

ビッグデータと機械学習を用いた
スマート除排雪に関する研究開発経緯

田中 譲

北海道リージョナル・リサーチ

CPS 研究室長

(北海道大学・名誉教授)

この報告書は、北海道リージョナル・リサーチ CPS 研究室が目指す「都市規模ソーシャル CPS 基盤技術の研究開発」に至る経緯を説明するために、平成 30 年 3 月に田中（当時、北海学園大学・客員教授）が札幌市雪対策室に提出した「ビッグデータと機械学習を用いたスマート除排雪の研究」の平成 29 年度報告書の内容をそのまま掲載したものです。

1. ビッグデータと機械学習を用いたスマート除排雪

1. 1 ビッグデータとは

ビッグデータという言葉が市民権を得て久しいが、その定義には未だ決定版がない。**データ量 (Volume)**が従来のソフトウェア・ツールで処理できる範囲を超え、データ入力や分析処理のためのデータ出力の**速度 (Velocity)**が格段に速く、対象のデータ源やデータ型が極めて**多様 (Variety)**であるという3点を意味する3つのVで特徴づけ、「ビッグデータは、高ボリューム、高速度、高バラエティの情報資産のいずれか（あるいは全て）であり、新しい形の処理を必要とし、意思決定の高度化、見識の発見、プロセスの最適化に寄与する」と定義されることが多い。さらに、データの**信憑性 (Veracity)**や、データからの**価値 (Value)**創生を特徴に加えて4Vとか5Vとも呼ばれる。最近では、Volumeの観点からはビッグデータと言えないが、多種多様なデータを扱い、複雑で高度な分析を要する場合もビッグデータ研究に含めている。このような定義では混乱するので、著者は、先端科学技術分野や社会基盤サービスの分野における、**ミッション駆動型(mission-driven)アプローチからデータ駆動型(data-driven)アプローチへのパラダイム・シフト**を象徴する言葉として「ビッグデータ」を捉えている。

今日の社会にあって、我々は様々な複雑かつ難解な課題の解決を迫られている。これらの解決において、従来は、仮説を先ず立て、実験やシミュレーションを通じて仮説検証を行い、これを繰り返して最適解を求め、実対象に適用するという手順がとられた。最初に特定のミッションがあり、その遂行のために情報やデータが収集され分析された。この関係が、今世紀に入ったところから大きく変わってきた。ウェブの発展や計測機器の発達により、取得可能なデータを特定のミッションに関連付けることなく網羅的に自動的に獲得蓄積し、この上で様々なミッションに対して仮説検証可能な多様な新しい課題を設定しその解決を図っていかうという**データ駆動型**の考え方である。

この変化をもたらした象徴的な契機として、以下の5つを挙げることができる。

- a. ウェブの発展とサーチエンジン利用の広がりによるコンテンツと意図データの膨大な蓄積
- b. DNA シーケンシング技術の急発展による自動データ取得と蓄積（これによりバイオ・インフォマティクスが急速に発展した。）
- c. スマート・フォンと車載ナビゲーションシステムの普及によるモビリティ情報の自動取得と蓄積
- d. 物のインターネット (IoT) の発達による実世界モニタリング・データの自動取得と蓄積
- e. 米国のクイズ番組「ジェパディ! (Jeopardy!)」での**IBM Watson**の勝利¹⁾ (Watsonはその後、公表された論文とデータのみを用いて、機能性化合物や新薬の合成法を発見し、

難病の新治療法を発見するという挑戦的課題の研究開発を刺激している。)

ビッグデータ応用の範囲は、ビジネスから、社会基盤サービス、先端科学技術研究開発まで広範である。ビジネスにおいては、様々なサービスを広範なユーザに対して提供することにより、種々のユーザ層の挙動の動向を推定するだけでなく、個々人の嗜好や行動の規則性の分析までも行い、顧客対応サービスの向上や、新しいサービス・ビジネスの展開に応用することが既に活発に行われている。これに対し、現在、今後の大きな発展展開を期待して注目を集め研究開発が盛んになっているのが、社会基盤サービスと先端科学技術研究開発におけるビッグデータ応用である。既にバイオやバイオ・メディスンの分野、スマート・シティの分野では、一部実用レベルでの展開も進んでいる。

1. 2 データ分析の目的と手法

データ分析の目的には、**予測**や**推定**と**記述**の2つがある。対象データ集合は、基本的には表の形式を持ち、表の各行に対応するレコードは単にデータと言われる。各列に対応する属性は変数と言われ、**説明変数**と**目的変数**が区別される。XとYを各々すべての説明変数とすべての目的変数からなるベクトル変数とし、この表を $Y=F(X)$ なる関数Fがある誤差範囲でサンプリングしたものと考える。関数Fを、サンプル点全体を用いた誤差評価（通常は平均2乗誤差）が最小になるように計算機構として求めることが**予測**であり、**回帰推定 (regression estimation)**と呼ばれる。最適なFの探索範囲を線形関数に限定すると**線形回帰 (linear regression)推定**となり、特定のカーネル関数に限定すると**SVR (support-vector regression: サポート・ベクタ回帰)**などの種々の**カーネル回帰(kernel regression)推定**が得られ、多層ニューラル・ネットワークに限定すると**ディープ・ラーニング(deep learning)**が得られる。

回帰推定は、ある誤差範囲で**順問題**を解いたことになる。これに対し、任意のYの値や、Yの最大/最小値に対して、対応するXの値や制約を求めることを**逆問題**という。これは**設計問題**や**最適化問題**に対応する。

Fの計算機構を求めることを**機械学習 (machine learning)**といい、元のデータ集合は**学習データ集合(learning data set)**と呼ばれる。Yが有限個の離散値をとる1変数の場合には、FはXの任意の値をYの異なる値に応じて**分類**する判別関数となる。学習では、Fが学習データをよく近似するだけでなく、データ集合全体に対して仮定される確率分布の下で、未知データもよく近似することが求められる。前者のみを目指すと**過剰適合**

(**overfitting**)になり、未知データの近似が悪くなる。後者を**汎化能力(generalization ability)**といい、**汎化誤差(generalization error)**で評価される。分類には、**決定木**も用いられ、ランダムに構成された多数の決定木を用いた**アンサンブル学習(ensemble learning)**により精度を上げる**ランダム・フォレスト(random forest)**もよく用いられる。

実応用のデータ集合では、説明変数ベクトルXに対し、対象分野の知識を用いて適切な

ベクトル変数 $G(X)$ を定義し、 $Y=F*(G(X))$ となる F^* を機械学習することで回帰推定の精度が顕著に改善されることが多々ある。実応用のデータ集合では、全体が単一の関数 F で近似できることは稀であり、そのまま回帰推定を行っても良い結果は得られないことが多い。回帰推定を行う前に、全体をうまく分割 (segmentation) して、個々の部分集合が別々の単一関数で近似できるようにする必要がある。この分割は、 X の値に関する制約としては記述できずに、この後に述べるクラスタリングやマイニングを適用した結果を用いて、クラスタの指定や、特定パターンを持つデータのみからなる集合として初めて規定できる場合もある。

データ分析のもう一つの目的は記述である。データ集合に隠されている本質的な線形関係を表す低次元の**潜在 (変数の) 空間(latent space)**を求めることや、データのクラスタ構造や**頻出パターン(frequent pattern)**を求めることが目的となる。潜在空間を見つけるには、**特異値分解(singular value decomposition)**や**非負値行列因子分解(non-negative matrix factorization)**等の行列分解が用いられ、後者の2つには各々、**クラスタリング(clustering)**と**パターン・マイニング(pattern mining)**が用いられる。

クラスタリングは、データ間に何らかの**類似性**に基づく距離を定義し、この空間で近接分布しているデータのグループを見つけてグループ分けを行う。予めクラスタ数を指定する**k-means法**や、階層的にクラスタリングを行う**最短距離法**などがある。

パターン・マイニングでは、**アイテム・セット・マイニング(item-set mining)**と**アソシエーション・ルール・マイニング(association-rule mining)**が有名である。パターン・マイニングでは、 Y は考えずに、 X のみを考え、各変数は2値変数と仮定する。そうでない場合は、 X の各変数 x_i に対して、その値域をいくつかの区間に分け、変数 x_i を区間数分の変数 $x_{i1}, x_{i2}, \dots, x_{ik}$ で置き換え、 x_i の値が区間 j に入る時、 x_{ij} のみが1で残りが0となるように定義すればよい。2値化した関係表では、各行を**トランザクション**と呼び、各列を**アイテム**と呼ぶ。これはこの手法が生み出された契機となった課題が、顧客の1回の来店 (トランザクション) 毎の買い物 (アイテム) リストを調べ、同時に購入される傾向が高い買い物の集合 (アイテム・セット) を見つけることであったことに由来する。頻出アイテム・セット A に対して、これと同時に購入される傾向が高い物の集合 B への依存関係 $A \rightarrow B$ を**アソシエーション・ルール**と呼び、頻出ルールを求めるのが**アソシエーション・ルール・マイニング**である。各パターンは、そのパターンを持つデータとそれ以外とを分割する指標の役割も果たす。頻出パターンは系列やグラフにおいても定義できる。**グラフ・マイニング(graph mining)**では、グラフ全体の中で、各所に頻出する部分グラフ構造のマイニングを行う。

