2017.1.11 CSIS-S4D 第 2 回公開国際シンポジウム
基調講演　参考資料（日本語抄訳）

Telcos data for development: public – private partnership for sustainable development

Zbigniew Smoreda, PhD
Orange Labs, France

スライド 1
（抄訳なし）

スライド 2：Content

- 本日のプレゼンテーションは、携帯電話のデータ（CDR データ）を用いた人々の行動分析に関する考え方の流れを追いながら話を進める。これが最終的に「data for development」というアイディアにつながっていく
- また、OPAL という産官共同プロジェクト　―　様々な開発分野のプレイヤーに豊富なデータを提供できる枠組み　―　についても紹介する

スライド 3：CDRs telcos billing records

- CDR データは、通信の時刻と相手、種類、通信に使われたアンテナの ID から構成される非常にシンプルなデータ。このデータの特長の一つが、データの形式がどのような通信事業者のものでもほぼ共通な、課金のために収集されているデータであるということ。大概、数カ月から数年に渡って保管されているデータである
- 携帯電話の急速な普及によって、世界の人口の大部分をカバーすることができると考えられている

スライド 4：CDRs telcos billing records

- ビッグデータの解析技術・環境の整備が進み、CDR データのような大規模データが可能になると、通信事業者の研究部門と研究者の間で数多くのプロジェクトが立ち上がるようになった
- 通信事業者の観点からは、携帯電話利用者と通信相手の情報がランダムにコード化され、アンテナの ID が緯度・経度に変換されていればデータは共有するには十分な形だと考えていたが、これについては話の後半でもう少し議論を続ける

スライド 5：CDRs telcos billing records

- CDR データは、空間分解能の高いデータではなく、また携帯電話ユーザーが電話を使用したときにしか記録が残らないが、それでもなお数百万人の人口の中にある何らかの繋がりや動きを把握することができるデータ
- CDR データを用いた研究は、主に次の 2 つに分類することができる：一つは人々のコミュニケーションや繋がりに関する分析、もう一つは人々の動きに関する分析

スライド 6：CDRs first trial

- まず取り組んだのが CDR データの可能性を示すためのデータの視覚化（World Music Day の例を紹介）

スライド 7：CDRs first trial

- まだベルギーのデータを使って、社会的な関係性の強さが物理的な距離と関連があることを明らかにした。また、通話時間の違いが、通話相手との関係性の違いを示唆することも明らかにした。
-

スライド 8：CDRs first trial

- また、フランスのデータを用いた別の研究では、CDR データから検出できる国内のコミュニティの空間的分布は、行政区画による国土分割と相似していることが明らかにされた

スライド 9：Quickly growing research

- 本の数年の間に、通信事業者と研究者の間で、多くの CDR データを用いた共同研究が行われ、論文が発表された

スライド 10：NetMob – community building

- 2010 年に V. Blondel（University of Louvain）と A-L Barabasi（MIT）が携帯電話データ分析に関するワークショップを MIT で行う企画を立ち上げた。

スライド 11：NetMob – community building

- このワークショップにより多くの研究者と通信事業者の R&D コミュニティが一同に介する機会が生まれた

スライド 12：How to accelerate?

- 通信事業者はデータを部外者が利用することで生じうるリスクについても最大の注意を払う必要があり、厳正なる審査を経た研究チームのみが、CDR を使ったプロジェクトを行うことができる、という枠組みでデータチャレンジを実施することとなった
- 結果、ヨーロッパのデータ（既にフランスとベルギー、ポルトガル、スペインのデータが手元に集まっていた）は使わず、アフリカの途上国、コートジボワールのデータを使ったデータチャレンジを行うことにした。利用可能なデータが限られる途上国での試みは、先進国で行う試みよりもより大きな社会的インパクトがあると期待された

スライド 13：D4D: data for development Cote d'Ivoire

- ワークショップのために 5 ヶ月分のデータが提供され現地で匿名化された後、フランスの Orange Data Center でデータチャレンジ用のデータ加工が行われた。最終的にデータチャレンジ用に提供されたデータは、集計化したデータ、もしくは一部の標本のデータのみとした

