都市などのキーワードに対して、地名辞典や地方行政区分地図などの地理空間情報から編集距離を用いて該当する地名の確からしさをスコアとして検索結果を列挙する部分と、キーワードごとに列挙されたスコアと検索結果を元に LPS の位置を総合的に判断する部分に分けられる。本研究では、地理空間情報として NGA の地名辞典や各国地方行政区地図等を持ちいる。

また、各国の排出量とプラント一覧から LPS の排出量が推計できるかどうかの検討も行う。そのために、プラント一覧に含まれているデータの検証を行い、そ

のデータを用いる妥当性の検証も行う。

本研究では、アジア 18 カ国及びアメリカに属する発電容量が 100MW 以上の火力発電所を対象として上記ツールを用いた検索を行い、結果の一部を既往の研究と比較した。また、プラント一覧に含まれているデータとその妥当性の検証を既往の研究を用いて行った。

2. LPS 位置推計検索

検索に用いるプラント一覧として Platts(2009a)及び Platts(2009b)に記載されている発電プラント一覧を使用する。Platts とは世界最大規模のエネルギーや天然資源の商品市況を扱う情報会社で、彼らが出す価格情報は全世界で広く取引参考価格として扱われています。

このデータにはプラント名、国・州・都市などの所在地情報、発電容量などが書かれており、これらはアルファベット表記されている。この中からプラント名、国名、州名、都市名を検索キーワードとする。また、発電プラント一覧のうち、火力発電所かつ発電容量が 100MW 以上の発電所を本研究では LPS として扱う。

LPS の数と総発電容量を表 1 に示す。

表 1 対象とする LPS の数と総発電容量		
	プラント数	総発電容量(TW)
アジア全体	1306	1.01
日本	204	0.17
中国	644	0.44
インド	151	0.12
(アメリカ)	1211	0.77

また、地理空間情報として表 2 及び表 3 に記載したデータを使用した。なお、日本の地理空間情報は地名の読み方が書かれていなかったため、日本郵便(2010)の郵便番号データに記載されている地名読みを使用してローマ字表記に変換した。

表 2 地理空間情報一覧(地名辞典)

情報名	発行者	記数数	精度	メモ
GEONames	NGA	7057742	分まで	世界各国(アメリカ以外)
AllNames	USGS	244954	分まで	アメリカのみ
位置参照情報	国土交通省	86812	大字単位	日本のみ
Wikimapia	Wikimapia	約6000	不明	発電関連を抽出

表 3 地理空間情報一覧(各国行政区地図)

情報名	発行者	使用国
国土数値情報行政区画データ	国土交通省	日本
Geoinfo.China digital 400	Geoinfo	中国
ACASIAN Maps	ACASIAN	マレーシア、ミャンマー、パキスタン、台湾、韓国
ESRI Data & Maps 9.3	ESRI	アメリカ
Agro Maps	FAO	その他

- (1) キーワードを用いた地理空間情報の検索と列挙
最終的なLPSの位置を推計するための判断材料を収

集するために各プラント一覧に含まれる各キーワードを元に地理空間情報一覧から関連する地理空間情報を検索し、列挙する。

地理空間情報一覧からの検索方法及び列挙方法は各プラント一覧に含まれる各キーワードと各種地理空間情報とを照合し、地理空間情報から十分な数の検索結果を優先順位付きで列挙する。優先順位として文字列同士の類似度を表す「距離」を基準として用い、キーワードに「近い」(類似している)検索候補の順位を高くする。

(2) 文字列同士の「距離」の計算方法

文字列同士の「距離」の計算方法として、レーベンシュタイン距離(編集距離)を用いて計測する。レーベンシュタイン距離とは片方の文字列に対して1文字の挿入、削除、置換を複数回行い、もう片方の文字列に変形する際に必要となる最小の編集回数で表される。