学習データのサイズは対象問題により様々である。ディープ・ラーニングでは 10^6 個以上のデータが用いられることが普通であり、説明変数もそれに応じて、 $10^2 \sim 10^3$ 個程度が用いられる。一方、組成原子の異なる無機単結晶材料のデータ集合を第一原理計算により求めるような場合には、学習データ数は精々 10^3 個のオーダーと少ない。このよう場合、独立

な説明変数の数 k が増えるに従い、学習データとなるサンプル点の k 次元空間での分布密度は急激に低くなり（データの**スパース性**という）、回帰推定の近似が悪くなる。これを「**次元の呪い**」という。この場合の独立な説明変数の数は精々 5～7 程度に抑えるべきである。少ないサンプル点で精度の良い学習を行うには、目的に応じて適切にサンプル点を選びながらサンプル数を徐々に増やす手法も用いられる。**ベイズ最適化 (Bayesian optimization)** や、**モンテカルロ木探索 (Monte Carlo tree search)** などである。これらを用いる場合も、対象データ集合が不均質な場合には、予め均質な部分集合にセグメンテーションを行っておくことが重要である。

ディープ・ラーニングは画像認識での成功が契機となり、種々の分野に応用されるようになった。これを用いた**アルファ碁**がプロ棋士に勝利したことは大きなインパクトを与え、Watsonのインパクトと共に、第3次人工知能ブームをもたらした。ディープ・ラーニングを適用する際の大きな課題は、学習データに膨大なデータが必要となることである。この解決策として、シミュレーションによって生成されたデータが学習に利用される。

1. 3 シミュレーション技術の発達とデータ同化

スーパーコンピュータとそれを用いた高性能シミュレーション技術の発展は、種々の物理現象の高精度なシミュレーションを可能にしつつある。近年特に注目すべきは、気象シミュレーションにおける、地球全体をも対象とする大規模化と精度の向上であり、材料物性における**密度汎関数理論**を用いた材料物性シミュレーションの顕著な発展である。

これらは、種々の物理現象を精緻な数理モデルによりモデリングし、高性能計算機を用いて高精度計算を行うことにより、実験計測や観測を通じて得られるデータを計算で求められるようにしつつある。しかしここには大きな課題がある。シミュレーションのためのモデリングには種々の変数が含まれるが、これらの中には初期値や境界値が計測や観測を通じては求められない変数が存在するために、偏微分方程式の解がユニークに定まらないことである。 X を観測や計測が不可能なこれらの変数も含めたシミュレーションの全変数を含むベクトル変数とし、 Y をシミュレーションにより求めたい特性を表すベクトル変数とする。シミュレーションを予測に用いるには、以下のように考えればよい。 X の時刻 $t-1$ における値の確率分布 $P(X_{t-1})$ を想定し、これを事前確率と考える。 X の種々の値に対する Y の t における値を求めるアンサンブル・シミュレーションによって条件確率 $P(Y_t | X_{t-1})$ が求まる。 Y の確率分布 $P(Y_t)$ を事後確率と見做すと、 $P(X_{t-1}, Y_t) = P(Y_t | X_{t-1}) P(X_{t-1}) = P(X_{t-1} | Y_t) P(Y_t)$ となり、 $P(X_{t-1} | Y_t) = P(Y_t | X_{t-1}) P(X_{t-1}) / P(Y_t) \propto P(Y_t | X_{t-1}) P(X_{t-1})$ が成り立つ。これを用いると、 Y の t での観測値を用いて、 Y_t がこの値を取る時の $P(X_{t-1} | Y_t)$ を求めることができる。時刻 t での観測結果 Y_t を反映させてこのシミュレーションを続行するには、 X_{t-1} の分布にこの分布 $P(X_{t-1} | Y_t)$ を使ってアンサンブル・シミュレーションを $t+1$ まで進め、 Y の $t+1$ での値 Y_{t+1} の確率分布を求めて Y_{t+1} の予測とすればよい。これを**データ同化**

(data assimilation) という。

離散系に対しても、人流や交通流のシミュレーション技術が発展してきている。しかし、これを車両から得られる実時間のプローブ・カー・データや、スマート・フォンから得られる実時間モビリティ・データと、都市規模の広域でデータ同化して、予測の精度向上を図る試みはまだ成功していない。高精度の交通予測や避難誘導の実現には必須の技術であり、今後の研究開発が望まれる。

1. 4 基盤技術の現状と課題

ビッグデータに対する基盤技術の発展も賑やかである。大規模データの管理検索のために、**コラム型データベースやnoSQL型データベース等の新技術**が出現し、データ型に関しても、**半構造型データや、データストリーム型データ、グラフ構造データ、非構造データ**も扱えるように拡張されてきた。各種データ型に対する多様なアルゴリズムが開発され、処理速度の改善も顕著である。データの入力、補正、管理、検索、分析、可視化のあらゆるサービスを**クラウド・システム**で提供し、システム開発・運用・維持のコスト削減を図る技術も実用化されている。

一方で、多くの研究者や実務家が、基盤技術と応用の間に大きなギャップが存在していることを指摘している。実応用における課題解決では、どのような検索と分析をどのような順に組み合わせれば有益な知見を得られるのかという問いに対して、普遍的答えはまだない。現時点では、データ・サイエンスは経験の積み重ねに負っており、欧米では人材教育に重点がおかれている。

分析シナリオが定まれば、**ワークフロー・システム**を用いて各種ツールを組み合わせることで実行させることは可能であるが、**分析シナリオを見つけること自体が研究対象であるような応用が現実には多く、分割と分析を試行錯誤的に繰り返しながら最適な分析シナリオを見つける探索的可視化分析過程**を支援できるシステム環境技術の研究開発が望まれている²⁾。

1. 5 ビッグデータ応用のメガトレンド

ここでは、**社会基盤サービスと先端科学技術研究開発**でのビッグデータ応用のメガトレンドを述べる。

社会基盤サービスには物理世界におけるものと、サイバー世界におけるものがある。後者では、サービス・トランザクションはすべて記録され、高度な分析の対象となる。前者においても、物理世界の状況を多様なセンサを通じて実時間でモニタリングし、サイバー世界に送ってそこで状況の把握と分析や最適化を行った後、アクチュエータや情報サービスを介して、直接的な制御や間接的な作業指示の形で、物理世界にフィードバックすることが考えられる。物理世界とサイバー世界をループで結合し、物理世界における社会基

盤サービスの効率化、最適化を図るシステムがソーシャルCPS(social cyber-physical system)として研究されてきた。

ソーシャルCPSは**社会のOS**と見做される。このような考えの源泉は20数年前まで遡ることができる。1991年に**David Gelernter**は**Mirror World**を出版し、都市や病院のような物理世界におけるあらゆる事と物、そして人々をそっくりサイバー世界で模倣し、前者の状態変化をモニタリングすることにより実時間で後者に反映し、後者でのモデルの挙動を分析することにより、後者の中で種々の最適制御を行い、その結果を前者に実時間でフィードバックすることにより、物理世界を間接的に最適制御するシステムの構想を提案した³⁾。**ミラーワールド**とは、サイバー世界の中で構築された物理世界のモデルである。

現在、都市のスマート化を目指した研究や、自動運転車の研究が進展している。自動運転車の開発では、個々の車のインテリジェント化だけでは解決しがたい課題も見え始め、道路などの基盤側にも何らかのインテリジェンスを埋め込む必要性が認識され始めている。これらの究極の姿はミラーワールドの構想に見ることができる。インフラ構造物の老朽化により維持管理技術の高度化が望まれているが、その究極の姿もミラーワールドの構想に見ることができ、サイバー世界では、モニタリング・データの分析とシミュレーションを組み合わせ、最適な維持管理プランが作成され、物理世界にフィードバックされる。

一方、先端科学技術研究では、種々の‘X’サイエンスが、‘X’インフォマティクスへとパラダイム・シフトを始めている。最初の変革はバイオ・サイエンスにおいて起こり、バイオメディカル・サイエンスがこれに続いた。気象、地震、宇宙、災害、感染症の各科学分野でもこの変革が起こりつつあり、現在、変革の真っ只中にあるのが材料物性、触媒、創薬、農業の科学分野である。

