

VOL.4
2022

長崎大学 情報データ科学部

NEWS LETTER

TOPICS

ビジネスプロセス検証

数学の幾何学・特異点

データサイエンス

秘匿計算

多目的最適化問題

実社会課題解決プロジェクト

PBL de Thailand (2022年9月5日～15日)

2022年9月5日から15日まで、泰日工業大学(タイ)への学生10名を派遣しました。現地の学生と交流しグローバルな視点で物事を考える力を育みました。



泰日工業大学にて(2022年9月13日 修了式)



長崎大学
NAGASAKI UNIVERSITY

第11回 情報データ科学部コロキウム

<ビジネスプロセスの検証および幾何学の世界へご招待>

| 日 時 2022年5月18日(水) 17:00～18:00

| 場 所 オンライン開催(Zoom)

| 参加者 45名

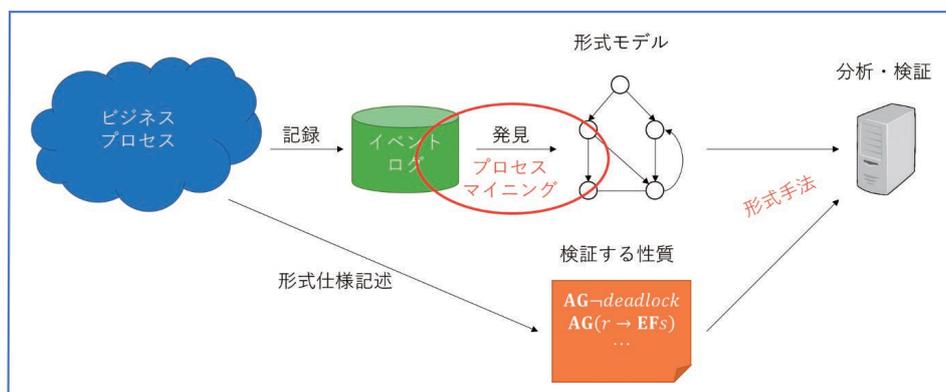
内容	講演者
講演1 プロセスマイニングアプローチによるビジネスプロセス検証	伊藤 宗平
講演2 また見つかった、何が、特異点が、データと溶け合う幾何学が	加葉田 雄太郎

プロセスマイニングアプローチによる ビジネスプロセス検証

長崎大学 情報データ科学部 准教授 伊藤 宗平

ソフトウェアは社会インフラの様々な場面で利用されており、その安全性を保証することが強く望まれている。ソフトウェアの安全性を保証するためのアプローチの一つに、ソフトウェアを数学的に厳密な言語で記述し、数学的あるいは機械的に検証する手法(形式手法)がある。一方、業務の手順であるビジネスプロセスを継続的に改善するためのプロセス分析手法としてプロセスマイニングと呼ばれるデータサイエンスの一手法がある。プロセスマイニングは、業務の実行履歴をイベントログとして記録し、そのログに対し種々のデータサイエンス的手法を適用することで、プロセスモデルを発見したり、業務が手順通り正しく行われているかどうかを検証したりすることが可能である。ビジネスプロセスの継続的改善において、(再)設計したプロセスが望ましい性質を持つことを保証することが強く望まれている。

