

ビッグデータ分析の社会実験事例

Cases : Big Data Analytics for Smarter Society



株式会社 NTT データ 技術開発本部 サービスイノベーションセンター

小西 一也

1997年3月茨城大学大学院修士課程修了。同年より、NTT データ通信株式会社（現株式会社 NTT データ）入社、2000年 NTT サイバースペース研究所勤務を経て、2003年より現職。特許情報処理、ビッグデータ分析等の研究開発に従事。



株式会社 NTT データ 技術開発本部 サービスイノベーションセンター

田代 裕和

2007年3月早稲田大学大学院修士課程修了。同年より、株式会社 NTT データ入社、2013年米国 NTT DATA, Inc. 勤務を経て、2015年より現職。開発自動化技術、開発プロセス、ビッグデータ分析等の研究開発に従事。

1. はじめに

従来、組織は様々な場面で優れた意思決定をするために、より多くのデータを分析しようとしてきた。現在、ICT 技術の発達により、組織内外の膨大なデータを高速に分析することが可能になり、さらに分析結果から将来の傾向を予測することも可能になりつつある。

NTT データでは保有するビッグデータ分析技術の研究開発と、その有用性を実証するための社会実験を推進している。本稿では、ビッグデータ分析の参考事例として、NTT データが海外で実施した 2 件の社会実験を紹介する。

2. インドネシア・ジャカルタ市洪水把握実証実験

2.1 背景

本実証実験は、2014年の総務省公募案件である「アジア・オセアニア地域における準天頂衛星¹のメッセージ機能の活用等に関する調査」の一環として実施した。

ASEAN 地域では洪水災害が多発しており、特にインドネシア経済を牽引するジャカルタ市では大きな経済損

¹ 特定地域の上空に長時間留まる人工衛星。天頂付近が開けていれば衛星との通信ができることから、災害時の情報伝達などへの活用が期待される。

失が発生している。具体的な被害の例として、交通遮断による救急搬送の遅延や物流の停滞などが挙げられる。そのため、ジャカルタ市の洪水災害対策は非常に重要な課題となっている。しかし、市内には 800 以上もの小規模河川が存在するため、河川改修や浚渫、排水路建設などによる洪水予防や、監視カメラや水位センサなどの設置による災害状況把握を、広域かつ詳細に実施することは困難という問題がある。一方、ジャカルタ市は世界有数の大都市であり、車両登録台数が 300 万台以上、ツイッター発信件数が世界一といった特徴がある。このような特徴を踏まえ、ジャカルタ市では車両やスマートフォンを移動センサとして捉えることで、広域かつ詳細に洪水災害状況を把握できる可能性がある。



図 2-1 ジャカルタ市の洪水の様子 (LewatMana 監視カメラ映像)

以上の背景を踏まえ、本実証実験ではジャカルタ市において時々刻々と発生する車両 GPS(Global Positioning System) データとツイッターデータを処理して、GIS(Geographic Information System) 上に表示する洪水把握支援方法を考案した。そして NTT データの大規模リアルタイムデータ分析処理のノウハウを活かして実験システムを構築し、洪水把握を支援できるかを調査した。実験には 2014 年 1 月にジャカルタ市で洪水が発生したときの実データを用いた。また、当時の各集落の代表者による 6 時間毎の洪水状況報告を取りまとめた洪水マップをシステム出力結果と照合し、洪水把握方法の有効性について考察した。

2.2 洪水把握支援方法

洪水により交通支障が発生すると、発生箇所には車

両が進入できなくなることが予想され、またその周辺で洪水に関連する内容のツイッターが発信されることが想定される。そこで本実証実験では、車両 GPS データを処理して交通支障の発生箇所を推定し、ツイッターデータから洪水に関する内容のものを抽出するとともにその位置を推定する方法を検討した。そして、各推定結果を GIS 上に重畳表示する実験システムを構築した。本システムの全体像を図 2-2 に示す。本システムの表示を確認することで、地図上で交通支障の発生箇所を把握でき、さらにその周辺で発信されたツイッターの内容から原因が洪水であることを判断できるようになることを見込む。

