3ZE-01

データサイエンスのプロセスと業務評価モデル

清洲　正勝†　　沼尾　雅之‡

電気通信大学情報理工学部†　電気通信大学大学院情報理工学研究科‡

1. 序論

情報通信技術の進化により，ユビキタスコンピューティング，IoTによってあらゆる場所と物からデータを取得できるように発展している．膨大なデータから価値を見出し，社会に貢献するための有用な情報に変えるデータサイエンスの分野がますます重要性を増している．

ビジネスインテリジェンスシステム（BIS）の要素の一つである業績管理（BPM）において，費用対効果に一定の基準が存在しないデータサイエンスでは，今後多くの判断基準が必要となる．

そこで，本研究では，データサイエンスの汎用的なプロセスモデルを考え，データサイエンスチームのそれぞれの役割に対して貢献度を分析し，意思決定を行うモデルを提案する．

2. 関連研究

2.1 ビッグデータ

リサーチ・アナリストのDouglas Laneyは，ビッグデータという言葉を発案し，量（Volume），速度（Velocity），多様性（Variety）3つの特性を持つデータであると定義した．[1]

2.2 プロセスモデリング

ビジネスプロセスモデリング表記法（BPMN）[2]は，オブジェクトマネジメントグループ（OMG）が策定するビジネスプロセスモデリング記法の標準規格である．BPMNは，OMGが策定する統一モデリング言語（UML）で策定されているアクティビティ図を発展拡張したものであり，業務のプロセスモデリングに適した図である．

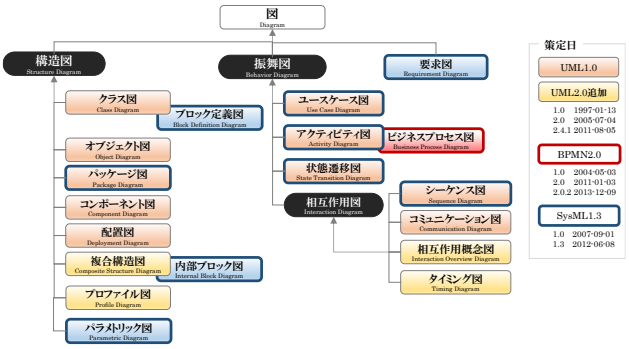

図 1. モデリングダイアグラムマップ

Process and Assessment Model of Data Science

Masakatsu KIYOSU † and Masayuki NUMAO ‡
† Faculty of Informatics and Engineering, The University of Electro-Communications. ‡ Graduate School of Informatics and Engineering, The University of Electro-Communications.
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585
† kiyosu@uec.ac.jp, ‡ numao@cs.uec.ac.jp

3. 提案

3.1 領域とデータタイプの定義

データサイエンスの領域を表現する4軸の象限図（図2）とビッグデータのタイプを表現する型式図（図3）を示す．

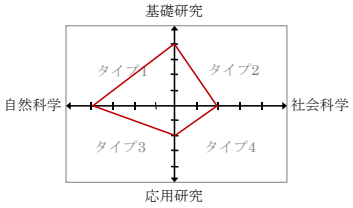

図 2. データサイエンス領域図

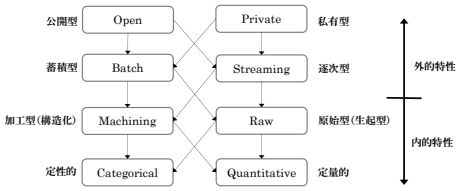

図 3. ビッグデータ型式図

3.2 作業分解構成の設計

CRISP-DM [3]に基づいて，データサイエンス独自のプロセスを定義し，作業分解構成（WBS）を設計する．（表1）

表 1. データサイエンスのWBS

フェーズ	ステージ	ステップ
ビジネス理解	要件定義	業務理解
		要求定義
		要件定義
	課題設定	問題発見
		課題設定
		仮説立案
		解決案選定
	業務計画	目標設定（KGI）
		要因設定（KSF）
		指標設定（KPI）
		体制構築
データ理解	環境構築	業務環境構築
	データ前処理	データ収集
		データ種別確認
		データ品質確認
	データ探索	データ特徴発見
		データ選択
データ準備（データマンジング）	データ修正（データクリーニング）	データ重複修正
		データ欠損値修正
		データ外れ値修正
	データ加工	データ結合
		データ抽出
		データ変換
		データセット生成
モデリング	モデル作成	モデル適用
		モデル調整
	可視化	図表作成

評価	模擬実験（シミュレーション）	実験環境構築
		ソフトウェア開発
		模擬実験
	解釈	知識検討
	報告	報告作業
		改善策提案
展開	適用	判断反映
		現場導入
		効果測定

3.3 チームの役割とスキルセットの定義

設計した WBS からデータサイエンスの汎用的なチームの 5 つの役割および，そのスキルセットを定義する．

1) プロジェクトマネジメント（PM）
2) マーケティング（MR）
3) データアナリシス（DA）
4) ソフトウェアエンジニアリング（SE）
5) インフラエンジニアリング（IE）