編集回数が少ないと二つの文字列が似ているということになり、編集回数が多ければ二つの文字列が似ていないということとなる。

この基準を元に、同じ場所を表す2つの文字列の「距離」は、スペルミスや表記ゆれがあつたとしても、大概が1文字、多くて2文字の差になると考えられ、別の場所を表す地名と比べて「距離」が近くなると考えられる。

(3) 検索結果からの位置推計手法

(1)で列挙する地理空間情報には、地名辞典から検索され、地名、地名を代表する位置(緯度・経度)、レーベンシュタイン距離を含んだ点情報と、地方行政区地図から検索され、地方行政区名、領域を表すポリゴン、レーベンシュタイン距離を含む領域情報の2種類が含まれる。

この各データに位置推計の判断基準となるスコアを付加する。スコアの付加計算式を式 1 に示す。

$$S_{i,k}^j = (SD^j - dist_{i,k}^j \times \gamma^j) \times \zeta_k^j \tag{1}$$

ここで

$S_{i,k}^j$: キーワード k を地理空間情報 j にかけて、列挙され

る i 番目の検索結果に付加されるスコア

SD^j : GIS データベース j に設定した信頼度

$dist_{i,k}^j$: キーワード k を地理空間情報 j にかけて、列挙さ

れる i 番目のレーベンシュタイン距離

γ^j : 地理空間情報 j に設定したレーベンシュタイン距

離の信頼度への影響係数

ζ_k^j : 地理空間情報 j とキーワード k との優先度

次に各領域情報が示す地方行政区内に含まれている全ての点情報のスコアに対して、それぞれの領域情報のスコアを加算する。

最後に点情報のうち、最もスコアが高い点の位置をLPSの位置とする。

3. 位置推計の検証

位置推計の検証は正解と考えられる位置情報を含んだプラント位置データを用意し、それと検索結果とを比較することで行った。

検証は日本、アメリカ、及び中国の一部プラントについて行った。

(1) 検証に用いたプラント一覧

検証には環境省(2005), EPA(2007), Ho *et al.*(2007)に記載されているプラント一覧を用いる。プラント一覧について表4に示す。

表4 検証用プラント一覧

	データの種別	データ数
環境省(2005)	事業種別排出源位置	727
EPA(2007)	発電所一覧及び位置	4998
Ho <i>et al.</i> (2007)	発電所一覧及び位置	52

(2) 検証方法

検証方法は同じ国、州、都市にあり、名称から同一のプラントと判断出来るものについて、両者の位置のズレを計算し、「発電容量基準で、指定したズレ以下に何割のプラントが収まるか」という基準で検証を行った。

なるべく小さい距離でより多くの発電所が収まればこの手法は他の地域でも同じ距離精度で検索可能と考えられる。

4. 結果と考察

(1) 位置推計の比較

a)環境省(2005)との比較

位置推計の結果と環境省(2005)とを比較した。図1にズレと発電容量の分布を、表5にその累積を示す。

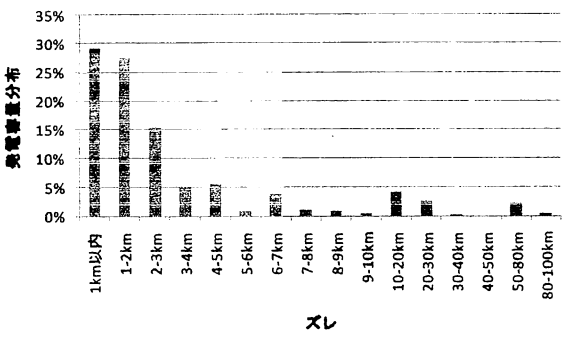


図1 発電容量の分布(環境省(2005)との照合)

表5 発電容量の分布(累積)