今後、社会基盤においては、IoTが発展し、物理世界におけるあらゆる物がインテリジェントになると共に、道路や建築物などのあらゆる社会基盤が常時モニタリングされて遠隔制御可能になり、その中で生活する人々もまた挙動や健康状態が常時モニタリングされるようになる。これらは相互に通信し合うと共に、これらすべての状態を実時間で反映した**ミラーワールド**がサイバー世界に構築され、今後起こりうる種々の状況をシミュレーションしながら、高度な分析によって予測を立て、最適な制御や誘導が物理世界に対してなされることになるであろう。このシステムは多層化かつネットワーク化され、行政によって管理されるものから、民間企業によって管理されるもの、個人によって管理されるものまで、個々に保全性を保ちながら、相互に連携できる仕組みが構築されるであろう。これからのスマート除排雪もこのような動向の中でシステム設計をしていく必要がある。

科学技術研究の分野では‘X’インフォマティクスへのパラダイム・シフトがあらゆる分野で進み、計測・観測機器や生産プロセスはセンサとアクチュエータを用いて完全にコンピュータ制御され、実験やシミュレーションの設定とデータ取得、仮説設定、データ分析による仮説検証、文献から抽出された知識ベースへの問い合わせ、報告書作成までのすべてがコンピュータ端末を通じて行われるようになり、**イン・シリコ・サイエンス (in silico**

science) の時代になる。あらゆるものがデジタル表現を介して扱えるので、コピーや出版・流通、再利用が可能となり、オープン・サイエンスが促進される。

1. 6 社会基盤サービスにおけるビッグデータ・アプローチ

スマート除排雪などの社会基盤サービスに関しても、道路と交通の状況、気象の状況を把握するために多数のセンサや気象レーダ、さらにはリモートセンシングなどを用い、大規模な観測・計測データを取得蓄積し、これらを用いた分析により、種々の問題解決に科学的な手法で取り組もうとする流れが形成されつつある。現状では、都市間を繋ぐ高速道路や国道などのルート上の除排雪や、特定のバス路線上の除排雪を対象とする取り組みが先行しており、都市全体をカバーする道路網全体の除排雪を対象とする取り組み例は国内外ともに極めて少ない。

このようなシステムにおけるアプローチ法は以下のようなになる。これらのモニタ・データは実時間で取得され、過去データとして蓄積される。対象のモデリング $Y=F(X)$ の機械学習において、 X にはその値を制御可能な変数と、そうでない観測のみ可能な変数が含まれる。前者を制御変数、後者を非制御変数と呼ぶことにする。除雪や排雪の実施に関するパラメータや、速度制限等は制御変数であり、気温や降水量は非制御変数である。前者に対応するベクトル変数を X_1 とし、後者のそれを X_2 とし、 $Y \approx F(X_1, X_2)$ と表す。 X には、現在値のみを用いる場合、現在値と共に昨日の降水量のような過去値を用いる場合、現在値と過去値に加えて明日の降水量のような未来値も用いる場合がある。推定すべき Y が未来値である場合は最後のケースになる。

以上を考慮すると、データ駆動型社会基盤サービスが目指す技術は以下の5種類に分類することができると考えられる。

- (1) 現状可視化分析
- (2) 状況対応知識発見
- (3) 現状対応意思決定支援
- (4) 想定状況対応計画支援
- (5) 状況予測対応意思決定支援

(1) は X と Y の現在の値のみを用い、現状の分析と可視化を行う。(2) は X と Y の現在値と過去値を用い、 $Y \approx F(X_1, X_2)$ を機械学習(回帰分析)により推定する。非制御変数 X_2 がどのような値の場合に、制御変数 X_1 に対してどのような値を設定すると、結果として Y の値がどうなったかを学習する。

(3) は(2)の推定結果 $Y \approx F(X_1, X_2)$ と、現在値と過去値を含みうる非制御変数 X_2 の値を用いて、 Y に対する目標制約を満たす制御変数 X_1 の値ないしは制約条件を求める。これは制御問題を解いて意思決定支援を行うことに相当する。

(4) では同様に回帰推定の結果 $Y \approx F(X_1, X_2)$ を用い、非制御変数 X_2 の想定可能なあら

ゆる値 v_2 に対して、 $Y \approx F(X_1, v_2)$ を求める。これを用いると、想定範囲内の各々のシナリオ $X_2=v_2$ 毎に、 Y に対する目標制約を満たす制御変数 X_1 の値ないしは制約条件を求めることができる。これはハザードマップの作成や、制御計画の策定支援に相当する。

(5) では、未来の値を含む非制御変数 X_2 の予測 v_2 をアンサンブル・シミュレーションと観測データのデータ同化などを用いて精度良く行い、(2) の結果を用いて $Y \approx F(X_1, v_2)$ を求め、制御変数 X_1 の種々の値に対して結果状況 Y を予測すると共に、 Y に対する目標制約を満たす X_1 の値ないしは制約条件を求める。これは気象等の制御不能な変数の将来予測を行い、これに基づいて対策のための制御問題を予め解いて意思決定を支援することに相当する。

(1) は監視システムや市民向け情報サービスとして用いられ、(2) は分析研究に用いられている。(3) は、将来予測を用いない監視指令／監視制御システムに用いられ、(4) は監視指令／監視制御の計画作成に用いられる。(5) は今後の気象変化などの将来予測を考慮した監視指令／監視制御システムに用いることができる。これらのシステムの実現はこの順に難しくなる。

X_2 には気象に関する変数だけでなく、例えば川の流量や土砂崩れのプロパティ等のように、その値の予測のためには、気象の変化に応じて別の数式を解くことや、シミュレーションを行うことが必要となるような変数も含まれることがあり、技術開発を一層困難にしている。

スマート除排雪のような社会基盤サービスに関するビッグデータの活用は、上述の(1)から(5)の順に拡大していると考えられる。(1)に関しては十分な台数のセンサが設置され、実時間モニタリングが可能になれば、可視化・分析には技術的困難はない。(2)に関しても過去データが十分に蓄積されれば、正常状況の範囲では問題なく実現できる。災害などの異常状況をも考慮した推定が可能となるためには、学習データの中に異常状況のデータを十分に多く含める必要がある。この課題の解決には、 F の機械学習に用いるデータの内、異常時の X と Y の関係に関しては、少ない観測データを補うために異常状況のシミュレーションを用いることが考えられる。例えば、水害の場合には、洪水のシミュレーションは地形情報を用いて精度良くなされるが、 Y に経済損失額や人身被害総数などが含まれる場合には、洪水シミュレーション結果から住宅地図情報や土地利用情報などを用いて損失推定を行うシミュレーションも併せて行う必要がある。(2)の実現性は、これらのシミュレータの開発の容易性に依存する。(3)は、(2)が実現できれば実現できる。(4)の実現には、(3)の実現が可能になることに加え、非制御変数 X_2 の想定可能な妥当な値域を決定できなければならない。(5)は、制御可能でない要素の将来の値を含む X_2 の予測値を精度よく求めることが必要である。これにはアンサンブル・シミュレーションと観測値のデータ同化を常時繰り返し行うことによって予測精度を常時保つ高度な技術の確立が必要である。このような技術は、理研で開発されているゲリラ豪雨予測技術に見られるように、現時点では京コンピュータのような最速のスーパーコンピュータを用いる必要がある。

1. 7 スマート除排雪へのビッグデータ・アプローチ

スマート除排雪は社会基盤サービスの一つと考えられ、2. 6で述べたような方針で、ビッグデータに基づく問題解決を図ることが将来の大きな方向性になると考えられる。

スマート除排雪に関しても、2. 6で述べたように、以下の5種類の技術が考えられる。

- (1) 現状可視化分析
- (2) 状況対応知識発見
- (3) 現状対応意思決定支援
- (4) 想定状況対応計画支援
- (5) 状況予測対応意思決定支援

最初の現状可視化分析は、データに基づき、エビデンスに基づいて道路と交通の現状を可視化提示したり、過去の除排雪作業の効果をエビデンスに基づいて評価し可視化提示することに対応する。

(2)の状況対応知識発見では、Yとして道路状況、交通状況、除排雪パラメータなどの推定を行うことが考えられるが、交通状況に関しては、プローブ・カー・データが実時間で入手可能な場合には推定の必要性がない。道路状況の推定に関しては、気象および交通データと、ABSの起動データから計算される道路の摩擦係数との間の関係を過去データを用いた機械学習により求め、現在の気象および交通データから、現在の道路の摩擦係数を推定することが考えられる。もしも、ABSデータが実時間で入手可能な場合には、直接に実時間ABSデータから摩擦係数が計算できるので機械学習は不要となるが、現状では、ABSデータの実時間入手はまだ困難な状況である。除排雪パラメータの推定に関しては、気象および交通データと、実施された除排雪のパラメータの間の関係を過去データを用いた機械学習により求め、現在の気象および交通データから、実施すべき除排雪のパラメータの推定を行うことが考えられるが、利用できる除排雪パラメータの過去データは、必ずしも最適な除排雪作業のパラメータにはなっていないので、これらに基づいて機械学習を行った結果は、最適な除排雪パラメータの推定値を与えることはできない。