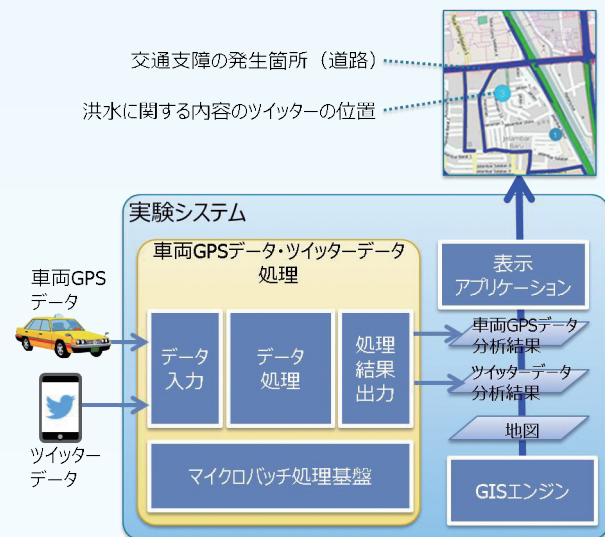


図 2-2 洪水把握システムの全体像

以下に、車両 GPS データ処理、およびツイッターデータ処理の概要を示す。また、時々刻々と発生するこれらのデータについて、各処理を遅延なく実行するためのマイクロバッチ処理基盤の概要も示す。

2.2.1 車両 GPS データ処理

本実証実験では車両 GPS データとして、約 16,000 台のタクシーに設置された各 GPS から 30 ~ 270 秒間隔で発信されるデータを利用した。この車両 GPS データにはタクシーの位置（緯度経度）情報や走行速度情報などがあるほか、本実証実験では位置情報と道路情報を突き合わせ、タクシーがどの道路上に在ったかを示す道路情報を付加した。

事前調査として、平常時と洪水発生時、各 6 時間の

タクシー分布を比較した。その結果、洪水発生時に洪水発生が報告されていた集落で、タクシー台数が減少していたことを確認した。そこで、各道路について平常時の平均タクシー台数を算出しておき、リアルタイムのタクシー台数と比較し、乖離が大きい道路を交通支障の発生箇所として推定する処理を実装した。ここで、平常時の平均タクシー台数については、2 か月間のデータから曜日ごとに単位時間あたりのタクシー台数の平均を算出した。リアルタイムのタクシー台数として、道路ごとに単位時間あたりの台数を算出し、平常時の同曜日、同時間帯の平均タクシー台数と比較した。

2.2.2 ツイッターデータ処理

本実証実験ではツイッターデータとして、洪水を意味するインドネシア語の「banjir」が出現したもの、およびジャカルタ市の BARAT 地区のおおよそ 10km 四方のエリア内で、発信位置情報付きで発信されたものを利用した。

洪水発生時、このツイッターデータは 1 日あたり 20 ～ 40 万件程度発信された。また、発信件数の時間変化を調査したところ、洪水発生の直後に極端に増加し、時間が経つと洪水の状態であっても減少することがわかった。この結果から、交通支障の原因判断に有効な、洪水に関連する内容のツイッターは、洪水発生の直後に抽出できると考えられる。

本実証実験では機械学習モデルを用いて、内容が洪水に関連する可能性が高いツイッターを抽出することとした。これは、一般的な洪水事象や液体が溢れる状態などを表すために「banjir」が用いられ、内容は実際に発生している洪水に関連しないツイッターを除去するためである。あわせて、洪水の発生場所とは異なる場所から洪水に関連する内容のツイッターを発信する場合もあるため、ツイッターに出現するランドマーク名から内容の位置を推定することとした。推定は、ツイッターをあらかじめ用意した約 1 万件のランドマーク名と位置情報のリストと照合することで行った。

2.2.3 マイクロバッチ処理基盤

洪水による交通支障の発生箇所を迅速に把握可能にするためには、時々刻々と発生する車両 GPS データとツ

イッターデータを、遅延なく処理し続ける必要がある。

本実証実験では、NTT データがこれまでに橋梁モニタリングなどの大規模リアルタイムデータ分析実証実験に適用してきた、ストリームデータのリアルタイム分析基盤である CEP(Complex Event Processor) を用いて、短時間データのバッチ処理を繰り返すためのマイクロバッチ処理基盤を用いた。