ズレ	発電容量
1km以内	29%
2km以内	57%
3km以内	72%
4km以内	77%
5km以内	83%
6km以内	84%
7km以内	87%
8km以内	89%
9km以内	89%
10km以内	90%
20km以内	94%
30km以内	97%
40km以内	97%
50km以内	97%
80km以内	99%
100km以内	99%

b)EPA(2007)との比較

位置推計の結果とEPA(2007)とを比較した。図2にズレと発電容量の分布を、表6にその累積を示す。

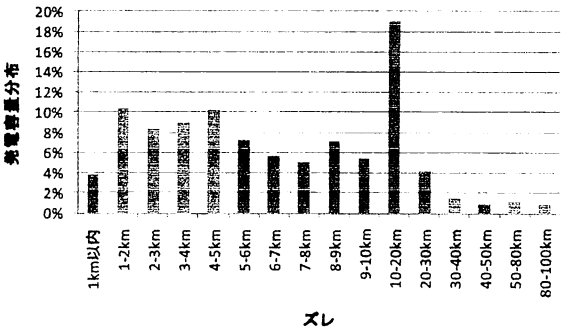


図2 発電容量の分布(EPA との照合)

表 6 発電容量の分布(累積)

ズレ	発電容量
1km以内	3%
2km以内	13%
3km以内	20%
4km以内	28%
5km以内	37%
6km以内	44%
7km以内	49%
8km以内	54%
9km以内	60%
10km以内	65%
20km以内	82%
30km以内	86%
40km以内	87%
50km以内	88%
80km以内	89%
100km以内	90%

c) Ho *et al.* (2007) との比較

位置推計の結果のうち Ho *et al.* (2007) に記載されている発電所について比較を行った。図 3 にズレと発電容量の分布を、表 7 にその累積を示す。

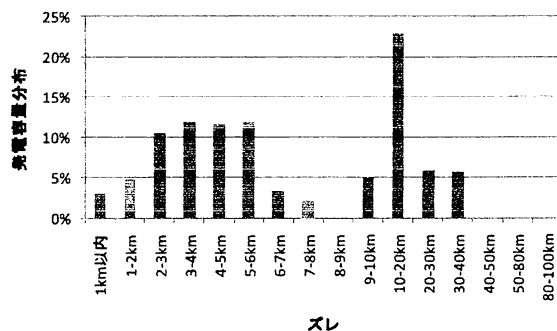
図 3 発電容量の分布(Ho *et al.* との照合)

表 7 発電容量の分布(累積)

ズレ	発電容量
1km以内	3%
2km以内	8%
3km以内	18%
4km以内	30%
5km以内	42%
6km以内	54%
7km以内	57%
8km以内	59%
9km以内	59%
10km以内	64%
20km以内	87%
30km以内	93%
40km以内	99%
50km以内	99%
80km以内	99%
100km以内	99%

(2) 検証

日本では 10km 以内に発電容量あたり 9 割の発電所が、アメリカ及び中国では 20km 以内にそれぞれ 8 割、9 割の発電所がおさまるといった結果になった。これは本研究の位置推計検索では 20km の精度で 8~9 割の LPS を推計できるということになる。

また、日本での位置推計検索の精度が他の 2 国より高い結果となった。これは、地名辞典のうち、位置参照情報の面積あたりの点数が多く、精度が他の地名辞典と比べて高いこと、加えて日本以外の地名辞典に書かれている緯度経度値の精度は分単位までしかないことが上げられる。地名の平均的な距離を表 8 に示す。

表 8 地名辞典の点間隔の平均値

国名	点間隔の平均(km)	文献
中国	5.6	GEONames
アメリカ	6.2	AllNames
日本	2.1	位置参照情報

この距離を上記検証結果の表 5, 6, 7 に当てはめると発電容量はほぼ 50% となる。

また、日本で位置参照情報を用いずに検索を行った時のズレと発電容量の分布を図 4 に示す。

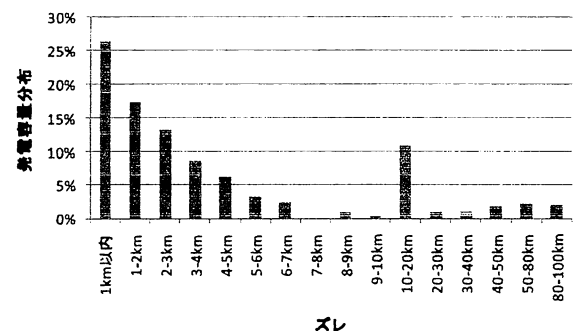


図 4 発電容量の分布(環境省との照合-位置参照情報無し)