除排雪作業のパラメータは、変数Yとして扱うのではなく、むしろ制御可能変数 X_1 の一部として扱い、気象、交通、除排雪実施のパラメータと、翌日の交通状況Yとの間の関係を過去データによって機械学習して $Y \approx F(X_1, X_2)$ を求め、渋滞が生じないYの値の範囲から、気象などに関する X_2 の現在値に対して、この関係を満たす変数 X_1 の値に対する制約条件を求める(3)の現状対応意思決定支援の問題として取り扱うべきである。残念ながら、この逆問題としての定式化による最適な除排雪パラメータの推定技術はまだ確立できていない。

(4)の想定状況対応計画支援に関しても、想定される多様な気象状況と交通状況の下での最適な除排雪パラメータの値の推定という逆問題としての定式化による解放は、現状ではまだ確立できていない。

(5) の状況予測対応意思決定支援では、未来の気象などの値を含む非制御変数 X_2 の t 時間後の予測 v_2 をアンサンブル・シミュレーションと観測データのデータ同化などを用いて精度良く行い、(2) の結果を用いて $Y \approx F(X_1, v_2)$ を求め、制御変数 X_1 の種々の値に対して結果状況 Y を予測すると共に、 Y に対する目標制約を満たす X_1 の値ないしは制約条件を求める。予測として用いる場合には、 Y には t 時間後の道路の摩擦係数や、交通状況などが考えられる。最適制御に用いる場合には、 Y は t 時間後の交通状況となり、求めるべき X_1 の一部として実施すべき除排雪のパラメータ値に対する制約が得られることになる。しかしながら、現状の技術水準では、 t 時間後の気象予測の精度が実用レベルからは程遠い。除排雪の意思決定では、1 回の実施毎にその成否により数億円から数十億円の損益の差が生じ、予測精度が高くないと実用性はない。

参考文献

- 1) Baker, S., 2012, Final Jeopardy: The Story of Watson, the Computer That Will Transform Our World, Mariner Books, Boston, MA, USA.
- 2) Tanaka, Y., 2014, Meme Media and Knowledge Federation for Exploratory Visual Analytics of Big Data, DASFAA Workshops, Bali, Indonesia, April, pp. 3-17.
- 3) Gelernter, D., 1991, Mirror Worlds: Or the Day Software Puts the Universe in a Shoebox: How It Will Happen and What It Will Mean, Oxford University Press, USA.

2. スマート除排雪実証実験

2章で述べたように、著者らは、ソーシャル CPS の考えに基づき、札幌市全域における、気象、交通状況、道路状況、除排雪実施状況などを実時間でモニタリングすることにより、これらのビッグデータを蓄積した過去データを用いた機械学習の結果を用いて、現在の道路状況の推定や、最適な除排雪作業に対する意思決定支援を行うことを、スマート除排雪と呼んで実証研究を行ってきた。本章では、これらの実証研究を報告し、そこから得られた現時点での知見に関して報告する。

2. 1 取得データの種類

著者らのグループがこれまで集積してきたデータの一覧を表1に示す。

表1 著者らのグループがこれまで集積してきたデータの一覧表

<ul style="list-style-type: none">□ 道路交通データ・一般車プローブ・タクシー・プローブ 5分毎実時間(富士通)・ABS作動位置情報・渋滞データ (道路交通情報センター)	<ul style="list-style-type: none">□ 独自取得実時間プローブ・カー・データ・除排雪作業車(H27, 28年度:25台):札幌市雪対策室 実時間作業位置情報取得実験協力・北海道中央バス(H27, 28年度:23台) 実時間運行トラジェクトリ取得実験協力・レーザ・レンジ・スキャナ による道積雪・堆雪状況 3D実時間計測(H27, 28年度:中央バス1台) 実時間運行トラジェクトリ+ 実時間3D計測実験協力
<ul style="list-style-type: none">□ 気象データ・北海道メッシュ降雪量 :日本気象協会(6年分)・XバンドMPレーダ (5分毎):河川情報センター・マルチセンサーデータ :札幌総合情報センター(10年分)	
<ul style="list-style-type: none">□ 行政データ・地下鉄旅客数データ:札幌市交通局(9年分)・除排雪実施記録, 除排雪に関する苦情電話記録 :札幌市雪対策室(5年分)・交通事故情報(10年分) :北海道警察本部	

これらは、気象データ、交通データ、道路状況データ、除排雪記録データ、事故データに分類することができる。

気象データは、河川情報センターから5分ごとに受信しているxバンドMPレーダのデータと、札幌市が設置している気象マルチセンサステーションのデータの2種である。後者は過去データのみで実時間配信サービスは受けていない。

交通データに関しては、富士通からサービス提供を受けているタクシー2社のプローブ・カー・データを元に道路リンクごとの平均速度、台数などの統計データに換算されて5分

ごとに配信されるデータサービスと、1年ごとにまとめて過去データとしてホンダから提供されているプローブ・カー・データを用いたほか、独自に開発したスマートフォン・アプリを20台の中央バスの路線バスと、ボランティアの自家用車に搭載して得た実時間のプローブ・カー・データを用いた。

道路状況に関しては、独自に開発したスマートフォン・アプリをボランティアにダウンロードしていただき自動車に搭載していただくことで、ドライブレコーダ・サービスを提供すると共に、位置情報と共に、3次元加速度センサのデータとカメラ画像のデータを実時間で取得した。今年度からは、データ回収式のデジタル・タコメータとドライブレコーダをゴミ収集車20台に搭載して、プローブ・カー・データと道路画像の取得も行った。データ回収はゴミ収集車がセンターに帰還の際に無線でまとめて回収するようにした。また、中央バスの1台のバス車両の前面に、下方に向けて270°をレーザ・レンジ・スキャナでスキャンし、実時間で道路の3次元形状の計測取得を行った。このデータを用いると道路の有効路幅や両サイドの堆雪の高さなどを求めることができるが、この機器を搭載した車両を札幌市全域をカバーするように連日走行させてすべての道路リンクの計測モニタリングを常時行うことは不可能なので、このデータは分析や推定の検証を行うためのグラウンド・トゥールズ・データとして用いるために取得された。また、道路の摩擦係数を計測するために、ABSの起動データ（位置と継続時間、初期速度、終了速度を含む）を冬期間4か月分に関して、トヨタから特別に提供していただいた。これは市内を走るLexus車から取得されたものである。

除排雪記録に関しては、北地区の除排雪車20台に独自開発のスマフォ・アプリをローディングしたスマートフォンを搭載し、実時間でプローブ・カー・データを取得した。作業内容に関してはオペレータがボタンを押すことにより記録できるようにした。

この他にも、北海道警察から過去10年分の人身事故の番地データを取得し、札幌市からは地下鉄の各駅での乗降客数データと、除排雪に関する電話での市民からの苦情のデータを取得しているが、これらは分析や推定には未活用である。

2. 2 分析の目的

スマート除排雪における本研究の目的は相互に関連する以下の3点にまとめられる。

- G1 特定状況下での過去の除排雪作業に関する定量的評価とその情報可視化
- G2 現在の状況に対する実時間アセスメントと除排雪作業に関する意思決定支援
- G3 将来の状況に対する予測・推定とそれに基づく除排雪作業計画の最適化

実施された除排雪作業の定量的評価とその情報可視化は、従来、経験と勘に頼ってきた除排雪作業を定量的に評価し、問題点や効果的除排雪法を探索するには必須の技術である。現状では、市の担当者の意見としては、情報開示による苦情の増加を恐れるあまり、積極的な

賛同は得られていないが、このような情報を、実施された除排雪作業のコストとそれによる交通改善状況の定量的評価と共に市民に情報開示することにより、むしろ、徐排雪作業にたいする市民の理解が深まり、市と市民の共感に基づいて、よりよい除排雪作業の実施法に向けた議論が進むものと考ええる。

気象状況および、交通状況や道路状況の実時間モニタリングによる分析可視化と、それに基づく意思決定支援は、経験と勘に基づく意思決定に代えて、エビデンスに基づいた意思決定を可能にし、道路リンクごとの除排雪の緊急度を明らかにし、除排雪作業実施の優先度を付与することも可能にする。現在の除排雪出動の基準となっている10cm以上の積雪という条件は、その単純さゆえに、実運用においては除排雪実施の必要条件にはなっているものの、十分条件としては用いられてなく、往々にして、監督者の経験と勘によって作業実施が見送られることも多々ある。この点は、気象と交通状況の実時間モニタリングに基づいた、定量的な現状分析結果の情報可視化提示が行われることにより、著しく改善されるものと考ええる。