CEP を用いることで、処理の時間間隔は非常に短くすることができるが、あまり短くすると統計値の信頼度が低下してしまう。例えば平均タクシー台数を考えたとき、月曜日の 7 時 0 分から 8 時 0 分の 1 時間のタクシー台数は各週で似た傾向を示す可能性が高いが、7 時 0 分から 7 時 5 分の 5 分間のタクシー台数は各週でばらつく可能性が高い。本実証実験では試行錯誤の結果、統計値の信頼度が高くなるように、車両 GPS データ処理の時間間隔を 1 時間、ツイッターデータ処理の時間間隔を 20 分に設定した。そして、マイクロバッチ処理基盤を用いることで、これらの時間間隔で遅延なく処理し続けられることを確認した。これらの時間間隔は、各集落の代表者による洪水状況報告の 6 時間間隔の取りまとめと比較し、洪水把握をより迅速にできる可能性がある。

2.3 実証実験

実証実験では、洪水発生時の車両 GPS データおよびツイッターデータを再生して、洪水把握システムで収集、処理、結果表示し、洪水マップと照合することで、交通支障の発生箇所や洪水に関連するツイッターが表示される場所と、実際に洪水が発生した集落との関係を確認した。

まず、車両 GPS データ処理結果の例を図 2-3 に示す。図中の赤破線枠に示すように、実際に洪水が発生した集落の道路においてタクシー台数の減少が見られ、交通支障の発生を推定できることを確認した。

次に、ツイッターデータ処理結果の例を図 2-4 に示す。この例に示すように、実際に洪水が発生した集落における交通支障発生道路の周辺のツイッターを閲覧することで、支障発生の原因を洪水と判断できることを確認した。

これらの結果から、洪水把握方法によって、経済的影

洪水発生集落の道路で
タクシー台数の減少が見られた

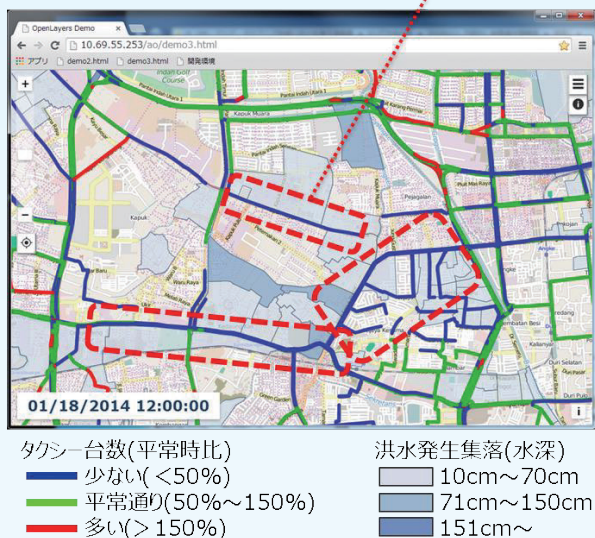


図 2-3 車両 GPS データ処理結果の例

交通支障が発生した道路の周辺のツイッターから
洪水事象が発生していることを確認できた



図 2-4 ツイッターデータ処理結果の例

響の大きい「洪水により交通支障が発生している道路」を、迅速かつ詳細に把握できる可能性を見出した。本システムを利用する場合に想定されるステップは以下の通りである。

- [1] 本システムが、「banjir」出現ツイッターの発信件数の時間変化から洪水発生を自動検出
- [2] 利用者が本システムの出力を目視し、平常時と比較してタクシー台数が減少している交通支障発生道路を把握
- [3] 利用者が本システム上で交通支障道路の周辺のツ

witterを閲覧し、交通支障の原因が洪水かどうか判断

2.4 今後の計画

紹介した洪水把握支援方法は、ジャカルタ市の地域防災庁などから高い関心を示され、実用化に向けた期待の声が寄せられている。今後、各データ処理の高度化や、監視カメラや標高データなどの他の情報ソースの拡充により、処理精度を向上させるとともに、それに応じた処理性能のさらなる向上を図り、ASEAN 地域の各都市も視野に入れ、実用化に繋げていく予定である。