スライド 14：D4D: data for development Cote d'Ivoire

- データチャレンジの発表は一般的な学術会議のような形式で行われたが、実態としてはハッカソンに近いものだった

スライド 15：D4D: data for development Cote d'Ivoire

- 応募は予想を遥かに超え、世界中から 263 チームの応募があり、最終的には約 80 チームが研究成果をレポートとして提出した

スライド 16：D4D: data for development Cote d'Ivoire
（抄訳なし）

スライド 17：D4D: data for development Cote d'Ivoire – first prize
- 優勝チーム（first prize）はバーミンガムの研究グループだった。彼らは地域間の人口流動により国内の病気感染がどのように広がっていくかをモデル化した。これにより、人の動きや、さらには情報の拡散プロセスが感染症の流行にどのような影響を与えるのかが示された
- この研究は非常に革新的であり、保健・公衆衛生分野の行政機関と通信授業者両者にとって有用なものであると受け止められた

スライド 18：D4D: data for development Cote d'Ivoire – development prize
- 開発賞（development prize）は IBM Dublin の AllAboard というプロジェクトだった。ここで提案されたのは、交通ネットワークの改善とユーザー満足度向上を目指した、CDR データを用いた公共交通計画の最適化を行うためのシステムである

スライド 19 : D4D: data for development Cote d'Ivoire – AllAboard applied to Abidjan, Ivory Coast
- CDR データは都市に住む人々の流動の起点と終点（origin - destination）を抽出し、そこから交通ネットワークの利用者数を計算するために用いられた。また、新たな交通ルートを提案することにも利用された。
- AllAboard により提案された最適化モデルでは、既存の交通ネットワークの利用者数を増やし、所要時間と待ち時間を短縮することでユーザー満足度を向上させるための分析を行うことができ、実際にアビジャンの SOTRA という既存の交通ネットワークの改善のために試用された

スライド 20: D4D: data for development Cote d'Ivoire
- その後 50 以上の論文がジャーナルで発表され、このデータチャレンジは多くの学術的コミュニティの関心を集めた

スライド 21：D4D: data for development Cote d'Ivoire
- D4D チャレンジは NetMob のコミュニティを超えて多くの研究者を一同に介する機会を創出した。この試みはオープン・イノベーションという観点からは大成功を収めたといえる。一方で、我々は、提案を超えた実際のプロジェクトについては、アビジャンでは一つも実現できなかった。
- 振り返れば、恐らくコートジボワールという選択は最善のものでなかったかもしれない。なぜなら、この国は当時内紛がやっと終結し、政情的にも不安的な国であったため、我々のプロジェクトを実施に移す段階になった時に必要な、現地のカウンターパートを探すのが非常に困難であったからである。

スライド 22：Second D4D Senegal: do it differently
- コートジボワールでの経験をもとに、よりパワフルな現地のカウンターパートの確保を目指して我々が選んだのはセネガルだった。なぜならセネガルには、Sonatel という通信事業者を通じて Orange も事業を展開しており、

また Sonatel は D4D チャレンジに関心を抱いていたからである。さらにアフリカで最も発展した国の一つであり、優れた大学や多くのスタートアップ企業が存在し、行政機関も優秀であるという、新たな事業を行うための環境が整っている国でもあった。

スライド 23：D4D Senegal

- Sonatel のアレンジを経て、我々は早速プロジェクトの説明のために現地入りし、D4D チャレンジを行う上での法的な問題点の確認と、チャレンジ終了後実際にプロジェクトを実施するための協力を要請するため、Personal Data Protection Commission と議論を重ねた

スライド 24：D4D Senegal

- また、データチャレンジを通じて得られた結果に関心を持ちそうな機関を事前にリスト化し、事前に面談を行う等の準備を行った。Sonatel の CSR 部門を通じて、多くの省庁や統計局、食糧保障プログラムにコンタクトを行い、まだセネガル国内の全ての大学にも D4D の活動を紹介した。同時に倫理委員会を立ち上げ、CDR データをデータチャレンジに利用する際のリスクを最小限に抑えられるよう、環境整備を行った