位置参照情報を使用しなかった場合の結果は使用した場合(図 1)と比べて 10~20km に図 2, 図 3 と同様のピークが出現した。

位置推計検索の精度が 20km となった理由として、地名辞典の精度及び情報量がボトルネックになっていると考えられる。また、日本の「位置参照情報」と同程度の情報量と精度を持つ地名辞典が各国に存在すれば、10km 程度の精度で位置を推計可能と考えられる。10km よりもさらに精度を向上させるためにはより精度や情報量が高い地名辞典やプラント一覧を用いる、10km 以下は衛星写真を用いて探し出す、大気汚染物質や二酸化炭素の衛星観測データを用いて推計結果の

補正を行う等が考えられる。

表 9 発電容量検証結果

ズレ率	収まった発電容量
1%以内	47%
2%以内	51%
3%以内	55%
4%以内	58%
5%以内	60%
6%以内	63%
7%以内	66%
8%以内	68%
9%以内	70%
10%以内	72%
20%以内	85%
30%以内	89%
40%以内	91%
50%以内	92%

5. プラント一覧を排出分布の作成に用いる妥当性の検証

LPS による健康影響を正確に把握するためには、LPS の位置だけでなく、LPS の排出量も把握する必要がある。そのためには実際の排出量やその時間的推移といったデータが必要となるが、今回用いた発電プラント一覧には各発電所の発電容量や燃料種別は書かれているが、二酸化炭素や大気汚染物質の排出量は書かれていない。著者らは各プラントの排出量を推計する際には国ごとの総排出量を各国エネルギー統計などから推計し、LPS データから得られる発電容量や燃料種別を配分指標として各プラントに割り当てるという手法を行ってきた。しかし、州名都市名に不確かな情報が含まれているのと同様に発電容量及び燃料種別も間違っている可能性がある。

そこで本研究ではアメリカの EPA(2007)に記載されている発電容量と燃料種別を用いて今回用いた発電プラント一覧に記載されている発電容量及び燃料種別の検証を行い、発電容量を配分指標として使う妥当性についても検証を行った。

(1) 検証方法

発電容量の検証方法は発電プラント一覧に記載されている各発電所の発電容量と、EPA(2007)に記載されている発電容量との差の絶対値を、EPA(2007)の発電容量で割り、これを「発電容量のズレ率」と定義した。このズレ率を用い「プラントの発電容量あたり何割が特定の値以下の「ズレ率」に収まる」という基準で検証を行った。

燃料種別の検証方法は「プラントの発電容量あたり何割の燃料が一致した」という基準で検証を行った。

実際の排出量は年間発電量と相関するが、発電容量と相関するかどうかは分からない。そこで、発電容量と年間発電量とを比較した。EPA(2007)には発電容量と実際の年間発電量と、二つを元にして算出された年あたりの稼働率が記載されているため、稼働率に対すると発電容量の分布の関係を調べることで発電容量を配分指標として使用する妥当性を検証した。