3. 中国・吉林市渋滞緩和実証実験

本章では、中国・吉林市において ITS (Intelligent Transport System) 分野でビッグデータ分析を活用した実証実験について説明する。先進国や発展途上国を問わず社会問題となっている渋滞問題を交通ビッグデータ分析による渋滞予測と信号制御で解決可能であることを、渋滞問題が社会問題化している中国・吉林市において実証した [1][2]。

3.1 背景

交通渋滞による日本全国の損失は、国交省の統計によると、平成 14 年度で年間約 38.1 億人時間にも上ると試算されており、膨大な時間的、経済的な損失が発生している。今回実証実験を行った中国・吉林市においても、渋滞問題とそれによるバス運行の遅延が社会問題化していた。交通渋滞は化石燃料消費や CO2 発生等による環境負荷の要因にもなるため、現在、日本のみならず、世界各国で渋滞緩和・解消を行うための様々な取り組みが行われている。

渋滞緩和・解消を行うには、都市計画に基づいて道路や交差点の改良を行い、交通容量を拡大することが最も効果的であるが、多額の費用がかかることや、物理面また法制度面の制約からこれらの対策実施が難しい場合も多い。そこで、IT を活用することにより既存の設備を効率的に活用し渋滞緩和を目指す ITS の取り組みがある。日本では、ITS の活用を先進的に進めており、道路センサ等を用いて収集した交通情報を元に、事故状況

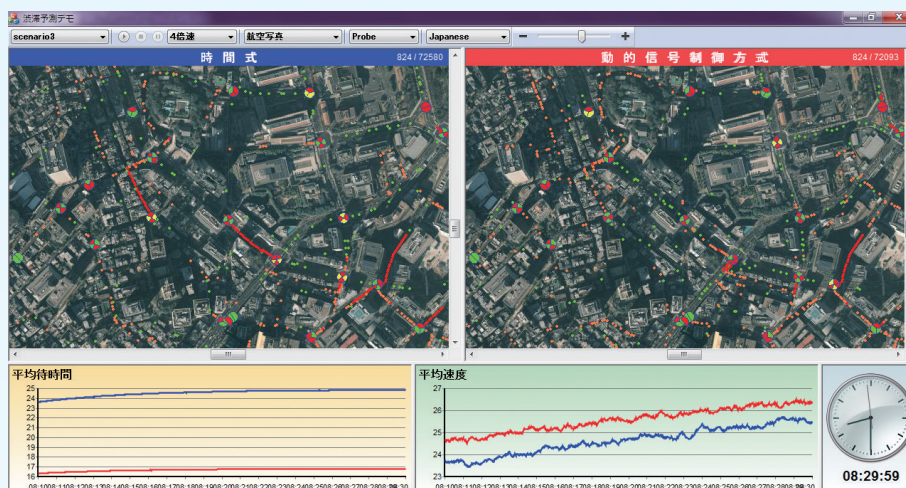


図 3-1 交通シミュレーションの画面
(東京赤坂付近)

や渋滞状況表示し、渋滞予測情報等を配信する、見える化の施策が実現されている。また最近では、道路センサ情報以外に、車両 GPS データと呼ばれる、車両の位置情報や速度等を含むデータの利用も活発に検討されている。これらは、近年ではスマートシティを実現する重要な技術として注目されている。

スマートシティ実現のためには、既存の設備を最大限に活用すべく、交通に関する多様なデータを元に最適な渋滞緩和施策を決定する必要がある。NTT データは、こうした状況を鑑み、交通シミュレーション技術を活用することによって渋滞予測・信号制御を行う技術を開発した。交通シミュレーション上で、複数の渋滞緩和施策に対する効果を比較できれば、最も効果の高い施策を選択することが可能となる。

3.2 NTT データの交通シミュレーション技術

3.2.1 交通シミュレーションの概要

NTT データの交通シミュレーションは、マルチエージェントと呼ばれる方式を採用している。この方式では、車、信号、道路、交差点等、交通状況に影響を与える対象をエージェントと呼び、このエージェントをコンピュータ上で動作させる。各エージェントの動作ルールは、現実の車や信号機等の動きを基に設定する。このエージェントを、コンピュータ上の仮想的な道路ネットワーク上で大量に動作させることで現実の渋滞状況を再現することができる。さらに、数十分先～数時間先等、将来の状況を再現できれば、渋滞が発生しやすい道路や時間

