

職務経歴書

2021 年 3 月 19 日

氏名: 柴田 怜

<ビジョン>

データサイエンス並びに動学的応用一般均衡モデルを応用し、様々な社会問題を解決する新規プロジェクトを提案していきたい。

<応募背景>

日本は現在、2050 年問題に直面している。

具体的には、AI-RPA の国内技術力が少子高齢化に伴う人口減少を補完するに至る見通しは立っていないうえ、気候変動等の環境影響も年々深刻化しており、その解決手段の提案が求められている。

この問題の所在は、データサイエンスの社会実装の遅れ、第二に環境政策における経済性の欠如と考えられる。

例えば、AI 研究開発は既に行われているが、具体的な社会問題への実用化は IT 企業によって行われている事例は限定的であり、RPA も Excel の入力等の事務作業を代行するに留まっている。

加えて、環境政策についても、研究レベルでこそ、2018 年にノーベル経済学賞を受賞した DICE(Dynamic Integrated Climate-Economy)モデルや動学的応用一般均衡モデルが開発され、環境・経済・雇用影響を同時に実証分析・予測することができる一方、日本政府がこれを ESG(Environmental, Social, and Corporate Governance)投資や環境税等に応用した事例は現在確認することができない。

以上から、データサイエンスの応用又は動学的応用一般均衡モデルの開発を以て日本社会の抱える諸問題を解決することを目的として、この度、コンサルタント職、将来的にはその PM を希望している。

<職務要約>

- データサイエンスを用いた解決手段の提案経験は、通算 6 回である。
 - 機械学習を用いた営業向け販促資料及びその予測モデルのアップロード
 - 小売店における来店人数予測
 - 半導体監視センサー値の閾値に関するアルゴリズム開発
 - 技術顧問
 - フードロスに向けたレコメンデーション・サービスに係るスマートフォン・アプリケーションの開発プロジェクト
 - PM に対データサイエンスに基づき、下記を例に助言した。
 - ◇ 具体的にどのようなレコメンデーションを行うことで、フードロスを解決し得るか?
 - ◇ 契約先の提供データを用いた場合、理論上、開発環境上、このレコメンデーションを全うし得るか?
 - ◇ 他のメンバーに対しては、Python のデバッグ処理を代行するとともに、GCP 又は AWS 等の利用に関する事前調査を指示した。
 - 採用面接補佐官
 - データサイエンティスト職の希望者に対する選考に際し、役員向けに助言した。
 - ◇ 面接時、理論と実装について下記を例に質問した。
 - ロジスティック回帰分析について説明されたい。
 - Inner Join と Left Join の違いについて説明されたい。
 - ◇ 明確な返答は一度もなく、すべて不採用を役員に助言し、役員は不採用と通知した。
- 主な提案業界は、製薬、マーケティング、大手 Sler である。
- データサイエンティストとして新卒入社し、様々な言語・ツールを用いた開発を 3 年以上経験した。

＜職務内容＞

総括	
課題	<ol style="list-style-type: none"> 顧客は、大量のローデータを管理する一方、プロジェクトに活用しきれていない。 予測精度が低く、問題解決に寄与しきれていない。 役員並びにPMらのデータサイエンスに関する知識不足
方法	<ol style="list-style-type: none"> 予測モデルについては、評価関数を基により精度を向上した。 技術を提案する際は、相手の知識水準と必要性に鑑みて助言した。 設計・開発における各段階にて、認識を合せた後、メール等にてエビデンスを残し、齟齬のないようにした。 採用面接補佐官を担当した際は、人件費と単価を予想して理論・実装に関する質問の要求水準を設定した。
提案	<ol style="list-style-type: none"> 評価関数については、機械学習における有効性及び現在のプロジェクトにおける必要性・実現性を開発責任者らに説明した。 採用面接補佐官としては、特に数学に関する知識と実装に取り組む姿勢を重視して採用可否を決定することを提案した。
結果	<ol style="list-style-type: none"> 相手の理解度に応じた説明により、多くの提案は受け入れられた。 顧客からも、データサイエンスに係る技術力及び課題解決への応用力を高く評価されることが多かった。 尚、2社ともSES(System Engineering Service)であった為、プロジェクトにおける金額上の寄与度は確認することができなかった。