3.4 汎用的なプロセスモデルの設計

設計した WBS と役割の定義から汎用的なプロセスモデルを BPMN で設計する．WBS のステップをタスクとして設計したプロセスモデル図の抜粋を示す．（図 4）

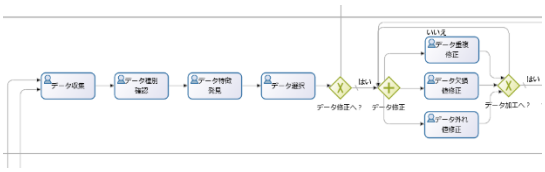

図 4. 汎用的プロセスモデル一抜粋

3.5 貢献度分析表の構築

プロセスモデルから責任分担表（RAM）を作成する．（表 2）

表 2.　責任分担表の例

フェーズ＼役割		PM	MR	DA	SE	IE
1	ビジネスの理解	R	F	A	I	C
2	データの理解	A	R	A	C	C
3	データの準備	I	F	R	A	R
4	モデル作成	I	A	R	F	I
5	評価	F	R	A	I	I
6	展開	R	F	C	R	A

R：実行責任 / F：追従責任 / A：説明責任 / C：相談対応 / I：情報提供

定数である創造した価値，変数である責任分担数および責任分担の加重値による貢献度分析アルゴリズムを設計し，貢献度 K を表す貢献度分析表を構築する．（表 3）

表 3. 貢献度分析表の例

フェーズ＼役割	PM	MR	DA	SE	IE	合計
ビジネス理解	4.88	3.90	3.41	1.22	2.44	15.85
データの理解	3.41	4.88	3.90	2.44	2.44	17.07
データの準備	1.22	3.90	4.88	3.41	4.88	18.29
モデル作成	1.22	3.41	4.88	3.90	1.22	14.63
評価	3.90	4.88	3.41	1.22	1.22	14.63
展開	4.88	3.90	2.44	4.88	3.41	19.51
K	19.51	24.88	22.93	17.07	15.61	100.00

4. 実験

4.1 検証

本研究室で行われた大学と企業の共同研究［4］（以下，プロジェクト）を対象として，プロセスモデルのタスク（WBS のステップ）レベルの思考実験を行った結果を示す．

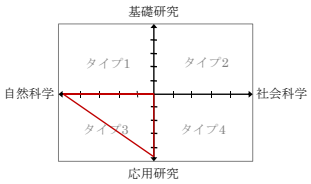

図 5. プロジェクトのデータサイエンス領域図

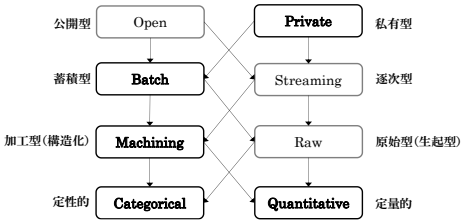

図 6. プロジェクトのビッグデータ型式図

表 4. プロジェクトの貢献度分析表

タスク＼役割	PM	MR	DA	SE	IE	合計
業務理解	1.05	0.84	0.52	0.16	0.16	2.73
要求定義	1.05	0.84	0.52	0.16	0.16	2.73
現場導入	0.16	0.16	0.16	0.16	0.16	0.79
効果測定	0.16	0.16	0.16	0.16	0.16	0.79
K	20.63	26.88	27.45	13.65	11.39	100.00

4.2 評価

プロジェクトを汎用的なプロセスモデルでトレースした結果，プロジェクトの内容が，汎用的プロセスモデルに準じていることが分かった．また，貢献度は MR と DA が高かった．しかし，研究論文からは実際のプロセスが分からず，ループバックのトレースが難しいことが分かった．

5. 結論

成功事例であるモデルケースでは，汎用的プロセスモデルにおいて抑えておくべき項目を踏襲して遂行されているといえる．したがって，本研究の汎用的プロセスモデルは，失敗に終わったプロジェクトの分析にも活用でき，貢献度分析は限られた人的資源を上手く活用するために用いることができると想定される．実験の結果から，本研究を業務管理や業績管理のコンピュータシステムに実装し，具体的なプロジェクトの積み重ねによって評価をすることが，本研究の価値を明確にするものであるといえる．

〈参考文献〉

[1] META Group, “Application Delivery Strategies File: 949”, Douglas Laney, 6 Feb. 2001.
[2] OMG, “Business Process Model and Notation”, 2004.
[3] Shearer C., “The CRISP-DM model: the new blueprint for data mining” J Data Warehousing, 2000.
[4] 沼尾雅之，松尾総一郎,『プロセス産業のための履歴テーブルに基づく品質分析法の提案』, DBSJ Journal, Vol.10, No.1, June 2011.