帯を事前に把握することが可能となる。

渋滞発生箇所が特定できれば、具体的な対策を検討することができる。対策としては、道路建設等、都市計画による対応や、信号設定や道路規制等設備による対応、カーナビゲーションシステム等への情報提示による対応等が考えられる。我々は、都市計画よりも短期間で実現でき、情報提示よりも積極的な働きかけが可能な、信号制御により渋滞緩和を行うアプローチが有効だと考えている。日本の交差点に設置されている信号の多くは、一定時間で表示が切り替わる時間式の制御が採用されているが、その青時間や複数信号期間での青時間の開始時間のギャップ等のパラメータを複数試行し、最も渋滞緩和された設定を選ぶことができる。

このように、交通シミュレーションを用いて、様々な状況での渋滞状況を予測し、効果的な対策を選定することで、安全かつ簡易に渋滞緩和を実施することが可能になる。

3.2.2 NTT データの交通シミュレーションの特徴

交通シミュレーションはその有用性から、既に製品化されているものも含め多くの取り組みが行われているが、NTT データでは、幅広いニーズへ応えるため、安価なセンサやサーバを用いて、様々なシナリオへ適用可能なソリューションの開発を目指している。ここでは、技術の特徴として車両 GPS データによる処理、サーバの分散処理について説明する。

① 車両 GPS データによる処理

道路に設置されている埋め込み型等のセンサは新規設置に大きなコストがかかるが、車載機器や携帯電話から収集する車両 GPS データは安価に大量のデータを収集することができる。そのため、車両 GPS データを用いて、エージェントの動作ルールを構成することでデータ収集コストを低減することが可能となる。

車両 GPS データにも、交通量の密度が低い道路では十分な情報が得られないことや、GPS 誤差による位置修正や補完が必要である等の課題は存在する。しかし、道路センサがない道路からもデータを取得可能なため、地図に対するデータのカバー率が高いことや、個別車両の把握が可能のため右折や左折等の分岐情報も利用できること等、大きなメリットもある。NTT データでは、これらの特性を正確に把握することで、安価かつ効果的なデータ収集を実現している。

② 計算サーバの分散処理

交通シミュレーションの計算処理では、地図の範囲やエージェント数等に応じてサーバにかかる負荷が変わ

る。渋滞が発生するような車の密度が高い状況を、広範囲でシミュレーションする場合、スーパーコンピュータ等高スペックなサーバを用いなければ実施が困難である。そこで、図 3-2 にあるように、地図をメッシュという処理単位に分割することで、複数の汎用サーバを用いての分散処理を実現した。検証では、9 台のサーバを用いて都内にある 100 万台規模の車の渋滞の予測をニアリアルタイムに実現した。

分散処理では、1 つのメッシュをサーバの 1 つのコアに割り当てて計算している。そのため、メッシュの分割はメッシュ間の計算量が平準化されるように行われる必要がある。しかしながら、メッシュの分割を計算するために、長時間を要してしまえば、分散処理による高速化の効果が意味をなさなくなってしまう。そこで、図 3-3 に示す NTT 持株研究所で開発されたグラフデータ分析処理技術 (Grapon) を用いることにより、道路ネットワークデータの分割処理を高速化し、リアルタイムに交通シミュレーションが実施できるようになった。図 3-3 に示す通り、グラフデータ分析では交通量増加により処理量が増加する地域を分析し、交通流が多い地域