＜職務詳細＞

2017年4月1日～2020年7月31日 株式会社セラク

■ 事業内容:IT アウトソーシングサービス

■ 資本金:2 億 9,639 万 9,500 円

■ 売上高:63 億 1,700 万円

■ 従業員数:2,479 名

■ Data Scientist

＜役割＞

- PMらの要求に応じて開発やリポーティングを行う。
- 必要に応じた提案を行う。
- 現場の同僚らとソースコード等のクロスチェックや設計の認識合わせを行う。

＜成果＞

1. 主に出力した.csv やグラフを成果物として納品した。
2. その説明資料を PowerPoint にて作成した。
3. 顧客に成果物に関するプレゼンテーションを行った。

主要プロジェクト	業務内容	開発環境	期間
機械学習を用いた営業向け販促資料	<ul style="list-style-type: none"> - ランダムフォレストを実装し、提案資料を PowerPoint にまとめた。 - DB からローデータを取得した。 	<ul style="list-style-type: none"> - Python3 - Redshift SQL - RStudio 	4 箇月
予測モデルのアップデート	<ul style="list-style-type: none"> - 製品拡張に伴い、LightGBM を提案・実装した。 - KPI リポーティング(英文) 	<ul style="list-style-type: none"> - PowerPoint - サクラエディタ 	
小売店における来店人数予測	<ol style="list-style-type: none"> 1. ローデータの集計処理業務 - 顧客の店舗における映像データと POS(Point of Sale)を結合した。 - 仕様通りの集計処理を行った。 2. 集計データを用いた来店人数予測 - 来店人数に機械学習(ニューラルネットワーク回帰)を用いた。 - 来店人数を予測した。 	<ul style="list-style-type: none"> - RStudio - Python3 - GitHub - Tableau - Excel - PowerPoint - サクラエディタ 	2 年間
高速道路情報の非正規化に関するクエリ作成	<ul style="list-style-type: none"> - 高速道路情報に関する各テーブルを非正規化した。 - BI ツールに向けた前処理を行った。 - 成果物納品に伴い、契約満了となった。 	<ul style="list-style-type: none"> - MySQL - Excel - サクラエディタ 	2 箇月
半導体監視センサー値の閾値に関するアルゴリズム開発	<ol style="list-style-type: none"> 1. 半導体の自動警告機に関する異常値検知 - 半導体の自動警告機のセンサー値に生じる異常値をメンテナンスと事故に区別すること。 - 残差解析によってモデルの妥当性を検証した。 - 単位根検定(ADF 検定)により、対数差分系列の異常値のうち、定期的なメンテナンスと事故を区別した。 - SARIMA モデルに学習データに用い、検証データを予測し、実験した。 2. フローチャート作成 - R 言語のパッケージである changept を他の言語にて再現する為、フローチャートを作成した。 - コーディングの技術を向上することができた。 3. グラフの分類 - パーツ毎に類似した傾向を持つセンサー値を分類することを為、1 標本 t 検定を行った。 - p 値を基に、クラスター分析を行った。 	<ul style="list-style-type: none"> - RStudio - PowerPoint - Excel - サクラエディタ 	6 箇月
特許管理システムのヘルプデスク	<ul style="list-style-type: none"> - 顧客からの問い合わせメールに対応した。 - システムの仕様や作動等を確認し、回答した。 	<ul style="list-style-type: none"> - Excel (VBA) - PowerPoint - Word - サクラエディタ - 7-Zip 	4 箇月

2020 年 9 月 1 日～2021 年 2 月 28 日 株式会社 EDIX

■ 事業内容:SES(System Engineering Service)

■ 資本金:300 万円

■ 売上高:不明

■ 従業員数:60 名程度

■ Lead Data Scientist

2020 年 9 月～2021 年 2 月 2/28(退職済みである。)

<役割>

1. 技術顧問
 - 客先常駐先のプロジェクト・マネージャーの下、Data Science に係る技術顧問を SES(System Engineering Service)として務めた。
 - 例えば、「契約先から提供されるデータを以てこのレコメンデーションを機械学習の理論上、開発環境上、全うすることは可能か? 不可能ならば、何を追加すれば可能か?」などである。
 - この為、営業会議や契約会議に同席した。
2. 採用面接補佐官
 - 株式会社 EDIX における Data Scientist 職の希望者に対し、役員向けに技術的側面から質疑応答を行った。
 - この役員は、私の助言を参考に採否を決定した。

<成果>

1. 技術顧問
 - プロジェクト進行にあたり、AI 研究開発や理論統計学の知見を以て補佐している。
 - 現在進行形のプロジェクトにつき、成果については不明である。ご了承されたい。
2. 採用面接補佐官
 - データサイエンティスト職の求職者に対し、理論・実装ともに、技術レベルを図る質問を行い、回答を求める形式を取った。
 - 例えば、ロジスティック回帰分析や Inner Join と Left Join の違いについて説明されたい旨を質問した。
 - 残念ながら、具体的な返答はなく、開発未経験と判断し、役員向けに不採用を進言した。

