

地理空間データマイニングのための環境構築及び

データクレンジングツールの開発

井口 祐喜[†] 和田 雄次[†] 樋口 博[‡]

東京電機大学情報環境学部[†]

三菱電機株式会社 IT 宇宙システム推進本部[‡]

1. はじめに

近年、技術の向上により GPS や RFID を利用することで個体の位置情報を収集することが可能となってきた。将来的にこのような技術を使ったサービスでは大量の位置データを発生することが予想される。一方莫大な量のデータから有用な情報を発見する手法としてデータマイニングが存在する。本稿では個体の位置情報に対してデータマイニングを行うための環境構築、またデータを分析するために必要となるステップの一つであるデータクレンジングツールについて述べる。

2. 地理空間データマイニング

個体の位置情報と地理情報をデータマイニングすることを地理空間データマイニングと呼ぶ。ユビキタス社会により実現される位置情報ベースのサービスから生まれる位置データに対して地理空間データマイニングを行うことで様々な有意義な法則を発見することができると思われる。

3. 地理空間データ収集・管理環境の構築

地理空間データマイニングを行うためには個体の位置情報を収集する必要がある。本研究では測位を行うデバイスに GPS 内蔵 PDA を使用した。なお、この PDA は無線 LAN ネットワークに接続することができる。収集した

データは無線 LAN を介してアプリケーションサーバへと送られる。アプリケーションサーバにはデータベースを操作するサービスが稼動している。クライアントはこれを利用してデータをデータベースに格納する。データの保存形式は汎用性を考慮し XML を使用した。

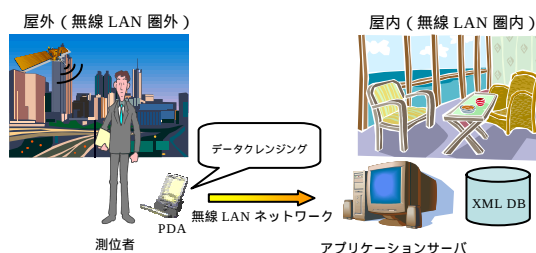


図 1 地理空間データ収集・管理環境

4. 地理空間データクレンジングツール

データクレンジングとは一般的にはデータの欠損値、空値を修正する処理のことである。この処理は分析を行う直前に行うことが一般的であるが測位者がデータに異常を見つけた場合、すぐにデータを修正できるようにクレンジング機能をクライアントに実装した。

GPS を利用して位置情報を収集した場合、マルチパスや誤った測位によりデータが汚れていることがある。このような汚れを削除、修正することでより良い分析結果を出すことが可能になる。そこでクレンジング機能として

マルチパスを修正する機能を実装した。マルチパスは前後のレコードとの距離で判定を行う。修正には前後のレコードの中間となる位置を使用した。

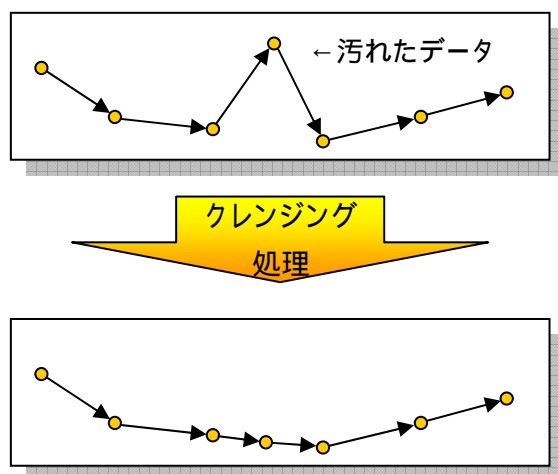


図2 クレンジングの例

5. 構築環境

本研究では測位を行う範囲を東京電機大学千葉ニュータウンキャンパス内に限定した。これはキャンパス内のほぼ全ての建物には無線LANが配置されているため測位者は屋内に入ることによってデータをサーバに転送することができるからである。ただし使用する機材がGPSなため測位を行えるのは屋外のみとなる。

6. システム構成

本システムは図3のような構成になっている。位置情報を収集しクレンジングを行うPDA、PDAから送られたデータをDBに格納するアプリケーションサーバ、アプリケーションサーバに受け渡されたデータを格納するデータサーバから構成されている。PDAにはMitec社のMio168を、アプリケーションサーバにはマイクロソフト社のインターネットインフォメーションサーバを、DBMSにはNCホールディングス社のNeoCore XML Management Systemの無償版であるXprioriを使用した。

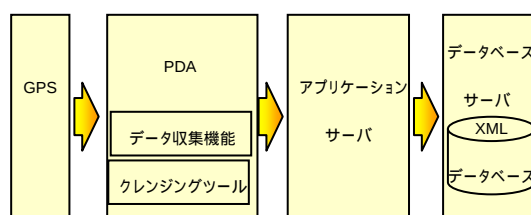


図3 システム構成

7. 今後の課題

今回構築した位置情報収集環境により地理空間データマイニングを行うための環境は屋外についてはある程度整った。次は複数端末を使い大規模なデータ収集を行い、実際にデータマイニングを行いたいと考えている。そのためにもクライアントの完成度をより高く使いやすいものへと改良する予定である。また、クレンジング機能についてもマルチパスの修正方法に中間の値以外の修正方法の実装などを行いたいと考えている。

8. まとめ

本研究により地理空間データマイニングを行うための環境を構築することができた。また、収集したデータに対してその場でデータクレンジングを行うことで最初からある程度きれいなデータをサーバに送ることが可能となった。今後の課題としては構築した環境を利用し、地理空間データマイニングの可能性について研究していく予定である。

参考文献

- [1] 土屋 淳、辻宏道 著 新・やさしいGPS測量 社団法人 日本測量協会 2001年
- [2] 松本 雄一、土井利次、仲瀬明彦、波田野寿昭 地理データに対するデータマイニングの適用と応用に関する研究 2003年 国土交通省、国土地理院 詳細地理情報を用いた実用的GISアプリケーションの開発に関する研究報告書