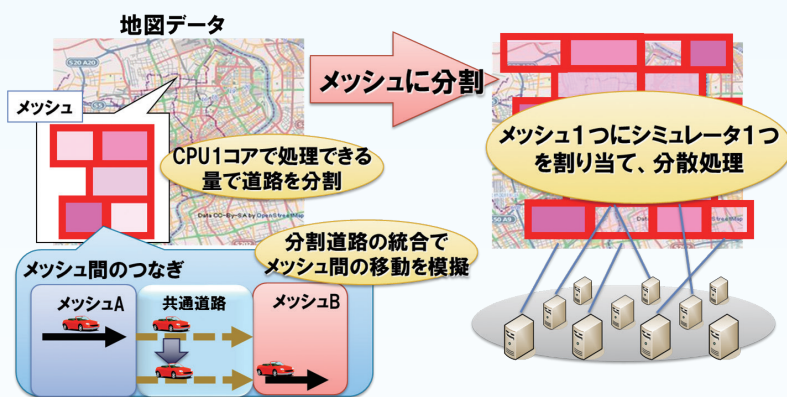


図 3-2 分散処理のイメージ

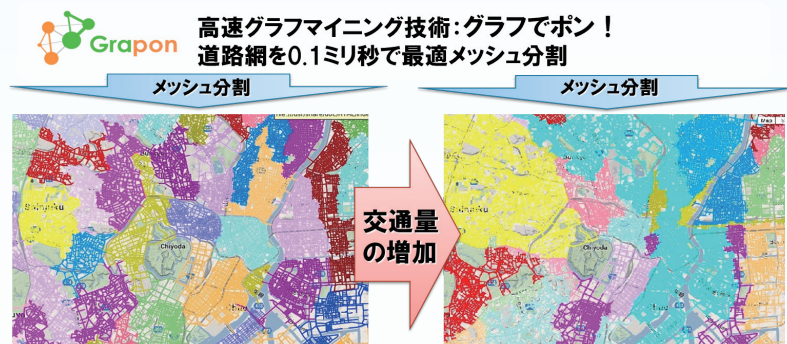


図 3-3 NTT のグラフデータ分析処理技術 (Grapon)

交通シミュレーションによる信号パラメータ最適化のイメージ

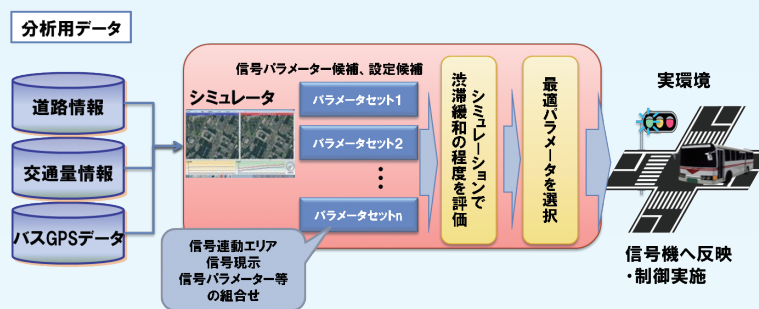


図 3-4：交通シミュレーションに基づく信号制御

は小さく、交通流が少ない地域は大きくメッシュを分割する。これにより、並列分散処理における処理量の平準化が達成される。

3.3 実証実験

NTT データは、交通シミュレーション技術を用いた信号制御技術の有効性を検証するため、2014 年 11 月 10 日から 12 月 12 日までの約 1 カ月間、中国・吉林市において、吉林市ならびに吉林市の現地企業と協力して、現地において実証実験を実施した。なお、本実証実験は、吉林市にとって実際の交通環境に適用する ITS の実証実験として初の取り組みとなる。

3.3.1 実験の背景

渋滞緩和には、日々刻々と変動する交通量を勘案し信号制御の設定を適切に行うことが基本となるが、信号制御の設定変更によって万が一渋滞が発生した場合に市民生活に大きな影響を与える。吉林市においては、これまでは信号制御の設定を変更することによるリスクの発生を懸念する声があった。また、信号制御を決める信号パラメータ（青→黄→赤→青の一連の流さを決めるサイクル長、青時間の配分を決めるスプリット、信号間のずれを決めるオフセット）が多岐にわたり、最適なパラメータを選択することが困難という課題もあり、信号制御の設定が十分に行われてこなかった。これらにより吉林市の中心部においては渋滞が発生し、市民生活の足であるバス運行時間の遅れに繋がっていた。

信号制御におけるこれらの課題に対し、交通シミュレーションを適用し事前に効果検証を行うアプローチが、信号設定変更のリスクを下げ、最適な信号設定の選

択を支援する上で有効であることを吉林市当局に説明し、警察の監視や交通データの提供等、本実証実験の支援を受けることができた。

3.3.2 実証実験の環境

本実証実験では、車載端末を通じて収集されたバス車両約 200 台の車両 GPS データと道路・交通量調査の統計情報を組み合わせて、交通シミュレーションを行った。交通シミュレーションでは、渋滞緩和につながる複数の信号パラメータを設定し、その渋滞緩和の効果を評価した。その中で、渋滞緩和効果が見られた信号パラメータを用いた信号設定を、吉林市中心部の 6 力所の交差点の信号機に反映した。効果検証では、交差点通過平均速度・バス運行時間等、定量的な指標を信号制御の最適化前後で比較し、渋滞緩和とバス運行時間の改善効果を評価した。