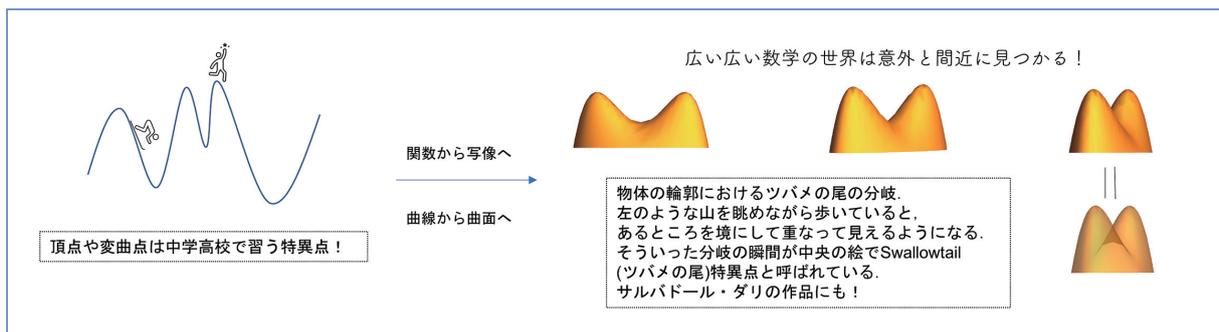
本発表では、プロセスマイニングと形式手法という異なる二つの技術を組み合わせることで、業務プロセスの形式的検証を行う枠組みについて紹介し、具体的なビジネスプロセスの検証事例およびイベントログからプロセスモデルを自動的に発見する試みを紹介し、明らかになった課題と今後の方向性について報告した。



また見つかった、何が、特異点が、データと溶け合う幾何学が

長崎大学 情報データ科学部 助教 加葉田 雄太郎

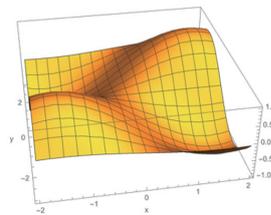
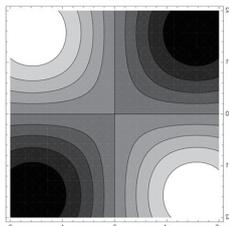
「数学の研究者です」と自己紹介をした時には不思議な顔をされることが多いです。これを読んでいる皆さんはどうでしょう？数学はテストのために勉強するもので、これ以上何を新しく研究する必要があるのだろうか、そもそも何の役に立つのだろうか、と思う人が多いのではないのでしょうか？例えば、皆さんが最近よく聞く「データサイエンス」の背後でもいろんな数学が活躍していて、きちんと理解するためには大学で新しく学ぶ数学の知識が必要です。そして、新しい技術が新しい数学を必要とすることも多く、それが数学の研究を推進する一つの動機と言えるかもしれません。さて、私はたとえば、数学の中でも幾何学や特異点論というのを研究しています。「形」に関することが大好きで、世の中にはどんな「形」があってそれらはどんな性質によって特徴づけられるのだろうか、そんなことに興味を持っています。意外に思う人もいるかもしれませんが、数学は「形」を表現するためのとても有用な言葉なのです。世の中の現象を知れば知るほど多様な「形」が現れて、それらを記述する言葉を見つける楽しみがある、そんな学問でもあるのです。



画像データは曲面ともみなせる！

グレースケール画像

・画像の各点 (x, y) に対して輝度値 $f(x, y)$ が与えられている



グラフ曲面

・3次元空間内の曲面 $(x, y, z) = (x, y, f(x, y))$

第12回 情報データ科学部コロキウム ＜消費者の行動データは宝の山＞

日 時 2022年6月15日(水) 17:00～18:00

場 所 オンライン開催(Zoom)

参加者 57名

内容	講演者
講演 マーケティング・サイエンスにおけるデータサイエンス	宮本 道子

マーケティング・サイエンスにおける データサイエンス

長崎大学 情報データ科学部 教授 宮本 道子

マーケティング・サイエンスでは、消費者の消費者行動から取得できるさまざまなデータを活用して、消費者に対してより最適化されたマーケティングを探る。ここでいうデータとは、消費者の年齢、性別、年収、家族構成などのデモグラフィックと、位置情報、購買履歴、アクセス履歴、センサーデータなどが含まれる。例えば、小売り大手のイオン・グループの子会社の「イオンマーケティング株式会社」は、日々、店頭から入る情報を反映したビックデータ、顧客アンケート情報、外部市場データによるマーケティング戦略をもとに、顧客ソリューションの提案を行っている。この手のデータはイオンのように、企業は自社で分析、調査し、それを基にビジネスにつなげて利用しているところが多く、なかなか研究用に提供されていない。