将来の状況に対する予測・推定とそれに基づく除排雪作業計画の最適化に関しては、シーズン単位の長期予測と、今晚の除排雪作業を実施すべきか否かを判断するための数時間から10時間程度後の状況に対する短期予測の2種類が考えられる。

除排雪作業は通常夜間に行われるので、その出動準備の判断は通常午後6時頃に行われる。出動の最終判断は遅くとも午後10時までには決定する必要がある。待機準備するだけでも莫大な人件費その他のコストはかかる。実施をする場合には除雪のみでも1億から数億円の費用を要し、排雪運搬までを行うと数10億円の費用を1晩の作業で要する。今晚、作業をすべきか、あるいは見送って問題ないかどうかは、午後6時から翌日未明までの降雪量に依存する。排雪運搬に関しては、作業後も継続的に大雪が継続するような場合は拡幅除雪までとし、排雪運搬は見合わせる必要がある。通常、排雪運搬はシーズン中に1回と考えられており、緊急事態が生じない限り、2回以上の実施は回避される。これは排雪運搬作業には除雪の80倍程度ないしそれ以上のコストを要するからである。予測に基づいて除排雪作業を見送った場合、予測が外れて降雪が続き交通渋滞をもたらした場合の経済損失も多大であり、除雪が行われなかったことにより、道路状況はそろばん道路を形成したり、路幅を狭くし、交通渋滞が一層深刻になり、これが周辺にも広がって、渋滞回避も不可能になるグリッドロックの状態が生じてしまう。このような状態になると、むしろ大災害状態と変わらず、復旧には数日以上を要することもある。従って、短期予測に基づく意思決定は、その成否によって、数億から数10億円のリスク・コストを伴うものとなり、予測精度が充分高くない限りは実用に供することはできない。残念ながら、数時間後の降雪状況を始めとする気象状況を、このような要求精度で予測できる技術は現状では未だ確立されていない。従って、実用の観点からは、短期将来予測に基づく除排作業に関する意思決定支援システムの開発は残念ながら現実的ではない。

予測とそれに基づく意思決定支援は、基礎研究としては成立し、それなりの学術研究成果

として論文発表などに結実させることは可能であるが、以上に述べたような理由から、現時点では、実用的な意味はないと結論できる。

長期降雪予測は、シーズンごとの除排雪予算の効率的な執行計画のために重要な技術であるが、気象のシーズン単位の長期予測技術も、現状の技術では要求精度で行う技術は未だ確立されてなく、実用の観点からは、このような長期計画支援システムの開発を目指すことも現実的ではない。

以上のことから、札幌市におけるスマート除排雪に関する本研究は、その目的として、G1とG2に焦点を絞って実施した。

除排雪作業において、最もコストがかかるのは、排雪運搬である。その中でも、排雪した雪を雪捨て場に運搬するコストが大きな割合を占める。この作業の最適化、スマート化には、

C1 雪捨て場への運搬ルート最適化

C2 特定ルートの特定区間の両側路側堆雪の排雪運搬に要するトラック台数の推定

の2点が考えられる。C2は、アイドリング待機をするトラック台数を少なくすることにも貢献すると考えられる。

C1の実施はコスト低減に大きく貢献すると考えられるが、現状では、委託業者と除排雪エリアごとに雪捨て場が固定的に決まっており、運搬ルートも交通の混乱を避けるために固定していると聞いている。従って、C1に関しての提案は、本研究の対象範囲を超えており、検討は見送った。この点に関しては、上記の制約を超えて、最適化の検討を行う意義があると担当部局が判断された場合には、改めて検討を進めたい。

C2に関しては、中央バスのフロントパネルに搭載したレーザ・レンジ・スキャナと同程度以上の仕様のもをパトロールカーに搭載し、同時に高精度のGPSを搭載すれば、これらから得られるデータを、昭文社の路幅データを含む地図データと照合することで、ほぼ10%以内の誤差で排雪運搬量の推定が可能と考えている。レーザ・レンジ・スキャナの取り付け位置に関して若干の工夫が必要と考える。本年度の研究においては、技術的な可能性を探るのみで詳細検討は見送った。

2. 3 除排雪実施の定量的評価とその情報可視化

図1は路線バスのルートと、赤で示される複数のバスの運行ダイヤと、左から2番目の運行ダイヤに相当するバスの実際の走行記録を青で示して比較したものである。これを見ると、路線バスが運行ダイヤから大きくずれて遅れの原因となっている箇所が、いずれも左折コーナーを開始点とする後続の少数の道路リンクであることがわかる。これは、左折の際、コーナーの除排雪が充分でないことから、コーナー部分の堆雪によって曲がりにくくなっていることと、見通しの悪さから横断する歩行者を注意して、速度を大きく落とさざるを得ない状況になるためであると考えられる。右折に関しては、右折車線が堆雪によって利用できなくなるような極端な場合を除いては、冬季とそれ以外での差は、左折の場合ほど大きく

はない。もちろん、右折車線が実質使えなくなったり、その実効的な長さが短くなるまでに雪の影響が出た場合には、右折による遅れも大きくなる。

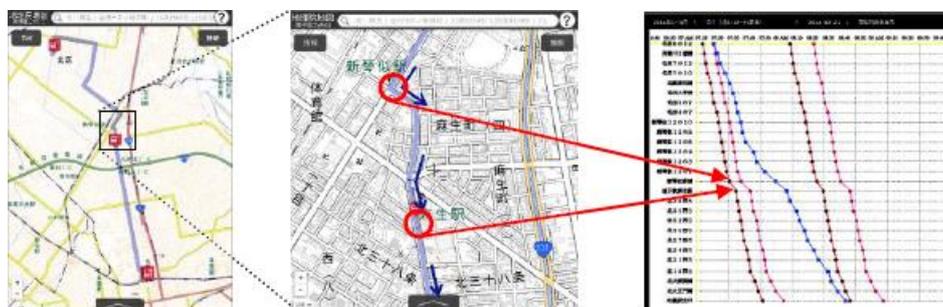


図1 路線バスの1ルートとそのダイヤと運行記録

(赤：複数のバスの運行ダイヤと、青：左から2番目のダイヤに対する実走行記録)

図2は、2016年の11月始めから2017年3月始めまでの積雪量の変化と、この間の3時間ごとの平均速度の低下率をヒートマップ表示したものである。X軸が時間でY軸が地図に示すルート上の位置に当たる。これを見ると、渋滞が交差点を中心にその周りに生じていることがわかる。

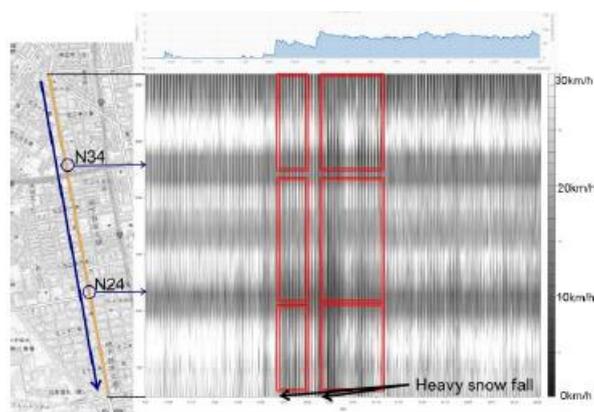


図2 2016年の11月始めから2017年3月始めまでの積雪量の変化と、この間の3時間ごとの平均速度の低下率のヒートマップ表示 (X軸：時間、Y軸：ルート上の位置)

交差点における交通の遅れが渋滞の大きな棋院になっていることと、特に、左折において交差点コーナーの堆雪が極端な速度低下の原因を作っていることが明らかとなったことから、担当者にもこのことを伝えたが、丁度、徐発担当者の方でも、コーナー除排雪の重点化を実施する計画があり、この機会をとらえて、除排雪車にスマートフォンを搭載して位置情報を実時間で取得すると共に、除排雪作業の前後の、左折の速度低下の割合の変化を定量的に評価してみた。図3は左側に1月28日～29日にかけて行われたコーナー重点除排雪の経路を示している。右上は作業前、右下は作業後の交通状況を可視化したもので、各交叉点ごとに、右左折と直進の12方向の平均速度を表示することができる。コーナー除排雪の対

象となった特定のコーナーに関して、東から南へ左折する交通流の平均速度を表示すると、上図では時速3kmであったのが、数では時速22kmに大きく改善されていることがわかる。子排雪対象となった他のコーナーでも同様の結果が得られた。このように、データに基づいてコーナー除排雪が左折時の速度低下を著しく改善し、渋滞緩和に効果的なことが実証できた。