- 実験期間：2014 年 11 月 10 日～12 月 12 日



写真 3-1 実証エリア付近の渋滞の様子

- 実験エリア：中国吉林省吉林市中心部
- バス路線数：8 路線
- 車両 GPS データ収集対象のバス：約 200 台
- 交差点：6 力所

- 信号機：約 60 機

3.3.3 実証実験の結果

① バス運行時間の改善

バス運行時間の改善度合を検証した結果、実験における信号パラメータの変更により、交差点間の運動や青信号が点灯している時間の長さを改善したエリアにおいて、バス運行時間が平均で 7%、最大で 27% 改善した。

② 交差点周辺の渋滞改善

交差点ごとの渋滞状況を分析した結果、実験において信号間の運動を考慮して信号を制御した交差点で、車両 GPS データから算出したバスの平均速度が 1 ~ 2km/h 程度改善しており、交通が実験前よりもスムーズに流れていることが分かった。以上の結果から、交通シミュレーションによって、交通渋滞が緩和できることを確認した。

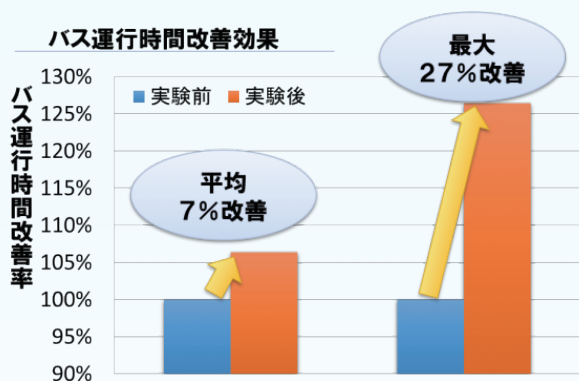


図 3-5：バス運行時間改善効果

なお、本実験にて顕著な渋滞の改善結果が得られたため、今回の実験で行なった信号制御の設定等は本実験後も吉林市で利用されている。本実験のさらなる展開として、実証エリアを拡大して最適化対象と範囲を広げ、タクシー GPS の活用等情報源を増やす等の取り組みによって、吉林市全体でさらなる渋滞緩和を実現することが期待されている。

3.4 今後の計画

本実証実験の成果をベースに、NTT データではさらなる検証を進め、信号制御最適化ソリューションの実用化を目指す。また NTT データではすでに車両 GPS デー

タ収集技術を開発しているが、データが不完全な場合にも対応できるようカメラ動画の分析やセンシングデータの組み合わせにより、交通状況を再現する技術開発に取り組んでいる。

今後は、2020 年に開催予定の東京オリンピックや世界各国で導入が進められているスマートシティ関連プロジェクトへの本技術の展開を図り、信号制御最適化ソリューションをベースとしたシステム構築により、社会全体のリソース最適化に取り組んでいく。

4. おわりに

本稿では、NTT データが実施したインドネシア・ジャカルタ市洪水把握実証実験と、中国・吉林市渋滞緩和実証実験の概要を通して、ビッグデータ分析の有効性の確認事例を紹介した。ビッグデータ分析を社会的に浸透させるためには、社会実験などの実証的な検討が不可欠である。本稿が、知財戦略の立案や知財管理の効率化等を実現するビッグデータ分析の検討の一助になれば幸いである。

参考文献

- [1] 中国・吉林市において、ビッグデータを活用した「渋滞予測・信号制御シミュレーション」の実証実験で渋滞緩和効果を確認 | NTT データ
<http://www.nttdata.com/jp/ja/news/release/2015/012303.html>
- [2] 将来の交通渋滞を予測・緩和する交通シミュレーション技術, 矢実貴志, 米森力, 道路 2015.6 Vol.891, pp.30-33