そこで、学生たちとマーケティング分析コンテスト等に参加し、それぞれの興味のある商品・サービスについて、業界背景や関連する文献調査を行った上、データの分析を行っている。データは非常に膨大な量である。あるコンテストでは、調査対象者：3,000サンプル、テレビ番組数：1万5千番組。調査対象商品：140ブランド、Webサイト：4,000サイトなど、分析対象となる変数の量も膨大となる。調査対象者が増えれば、より大きいデータとなる。そのため、十分な分析をするためには、データベースソフトや統計処理ソフトなどの活用、すなわちデータサイエンスの能力が必要となる。

ここで培ったデータサイエンスの能力は、将来、社会に出て、実際の企業データを扱うときにも大きな力を発揮してくれるに違いない。本学部からたくさんのデータサイエンティストが誕生することを期待している。

大分類	ファイル名称	データ形式	データの内容
アンケートデータ	メインデータ	CSV/SAS/SPSS	個人属性、チャネル利用、商品別の購入実態・購入意向など
	テレビ番組別視聴状況	CSV/SAS/SPSS	2012年3月1日～4月30日のテレビ番組(地上波)の視聴の有無
	雑誌閱讀状況	CSV/SAS/SPSS	2月25日～4月27日に発行された雑誌(695誌)の閱讀の有無
	Webアクセス頻度(11月、12月)	CSV/SAS/SPSS	3月1日～3月31日、4月1日～4月30日における個別サイトのアクセス回数(ファイルは2分割)
	データ定義	EXCEL	上記の各変数の定義説明用
出稿データ	テレビ出稿	CSV	テレビ番組別のテレビCM出稿データ(285598行のデータ)
	雑誌出稿	EXCEL	雑誌別の雑誌広告出稿データ(2821行のデータ)

出典：野村総研株式会社

第13回 情報データ科学部コロキウム

<安全と最適化：仮想空間における安全性と現実世界の最適化>

日 時 2022年7月20日(水) 17:00~18:00

場 所 オンライン開催(Zoom)