スライド 25：A process for Ethics review, involving both Orange staff and external experts set-up

- 今回のチャンレジでは、まず社会的、商業的、個人情報的なリスクのある可能性のあるプロジェクトについては、まず内部の評価委員会によって審査を行った。これに通過したプロジェクトのみが、外部の専門家の評価を受け、データにアクセスするための手続に入ることができるような枠組みを整備した。

スライド 26：D4D Senegal

- 今回 Sonatel は、前回のチャレンジと提供データの内容に一部変更はあったものの、一年分の CDR データを D4D チャレンジのための準備することができた。

スライド 27：D4D Senegal - April 2014 challenge launched

（抄訳なし）

スライド 28：D4D Senegal - April to August: more data and resources from donors

- D4D チャレンジの参加プロジェクトチームが二次データとして利用できるデータを集めるため、多くのデータを外部機関から収集した

スライド 29～37

（抄訳なし）

スライド 38

- 今後のために最も重要であったのは、3 つのテーマのためにゲイツ財団からの資金を確保したことであったと言える。彼らは、我々のデータセンターにあるデータにアクセスしてプロジェクトを続行している

スライド 39：D4D Senegal: Grants for implementation

- 保健、人口統計、農業の 3 分野についてそれぞれ 2 チームを選出し、合計 6 つのプロジェクトが実施されている。

スライド 40：D4D Senegal: 6 projects implementation

- セネガルでは D4D の結果報告会が行われ、D4D チャレンジの受賞者たちは、実際に現地に赴き、現地の行政機関や NGO と議論を行う機会がセッティングされた
- 現在実施されているプロジェクトは最終フェーズを迎えている。統計プロジェクトは過去に行ったコートジボワールの統計局の関心をよび、セネガルの次に、コートジボワールで同様のプロジェクトを転換する予定になっている

スライド 41：After D4D: the world of development needs Big Data

- この頃から、世界の発展のためのデータセット、について議論がなされるようになった。Y-A de Montjoye という数学者が、CDR データの匿名化は不可能であるというメッセージを発したのはこの頃である

スライド 42：Privacy warning: impossible CDRs anonymization

- ある論文で、150 万人分のデータを含む 15 ヶ月分のデータを使った時、少なくとも 4 つの時点の記録があれば、95%の確率で個人を特定できるという結果が報告された

スライド 43：The “privacy-utility” trade-off

- 事実、我々は個人の特定を不可能にする他に、データの精度を低くする、集計化処理行う、等の処理を行ってきた。一方で、プライバシーを保護しようとすればするほど、データの実用性が低下するというジレンマに直面していた。これは長期的に見れば、個人データを研究や開発に利用するための大きな障害となりうる問題である

スライド 44：A paradigm-shift in data protection

- この難題を打破するために、我々はデータとデータ利用者を完全に分離するシステム - Open Algorithm project (OPAL) – を提案した
- このシステムでは、データ管理者権限のあるものだけが、利用する目的に限りデータにアクセスする仕組みにより、個人情報の保護が保証される。またこのシステムから返されるデータは全て集計化された、個人レベルへのデータの再解析が不可逆なデータである
- 同時に、このシステムからアクセスできるデータにはノイズは含まれていない完全なデータあり、そのデータに対して、同様の処理を簡単に何度も繰り返すことがでる。さらにいかなるユーザーも利用可能なシステムである

スライド 45

（抄訳なし）

スライド 46：OPAL Project: query private data in a safe way

- このシステムを実際に利用して、我々は OD 表やアンテナ間の人口移動表等を作成することができ、また CDR データ以外のデータと組み合わせた分析を行うことも可能である
- この OPAL というシステムのポイントは、データが決して通信事業者の外部に持ち出されず、データの利用者がデータ利用をする際事前にデータの集計処理をしなくてよい、という点である
- このシステムに利用されているアルゴリズムはオープンソースであり誰でも利用することができる。また、このアルゴリズムの性能は、専門家と通信事業者による委員会により保証されている。このような方法で、我々はデータの秘匿性を確保しながら、実用性を高めるための試みを行っている

スライド 47～50

（抄訳なし）

以上