図3 1月28日～29日にかけて行われたコーナー重点除排雪の経路と、作業前後の交通状況の可視化（右上：作業前、右下：作業後、図では同じコーナーの左折の平均速度の改善を示している。

2. 4 現状の実時間アセスメントと除排雪作業に関する意思決定支援

2015年11月24日、札幌はそのシーズン初めての本格的な降雪を迎えた。降雪は午前6時ころより始まり、午前10時には、除雪作業の出動基準である10cmの積雪を超えた。しかし、作業監督者の経験と勘から、シーズン最初の11月の降雪は大湯になることはないとの経験則と、正午ごろに降雪が弱まったことから、札幌市はこの日の夜の除排雪を見送った。その結果、夕刻並びに午後9時以降の本格的な降雪により、翌日、札幌市の交通は完全にマヒし、市民からの苦情が増大し、新聞各紙も千歳市が除雪を見送らなかったことを受けて札幌市の対応を批判した。翌25日の日中には除排雪作業ができないことから、状況はさらに悪くなり、道路は各所でそろばん道路を生じ、結果的に状況の回復のためには、通常はシーズンに1回程度しか行わない排雪運搬作業をも行う必要が生じ、その年度の除排雪の費用を大きく膨らませ、178億円とした。これはそれ以前の11年間の中で4番目に大きい額となり、以降の除排雪年間予算を大幅に上げることになった。

図4はこの時の降雪量、積雪量の変化と、XバンドMPレーダによる雪雲の正午と午後6時の様子を示している。積雪量は午前10時から11時の時点で既定の10cmを超えていた。雪雲の様子を見ると、作業判断を行う午後6時の時点で、正午の状況に比べると、雪雲の広がり狭くなっているものの、札幌市の上空に集中して存在していることが分かり、除雪の必要性を示している。

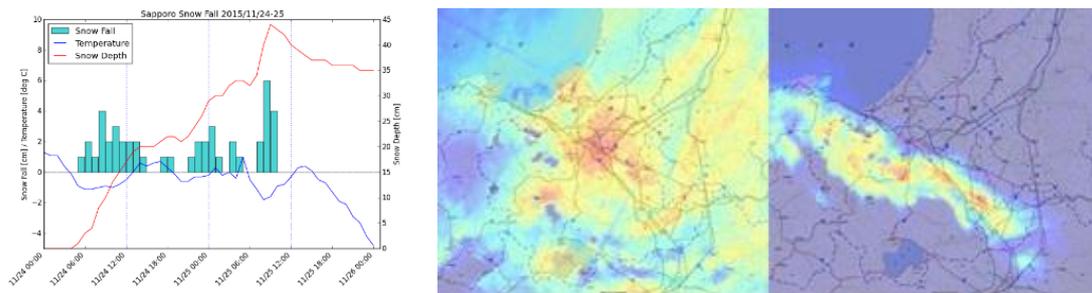


図4 2015年11月24日から25日にかけての降雪量、積雪量の変化と、XバンドMPレーダによる雪雲の正午と午後6時の様子

問題は、除雪作業の判断基準が積雪量 10 cmを超えた際という、非常に単純な基準に従っていることであり、そのために、作業監督者の経験と勘による判断が、この基準よりも優先されやすい点である。市全体をカバーする気象、交通、道路の状況に関する実時間モニタリングによって得られるデータを元に分析可視化を行い、データに基づいたエビデンス・ベースの判断を導入する必要がある。

図5の左図は、2015年11月24日午後6時の交通状況に関して、道路リンクごとの平均速度をある補正をしたのちに可視化したものである。平均速度をそのまま表示をした場合には、速度低下が降雪の影響によるものか、冬季以外にも渋滞を起こす道路リンクであるためかどうか判断できない。そこで、冬季以外の同じ道路リンクの同じ曜日、同じ時間帯の平均値と、現在の平均速度とを比較し、速度低下率を計算して色のグラデーションで表示している。赤が30%以上の低下、緑が15%の低下、青が5%以内の低下に相当する。この図を見ると、除排雪の出動判断を行う午後6時の時点で、冬季以外の平均速度に対して、30%以上の速度低下を伴う深刻な渋滞が、かなり広範な範囲で分散して生じていたことが分かる。所の中心部は殆ど全面的に15%以上の速度低下を起こしていたことも明白に観察できる。交通への影響は、午後6時の時点で、早急な除排雪の必要性を示していた。

図5の右図は2016年12月26日の交通状況を同様に可視化したものである。22日～24日にかけて強い降雪が続き、除排雪作業が連日行われたものの、とても追い付かなかった。この図を見ると、市の広範なエリアでほとんどの道路リンクが赤色になっており、冬季以外の平均速度に対して30%以上の重大な速度低下が生じていたことを示している。ブロックを囲む各道路リンクが渋滞し、渋滞回避を不可能にするグリッドロックが殆ど全域で生じており、もはや災害状態であることが見て取れる。

この2例から明らかなように、冬季以外の平均速度に対する速度低下率の可視化表示は、除排雪の緊急性とその個所を判断したり、除排雪作業の優先度を求めるのに非常に効果的な手法である。

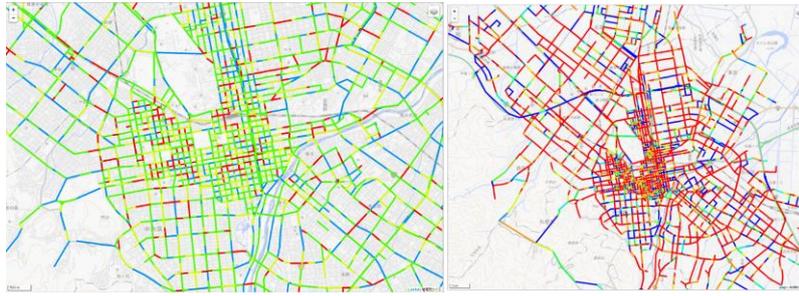


図5 冬季以外の平均速度に対する現在の速度低下率の道路リンクごとの可視化表示
 (左:2015年11月24日午後6時、右:2016年12月26日、赤:30%以上の低下、緑15%の低下、青:5%以下の低下)

2. 5 気象と交通の変化のみを用いた道路摩擦係数の推定

道路摩擦係数は、ABSの起動時間内における速度の低減量を起動時間と重力加速度で割ることによって得ることができる。札幌市全域をカバーするようなABSデータが実時間で得られるならば、市内全域にわたって摩擦係数の変化を道路リンクごとや、交差点の手前領域ごとに求めることは可能であるが、現状では、ABS起動データを実時間でモニタリングしている乗用車はごく一部のメーカーの高級車に限られている。しかし、数年にわたる冬季のABS起動の過去データが得られれば、これをその時の気象データや交通量データと関連付けて機械学習を行うことにより、気象と交通の状況変化に関する実時間データを用いて、現在の道路摩擦係数を道路リンクごとに求めることは可能である。本研究では、札幌における2012年12月から2013年3月までの冬季4か月間の1年分のABS起動に関する過去データ3万件をトヨタより提供していただき、気象データには、日本気象協会の1kmメッシュのデータを用いた。利用可能なABS起動データ数が3万件と少ないので、道路リンクごとの推定ではなく、1kmグリッド領域ごとの平均摩擦係数の推定を行うこととした。説明変数として、現在の気温、3日前までの日ごとの最高気温、現在の降雪量、3日前までの日ごとの総降雪量、現在の道路リンクごとの平均速度と平均台数、本日の周辺9グリッド領域の平均摩擦係数とABS起動総数の推定値の12変数を用いた。周辺9グリッド領域とは、対象グリッド領域に、これを取り囲む8領域を加えた3km四方の領域を指す。周辺9グリッド領域の平均摩擦係数とABS起動総数の推定値は、上記の12変数の内の最初の10変数を説明変数として用いた回帰分析で予め求めた。

この2段階回帰分析では、すべてのグリッド領域を最初に以下の3つのクラスに分類する。(1)凍結の場合、(2)雪が覆っている場合、(3)乾燥している場合の3つの場合である。この分類には、周辺9グリッド領域ごとに第1段階の回帰推定によって摩擦係数を推定し、その値が、0.2以下、0.2~0.5、0.5以上の3つのケースのいずれになるかによって、(1)(2)(3)の各場合に分類する。第2段階の機械学習は、これら3つの場合毎に

独立に行い、各々、独立に回帰推定を行う。また、ABS 起動は、ドライバーのブレーキ動作の違いにも大きく依存する。この問題を軽減するため、道路状況は 6 時間の間には大きくは変化しないとの仮定の下に、回帰モデリングには、直前の 6 時間の平均摩擦係数を求めて用いた。