(2) 検証結果と考察

a) 発電容量の検証

発電プラント一覧に記載されている各発電所の発電容量と EPA(2007)との発電容量とを比較した。

発電容量の比較結果を表 9 に示す。

「ズレ率」10%以内に 7 割、20%に 8 割 5 分の発電所が収まっており、用いたプラント一覧の発電容量には問題が無いと考えられる。

b) 燃料種別の検証

燃料種別の検証では発電容量基準で 87%の発電所の燃料種別が一致するという結果となった。

これより、用いたプラント一覧の燃料種別には問題が無いと考えられる。

c) 発電容量を配分指標として使用する妥当性の検証
稼働率ごとの発電容量の分布を図 5 に示す。

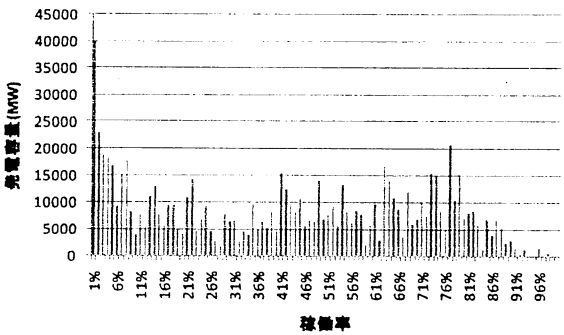


図 5 稼働率ごとの発電容量の分布

稼働率と発電容量には相関は無いと考えられ、発電容量と年間発電量とも相関は無いと考えられる。

よって、発電容量を配分指標として使用する場合には、個々の発電所については、実際の発電量との間に大きなズレが生じる可能性があることが明らかになった。稼働率等の情報が入手できない場合の LPS への配分方法については、今後さらに検討が必要である。

6. 結論と今後の課題

本研究は情報理論で使われる編集距離の考え方を利用した位置推計検索ツールを開発し、アジア地域の情報が不確かなプラントデータベースを対象に位置の推計を行い、推計手法の検証及び排出量推計手法とそのデータの検証を行った。

その結果、検証した3カ国において、20km以内に発電容量あたり8~9割の発電所が含まれる結果となった。

これにより上記の位置検索手法を用いると、LPSの位置が正確に書かれておらず、情報が不確かであった場合でも、LPSのプラント名、国、州、都市の名称情報を元に20kmの精度で位置を推計できることが分かり、20kmメッシュ程度であれば、LPSを考慮した正確な排出分布図を作成可能であることが分かった。また、10km精度など、これ以上の精度による位置推計検索には、より高精度で情報量が多い地名辞典が必要ということも分かった。

一方で、用いたプラント一覧では排出量推計のための妥当な配分指標は無く、発電所一覧を用いて排出量を推計するためには、発電容量以外の必要な情報を含んだプラント一覧が必要であることも分かった。

今後、より正確なLPSの位置推計のためには、より正確でデータ量が多い地名辞典やプラントデータベースか、別のアプローチが必要である。

謝辞: 本研究は、環境省地球環境統合研究推進費BC-088「統合評価モデルを用いた気候変動統合シナリオの作成及び気候変動政策分析」及び科研費(2136254)による研究成果の一部である。ここに記して感謝の意を表す。

参考文献

ACASIAN: ACASIAN Maps, *The Australian Consortium for the Asian Spatial Information and Analysis Network*, 2009
Daniel Jurafsky and James H. Martin: *Speech and Language Processing*, pp.73-76, *Prentice Hall*, 2009
EPA: The Emissions & Generation Resource Integrated Database, *United States Environmental Protection Agency*, 2007
ESRI: *ESRI Data & Maps 9.3*, *ESRI*, 2008
FAO: *Agro Maps*, *Food and Agriculture Organization of the United Nations*, 2003
Geoinfo: *Geoinfo.China digital 400, SuperMap*, 2000
Ho, S. Mun and Chris P. Nielsen: *Clearing the Air The Health and Economic Damages of Air Pollution in China*,

Massachusetts Institute of Technology, pp. B-12, 2007

国土交通省: 街区レベル位置参照情報, 国土交通省, 2006-2008

国土交通省: 国土数値情報行政区域データ, 国土交通省, 2007

環境省: 平成17年度大気汚染物質排出量総合調査, 環境省, 2005

日本郵便: 郵便番号データ, 日本郵便, 2009

NGA: *GEOnet Names Server, National Geospatial-Intelligence Agency*, 2010

Platts: *UDI World Electric Power Plants Data Base, Platts*, 2009a

Platts: *UDI International Electric Power Sourcebook, Platts*, 2009b

USGS: *Allnames, United States Geological Survey*, 2009

Wikimapia: <http://wikimapia.org/>