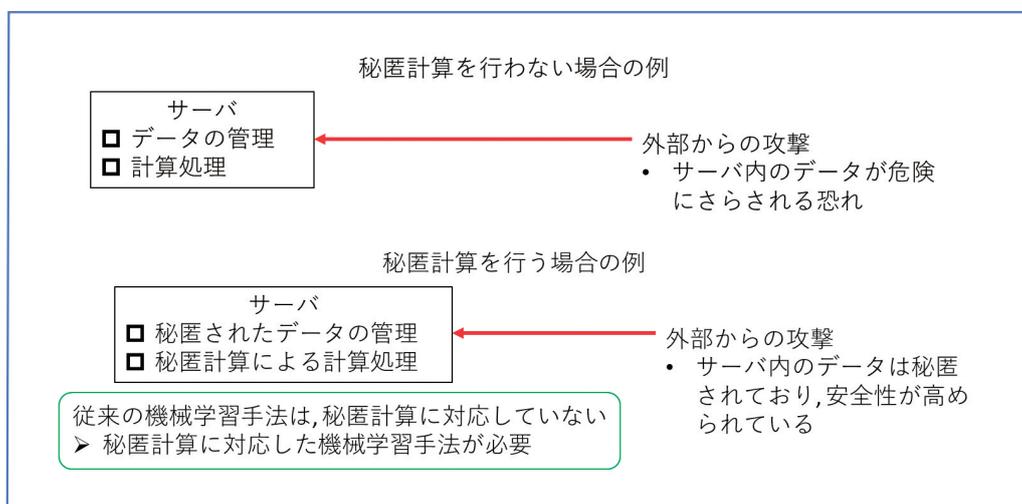
参加者 43名

内容	講演者
講演1 秘匿計算を用いた安全な機械学習	宮島 洋文
講演2 配送センターにまつわる多目的最適化問題	高田 寛之

秘匿計算を用いた安全な機械学習

長崎大学 情報データ科学部 准教授 宮島 洋文

近年、機械学習を用いたビッグデータの解析が広く行われている。機械学習においては、複雑かつ大規模な計算が必要となる場合があるが、そのような場合にはクラウドコンピューティングが用いられる。一方で、クラウドコンピューティングにおいてはデータを外部サーバに預ける場合、サーバに対する外部からの攻撃などによりデータが危険にさらされる恐れがある。そのため、データの安全性を高めることを目的とした研究が行われている。その一つとして、データを秘匿した状態を維持したまま計算処理を行う、秘匿計算が知られている。しかしながら、従来の機械学習手法はデータを秘匿していない状態のみを想定しており、データを秘匿した状態で計算処理を行うことができない。そこで、秘匿計算と組み合わせた機械学習手法を新たに提案した。この提案手法においてはデータが秘匿された状態で計算処理が行われるために、データを秘匿しない場合と比較してデータの安全性が高められている。また、この提案手法は機械学習の基礎理論に則っているため、高い精度を実現できると考えられる。そこで、数値実験を行ったところ、提案手法は従来手法と同等の精度を示した。



配送センターにまつわる多目的最適化問題

長崎大学 情報データ科学部 助教 高田 寛之

日本の少子高齢化による働き手不足の問題は、運送業界にも押し寄せている。従前のように各運送会社が人材を奪い合うことはもはや現実的ではなく、各社で協力しあって効率的に配送する必要がある。

この協力のための枠組みとして、仕事を長期的視点で仕事量と報酬の2つの観点において公平に配送員へ割り振る仕組みが必要になった。セイノー情報サービス株式会社では、この問題に対して、遺伝的アルゴリズムを適用したが、最適化が終わっているのかどうか判断できないという問題を抱えていた。

本講演では、この多目的最適化問題のパレート解集合を効率よく導くために、より簡単な問題に帰着させた。また、多目的関数を単一目的関数にするための重みパラメータの決定方法について、配送問題では仕事量の相場をロバスト線形回帰によって求めるアイデアを述べた。パレート解集合を得ることで、仕事量偏差と報酬偏差の間にあるトレードオフを明確にできることを示した。

質疑応答の内容について紹介します。人手、遺伝的アルゴリズムとの比較結果について。遺伝的アルゴリズムと本研究はほぼ同程度の公平さを達成。人手は偏りが発生していました。制約条件が追加された場合について。制約条件をラグランジュ緩和によって新たな目的関数に振替えできますが、その後には割当問題の標準形に帰着させることができないかもしれません。

図1 仕事(x:仕事量とy:報酬)の散布図と相場を割り出すためのロバスト線形回帰の結果

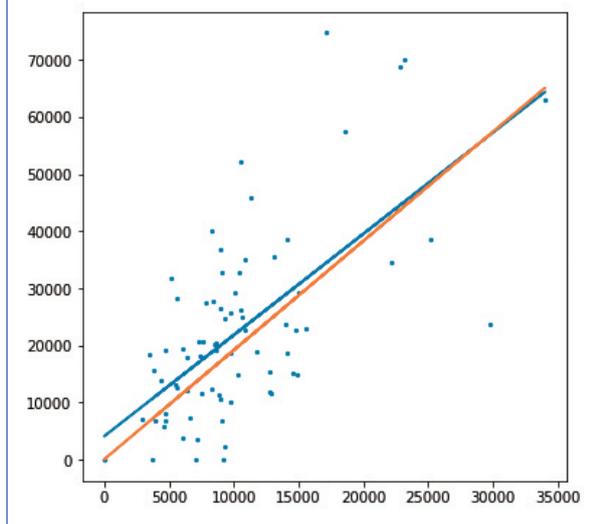
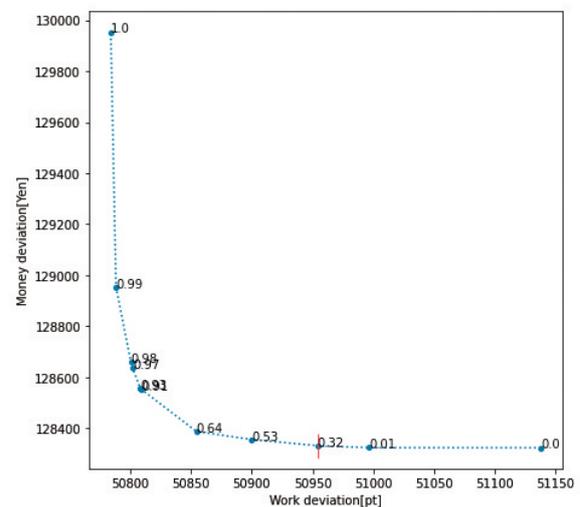