図 6 は札幌市中心部の 10 km×10 km の領域に対し、1 km グリッドの解像度で、現在の平均摩擦係数の値を、気象データと実時間タクシー・プローブ・カー・データを用いた交通状況から推定したものである。左は、2016 年 12 月 3 日午後 8 時、右は 2016 年 12 月 7 日午後 8 時の推定値である。赤、橙、淡い青、濃い青は、それぞれ摩擦係数が、0.15 以下、0.15～0.2、0.2～0.3、0.3 以上の値をとることを示している。3 日と 7 日を比べると、凍結領域が広がっていることが見て取れる。

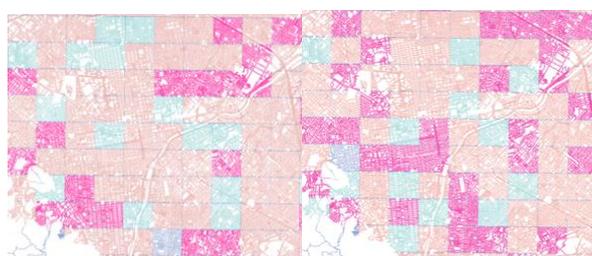


図 6 2 段階回帰モデリングによる路面摩擦係数の回帰推定（札幌市中心部 10 km 四方、1 km グリッドの改造度で推定）

残念ながら、現在利用している 3 万件のデータでは、解像度を上げて、道路リンクごとの推定を行うことは不可能である。これが可能になれば、凍結防止剤散布の道路区間を優先順位付きで提示したり、ドライバーに実時間で凍結路面区間を情報提示するサービスなどが実現可能になる。自動車メーカーでは、ABS データの実時間データ取得の範囲はその後拡大しており、協力が得られれば、十分な量の過去データの取得は現在は可能であると考えられる。将来的には、実時間 ABS 起動データの受信も可能になるのではないかとと思われる。そのためには、自治体として、メーカーに協力を要請することが必須である。

2. 6 有効路幅の推定

図 7 左上に示すのは、中央バスの路線バス 1 台のフロントに搭載したレーザ・レンジ・スキャナである。下側に向け、左右 270° をスキャンし、道路の路面と、両側路側の堆雪状況を 3 次元計測することができる。左下には独自開発のスマートフォン・アプリを用いて取得した走行経路と、3 次元計測結果を点群で表示した可視化結果を示している。路面の凹凸や路側の堆雪状況が見て取れる。このデータを用いると、任意の地点での道路の有効幅の時間的変化を求めることができる。右図は、3 つの異なる地点での有効路幅の変化を 2016 年 12 月 26 日～2017 年 3 月 8 日までにかんして示したものである。各チャートには、薄い

赤の領域が同時に表示されているが、これはこの道路の対面走行が可能な最低有効路幅の6 m以下の領域を示している。赤い曲線で示す有効路幅が、薄赤の領域に入ると、交通障害が生じることを意味している。このような計測は、比較的少ない限定数のパトロールカーや、多くの路線バスに同様の装置を問際してモニタリングを行うことにより、市全域での計測が可能になる。実際には、いずれの場合も多大な初期投資が必要となり実現性が低い。幹線道路のみに適用とか、グラウンド・トゥールース・データとして用いて、検証に用いたり、機械学習に用いて、別の説明変数のみから回帰推定を行うなどの利用法が考えられる。

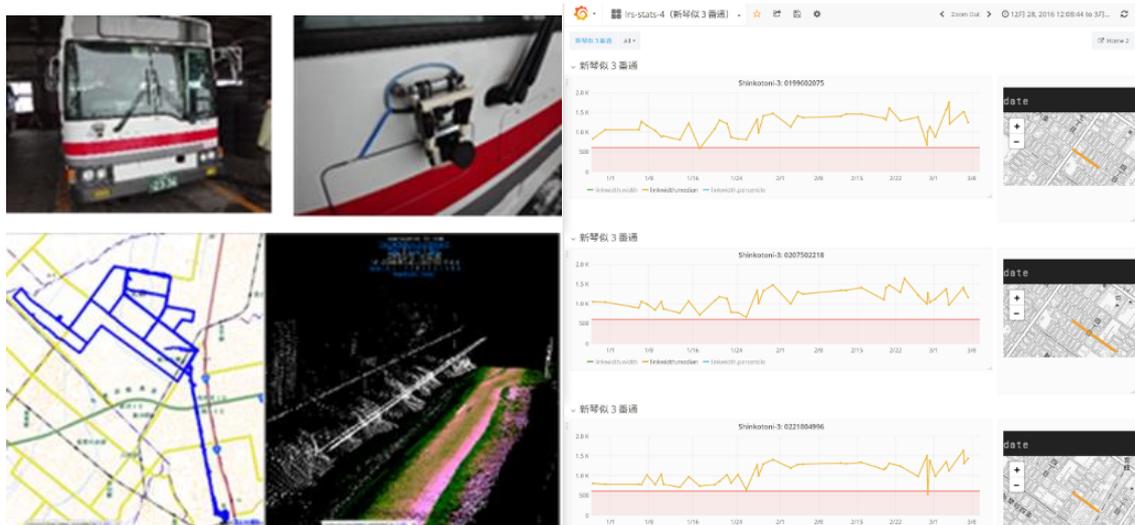


図7 中央バスのフロントに取り付けたレーザ・レンジ・スキャナと3次元計測結果の点群表示、このデータから3地点に関して得た2016年12月26日～2017年3月8日までの有効路幅の変化

2. 7 探索的可視化分析と統合監視分析指令センターの構想

スマート除排雪のためのデータ分析においては、過去の膨大なデータに基づいて、その中から、特定の状況と特定の作業判断のデータのみを用い、同じ状況での異なる作業判断のもたらす結果を比較分析したり、気象や交通、曜日や時刻の違いが結果にどう影響するかを見たり、気象の変化や除排雪の実施の有無により、渋滞の空間的広がりがどのように変化するかを分析したりすることが重要である。

このような分析は、通常、データの絞り込み（セグメンテーションという）と、絞り込んだデータに対する各種の分析を何度も試行錯誤的に繰り返して行われる。絞り込みの条件としては、説明変数の値に対する条件が課される場合の他に、分析結果に基づいて、特定のクラスターやマイニングされたパターンに対応するデータのみを絞り込みを行う場合もある。このような試行錯誤の分析過程を支援するには、探索的可視化分析システムと呼ばれるシステム技術が必要である。探索的可視化分析システムでは、データベース中の対象データを、

種々の異なる表現で可視化した可視化ビューと、特定の分析やマイニングを適用した結果を可視化した分析可視化ビューとを区別することなく、任意の数を同時に画面上に表示することができる。利用者はその内のどれに関しても、表示されている任意の可視化オブジェクトや、それらの任意のグループを画面上で直接指定することができる。システムはその指定に対応して即座に対応するデータのみへとデータベースのデータ集合の絞り込みを行い、その結果はすべての可視化ビューと分析可視化ビューに即座に反映される。

