**大量のデータを収集分析する方法の調査**

PMコース　矢吹研究室　1142046　小池由也

1. 研究の背景

　ここ数年で，情報を得るためにウェブサイトを検索する機会が増えた．最近では「ビッグデータ」という言葉も聞くようになり，改めて大量の情報に囲まれていることが確認できる．

　ビッグデータという言葉には明確な定義がないが，私の研究では「量(Volume)」「多様性(Variety)「速度(Velocity)」の「3V」で特徴づける[1]．インターネットの普及によりGoogleやYahoo!の検索欄には，この瞬間にも多くの人がさまざまな言葉を入力している．さらにTwitter，FacebookなどのSNSには画像やつぶやきが投稿されていて高速で多様かつ大量のデータが生み出されるようになった．また，スマートフォンから取得される位置情報データや，防犯カメラで記録される人間の表情や動きのデータも「高速で多様かつ大量に生み出される」の「3V」の特徴に当てはまる．

　企業でもビッグデータを活かして，商品の購買履歴やサイト内のアクセス情報を基に，商品を購入する際に他のおすすめの商品やバーナー広告の表示を行っている[2]．他にも利用者の店内の行動データを基に利用者の目に留まりやすい所に商品を陳列させて売り上げを20％増加させた企業もある[1]．

　現在では，インターネットの普及とコンピューターの処理技術の向上により「3V」のデータと相性の良い分析法・アルゴリズムが発達し必要な情報の蓄積と解析が以前に比べて簡単に行えるようになりデータの使い道が「法則を確かめる道具」から「法則を発見する道具」として使われるようになった[3]．そのため従来とは異なり精度ではなく量を重視し因果関係ではなく相関関係が重視されている．

　ビッグデータはそれ自体に価値があるケースは稀で収集・解析を行い意思決定に生かす「ビッグデータ・マネジメント」と呼ばれる活動を経て価値を見出す．ビッグデータ・マネジメントには大まかに分けて2つの応用パターンがあり，ありとあらゆるデータを集めそれらのデータ二階席処理を行い，起きている事象に一定の規則性を見出し，それを利用する方法か，ネットワーク上に飛び交うあらゆるデータを捕捉し，その中に事前に検討をつけたパターンにマッチするデータの流れを認識した際，あらかじめ決めた処理を行う方法である．

1. 研究の目的

オープンソースプロジェクトを行う場を提供している．GitHubを利用し，そこで公開されてるプロジェクトにおいて使用されているプログラミング言語を1ヵ月間分のデータを収集し人気の言語を解析する．

1. プロジェクトマネジメントとの関連

　　　プロジェクトに関係するデータを収集し解析した結果を利用し意思決定を行うことで意思決定によるリスクの低下が見込める．

1. 研究の方法
   1. ビッグデータがどのようなものか，またどのように利用されているのかを調べ利点，欠点とビッグデータが利用された事例を調査する．
   2. オープンソースプロジェクトを行う場を提供しているGitHubがどのようなものなのか，どのようなプロジェクトがどれほど存在しているのかを調査する．
   3. データの収集・解析を行うツールを調査し簡単なデータで実際に使用する．
   4. 実際にプログラミング言語のデータを収集し解析する．
2. 現在の進捗状況

ビッグデータとは「Volume」「Variety」「Velocity」の3つの特徴を捉えたものでプロジェクトでは主に意思決定の材料として使われる．

　調査するにあたってオープンソースプロジェクトをGitHubにしたのは，矢吹研究室ではGitHubの研究がされていて私が行う調査はまだ研究されていない項目だったためGitHubで公開されているプロジェクトを調査対象とした．

今回の研究で使用するツールは，「GitHub Archive」を用いてデータを収集し「Google BigQuery」を用いて収集したデータの解析を行う．

1. 今後の計画

|  |  |
| --- | --- |
| 計画表 | |
| 内容 | 日程 |
| データの収集・解析ツールの試運転 | 2014年1月～ |
| データの収集 | 2014年4月～ |
| データの解析 | 2014年6月～ |
| 解析結果の有用性の調査 | 2014年8月～ |
| 卒業論文執筆 | 2014年10月～ |

参考文献

[1]城田真琴．ビッグデータの衝撃－巨大なデータが戦略を決める，ｐ．22-25．東洋経済新報社．2012

[2]WORKSIGHT“誤解されている「ビッグデータ」の本質”．2013-3-15．

<http://www.worksight.jp/issues/287.html>

[3]ニッセイ基礎研究所“ビッグデータで何が変わるか？”．2013-11-8．

<http://www.nli-research.co.jp/report/report/2013/11/repo1311-c3.html>