図2 パレート最適解集合の結果(○の部分) :

点に付随する数値は最適解に変化があった最初の値を表す。x軸は仕事量偏差、y軸は報酬偏差を表す。仕事量偏差と報酬偏差のトレードオフ。赤い縦線は、ロバスト線形回帰を用いて決定した重みパラメータ。



2022年度 実社会課題解決プロジェクトの活動状況

現在、2年生は通年のプロジェクト(フェイズ3)として15のテーマに分かれて24チームが活動しています。多くの学生が授業外の時間も使って取り組んでおり、プロジェクトを自走させることが少しずつできてきた印象です。また、1年生は後期からのプロジェクトがスタートして、参画事業者からいただいた13のテーマに分かれて20チームが活動しています。

1年生の活動の様子

後期からスタートしたフェイズ2では、それぞれのチームの目標設定も明確になり、いよいよ活動が本格化しています。Raspberry Piを使って工場の見える化に取り組んでいるチームでは、初めて触る機材ですが、自分たちで調べて徐々に使い方のイメージができてきた様子です。相対的貧困とフードロスと同時に解決するためにはどんなシステムがあるとそこに関わる様々な人たちの支援になるかについて議論が熱くなってきたチームもあります。前期の活動で「NFTアートを使った観光促進」という提案をしたチームは、その案に乗ってくれた事業者とともに、実際の仕事としてのリアリティをもって活動が進んでいて、とても良い刺激を得られていると思われます。

2年生の活動の様子

通年で取り組んでいる2年生は、プロジェクトもいよいよ佳境に入り、チームごとの色もはっきりしてきた様子です。フィールドワークも活発に行われていて、様々な場所に赴いてそれぞれの活動に必要なデータや素材を集めています。島原の観光促進に取り組むチームはSNSで発信するための動画コンテンツの作成のため現地ロケ撮影を実施しました(写真1)。また、ICTを使って生産者を支援するシステムをつくろうと取り組んでいるチームは、長与町のみかん農家を訪問し、収穫前後の樹木の写真を集めました(写真2・3)。

フェイズ3には、選択科目であるプロジェクトCを履修する3年生も加わっています。1期生である先輩として、経験してきたことを踏まえたリーダーシップを発揮してくれていて、心強いものです。履修している3年生はみな、学外の参画事業者のみなさんと関わりながら進める本プロジェクトで自分自身の成長を実感しているようです。



写真1 南島原の龍石海岸での撮影の様子



写真2 収穫後で力を使い果たしているみかんの木



写真3 収穫後でも力があり来年の豊作が期待できそうなみかんの木

2つの木の違いを画像から判別できるか、チャレンジが続きます！



長崎大学 総合生産科学域事務部西地区事務課 総務第一係(情報系)
〒852-8521 長崎市文教町1-14 TEL.095-800-4101

詳しくはHPへ!

YouTubeチャンネル



情報データ科学部

HP <https://www.idsci.nagasaki-u.ac.jp/>