図8はスマート除排雪のためのデータ分析のために開発した探索的可視化分析システムである。

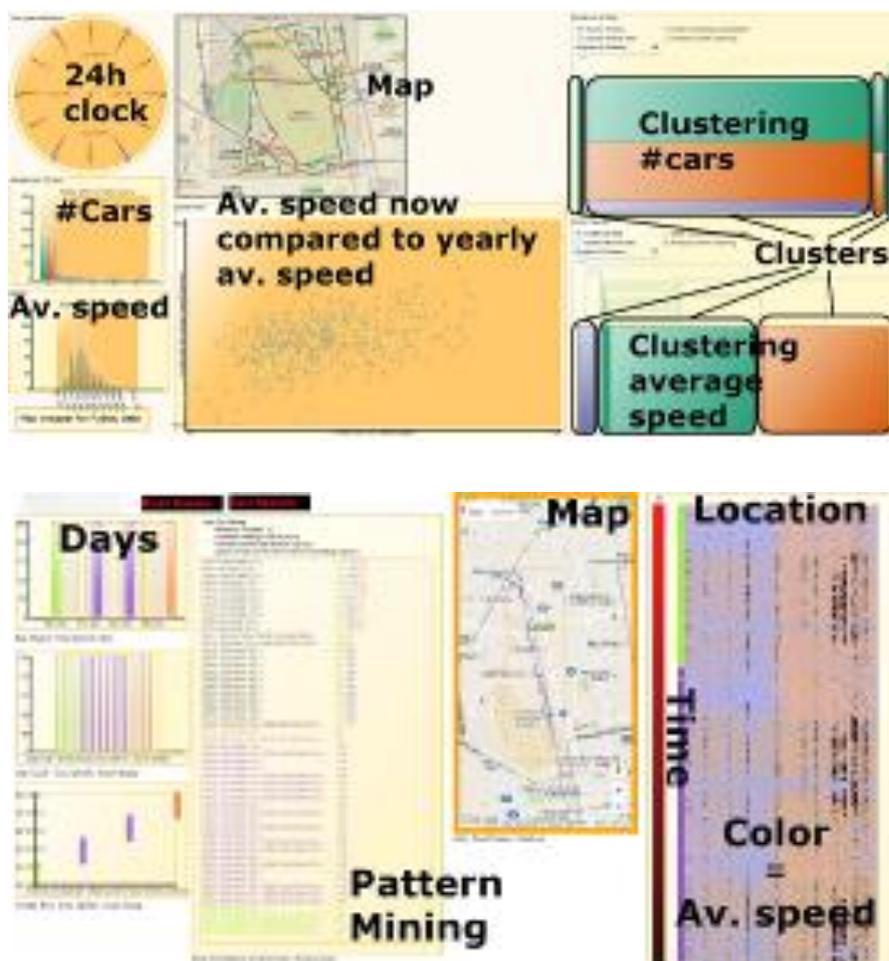


図8 スマート除排雪のためのデータ分析のために開発した探索的可視化分析システム

図8の上図では、対象データ集合は、24時間時計や地図を用いて、任意の場所の任意の時間帯のデータ集合へと自在に絞り込むことができる。平均速度のチャート上で任意の速度区間を指定すれば、その区間の平均速度を持つ道路リンクのみに絞り込まれる。クラスタリングは、道路リンクごとの平均台数と平均速度の2種類で2種類のクラスタリングが行われている。クラスタリング結果の可視化上で、任意のクラスタを選べば、そのクラスタに含まれる道路リンクのみに絞り込みがなされる。

図8の下図では、右端にヒートマップが示されている。X座標は道路リンク番号で、地図上で選ばれた領域内の道路リンクの集合を示している。Y座標は5分ごとの時間間隔を4日分示している。この4日間は左上のチャートで示されており、最初の日は無降雪、次の2日間は強い降雪、最後の1日は除排雪後の無降雪の日であった。ヒートマップのセルの色は、平均速度を表しており、黒が時速0 km、赤を最大速度としてグラデーションで表示している。青はその時間帯にデータがなかったことを示す。ヒートマップにおいて、平均速度に時速10 kmの閾値を設け2値化を行うと、各時間帯をトランザクション、道路リンクをアイテムとして、アイテム・セット・マイニングが可能である。この時求まるパターンのアイテム集合は、渋滞を同時に起こすことが多い道路リンクの集合を表している。このシステムは、探索的可視化し植えてむとしてじつ減されているので、左上の4日間を示すチャートで、降雪前の1日、降雪中の2日間、降雪語でかつ除排雪後の1日の3条件を選び、3つの異なる色で示すと、これらの絞り込み条件は他の可視化や分析結果にも反映され、同じく3色で対応部分が色分けされて表示される。パターンマイニングの結果を見ると、頻出パターンのリストが、その色で絞り込まれる道路リンク集合にのみ現れるパターンと、いずれの色の絞り込み集合にも共通に現れるパターンとの4種類のリストに分類されて表示されている。

このシステムを用いると、図9のように、降雪前に共起していた渋滞道路リンクの集合、降雪中に共起していた渋滞道路リンクの集合、降雪後で除排雪後に共起していた渋滞道路リンクの集合の3つに分けて独立に表示することも可能である。これらを比較すると、冬季においては、降雪前から渋滞の種となる道路リンクが存在し、降雪により、この種となる道路リンクを中心に渋滞区間が広がり、やがて飛び飛びの列島状であった渋滞区間が繋がり、大渋滞を生じるが、降雪が止み除排雪が行われると、元の列島状に戻ることが見て取れる。このような渋滞の種が何に起因するのかを見るために、同じ時間帯に計測した3次元の道路形状計測の3D点群表示を開いてみると、堆雪のすそ野に道路側への出っ張りが発達していることが分かった。継続的な違法駐車も同様の起因となる。ここから得られる仮説は、このような渋滞の種となり得るような、路側堆雪のすそ野の道路側への出っ張りを、降雪前にピンポイントで除排雪しておけば、大渋滞を防止できるのではないかということである。この仮説の検証のためには、ピンポイント除排雪の実施が必要であり、検証は未だ行えていないが、検証の価値は十分に高い。



図 9 探索的可視分析によって明らかとなった渋滞の種となる路側堆雪のすそ野のでっばりの発達

このような探索的可視化分析システムを巨大な共有スクリーンと、複数の分析担当者の手元の PC のスクリーンに自在に配分して表示できるような機能を装備した統合監視分析指令センターを実現し、札幌市のすべての交通状況、人流の動向、道路状況、気象状況、事故発生状況、状排雪作業状況、道路工事状況などをすべて統合的に監視分析し、除排雪作業の指令や、市民への回避誘導指令などに利用できるようにすることが望ましい。ゴミ収集者の運行監視なども統合することにより、ゴミ収集車に道路状況のセンシングを行わせてこれを収集し統合分析することも可能になる。

3. まとめと提言

データ駆動型のスマート除排雪に関する取り組みは、国内外ともに現状では都市間の感染ルートを対象としたものが殆どで、その内容は定点設置されたカメラや凍結センサーによる実時間モニタリングにより、除排雪や凍結防止剤散布の出動判断を行うものである。

都市の全域を対象としたものは、基本的には交通モニタリングシステムで、監視センターを設置し、大型スクリーンを用いて、定点カメラと定点設置センサにより交通状況の目視監視と、気象や道路凍結状況をモニタリングするものである。同時にプローブカーデータを用いるものもあるが、その規模は限定されており、本研究の実証実験で用いたような規模ではない。

データ駆動型のスマート除排雪の目標は、

- G1 特定状況下での過去の除排雪作業に関する定量的評価とその情報可視化
- G2 現在の状況に対する実時間アセスメントと除排雪作業に関する意思決定支援
- G3 将来の状況に対する予測・推定とそれに基づく除排雪作業計画の最適化

の3点になるが、G3は、短期予測、長期予測共に、将来の除排雪作業計画の最適化のために必要となる気象の予測精度が未だ十分ではなく、除排雪のように実施には非常に高いコストを要し、予測を誤って作業を行わなかった場合にも大きな経済損失を掃除するような作業の実施判断には、とても使えるような精度には達していない。G3に関しては、継続して基礎研究が必要であるが、実用化の観点からは、現時点としては対象外とすべきである。

G1に関しては、除排雪のアカウントビリティの向上のためにも、積極的にエビデンスに基づいた除排雪評価に関して市民への情報公開を行うべきである。市民からの苦情の増加を促進するのではないかと心配するばかりでなく、苦情に関しても、電話ではSNSを活用して、個々の苦情に関して、公開されたエビデンスに基づいて市民自体が議論できるようなシステム作りが有効ではないかと思う。

G2に関しては、本報告で紹介したようなシステムを現場で用いることにより、経験と勘に基づいた除排雪から、実時間でデータに基づいた現状把握による判断が可能になり、作業の優先順位まで考慮した判断支援が可能になると考える。

G1やG2を推進することにより蓄積される膨大なデータを用いて、種々の探索的可視化分析を行うことにより、特定の状況と特定の作業判断のデータのみを用いて、同じ状況での異なる作業判断のもたらす結果を比較分析したり、気象や交通、曜日や時刻の違いが結果にどう影響するかを見たり、気象の変化や除排雪の実施の有無により、渋滞の空間的広がりがどのように変化するかを分析したりすることが可能となる。本報告で示した渋滞の種の存在も、このような探索的可視化分析から見つかった。こうしたまだ知られていない新しい現象を見つけ出すことにより、ピンポイント除排雪の可能性の仮説のような、新しい除排雪手法に関する仮説を見つけ出すことが可能と考えられる。仮説の検証のためには、試験的な実施実験を行うためのエリアを設ける必要がある。

探索的可視化分析システムを用いた統合監視分析指令センターの設置も提案したい。このセンターの直属部隊として、分析に基づく種々の仮説を積極的に実証しうる除排雪知むとその実施エリアの設置も必要である。現状では、分析のために取得可能なデータは、現状の除排雪スキームに従った作業実施の結果データのみであり、現状のスキームから大きく逸脱した、新しい仮説に基づく作業に関するデータの取得はできない。従って、そのような作業の評価をデータの基づいて行うこともできない。根本に戻って、基本から最適な除排雪スキームを模索するためには、種々の仮説に基づく作業を実施し、実験データを取得できるような環境の整備が必要である。

本研究では、さっぽろ産業振興財団に協力して、財団が実施した総務省の「データ利活用型スマートシティ推進事業」の中で、「冬季プローブカーデータの収集・提供およびスマート除排雪サービス実証」の部分の研究計画策定と研究実施を行ったので、その報告書を付録として添付する。