主要プロジェクト	業務内容	開発環境
フードロスに向けたレコメンデーション・サービスに係るスマートフォン・アプリケーションの開発	<ul style="list-style-type: none">- 昼食時の混雑状況に関するビッグデータを解析し、需給調整を行うことで、フードロスを減少する事業- データサイエンスに係る技術顧問を務めた。- 分析系メンバーに GCP や AWS のレコメンデーション向けの API について事前調査を指示し、そのレビューを行った。	<ul style="list-style-type: none">- Python3- R- Git- PowerPoint- Excel- Teams

＜テクニカルスキル＞

- 開発言語
 - Python3
 - R
- IDE
 - Jupyter Notebook
 - RStudio
- DB
 - MySQL
 - BigQuery
 - Redshift SQL
- 機械学習
 - LightGBM
 - XGboost
 - t-SNE
 - k-means
 - 密度準拠クラスタリング
 - 主成分分析等
- 計量時系列分析
 - 多変量確率ボラティリティ変動モデル
 - 動学的パネルデータ分析
 - 偏グレンジャー因果性検定
 - 単位根検定
 - 共和分検定等
- 統計解析
 - 統計的有意差検定
 - 一般化線形モデル
 - 構造方程式モデリング
 - 因子分析
 - t 検定
 - 分散分析等
- 動学的応用一般均衡モデル
 - 環境・経済・雇用影響に鑑みた確率過程に基づく動学的マクロ計量モデルである。
 - 需要関数と供給関数に対し、確率偏微分方程式を立て、非線形計画法によって最適化する。
 - これを基に、経済・雇用予測や環境政策における経済性について実証的に提案することができる。
 - LCA(Life Cycle Assessment)にも応用可能である。
- 産業連関分析
 - 環境・経済・雇用影響に鑑みた静的な経済として、応用一般均衡モデルを形成する。
 - 地域経済や製品・部品等のミクロ単位について Excel VBA にて実装可能である。
 - LCA(Life Cycle Assessment)にも応用可能である。
- IT ツール
 - GitHub
 - Excel
 - サクラエディタ
 - PowerPoint
 - Word
- 文献調査・リポーティング(和英)

＜学位・称号・受賞・資格＞

- [上智大学大学院修士\(環境学\)](#)
 - [Total GPA3.74/4.00\(93.5%\)](#)
 - 専門
 - ◇ 環境経済学
 - 動学的応用一般均衡モデル
 - DICE モデル
 - ◇ エネルギー政策
 - 3E+S+M (Energy Security, Economic Efficiency, Environment and Macro Impact)
 - ESG(Environmental, Social, and Governance)投資
- [法政大学学士\(経済学\)](#)
 - 専門
 - ◇ 環境経済学
 - 環境税
 - 産業連関分析による LCA(Life Cycle Assessment)
 - 応用一般均衡モデル
 - ◇ 計量経済学
 - 重回帰分析
 - 対数変換
 - 一般化線形モデル
 - 計量時系列分析
- [Kaggle 2 Experts](#) (2021 年 3 月)
 - [Code Highest Rank: Top 0.2% \(338/161858\)](#)
 - [Code Solo 3 Silver + 13 Bronze Medals](#)
 - [Discussion Highest Rank: Top < 0.4% \(738/188376\)](#)
 - [Discussion Solo 77 Bronze Medals](#)
- [統計科学研究所統計データ分析士 2 級\(第 1212 号 2017 年 2 月 28 日\)](#)
- [UiPath Academy RPA Developer Foundation \(2019 年 4 月\)](#)

Kaggle の成績要旨

Competitions	Top	Teams	Notebooks
Predicting Molecular Properties	21%	553/1636	
ALASKA2 Image Steganalysis	25%	272/1095	
M5 Forecasting - Uncertainty	26%	237/912	
TReNDS Neuroimaging	32%	336/1051	
OpenVaccine: COVID-19 mRNA Vaccine Degradation Prediction	33%	532/1636	1 Bronze Medal
Riiid Answer Correctness Prediction	34%	1133/3406	3 Bronze Medals
Jane Street Market Prediction			3 Silver + 6 Bronze Medals
Lyft Motion Prediction for Autonomous Vehicles			1 Bronze Medal
HuBMAP - Hacking the Kidney			1 Bronze Medal
Rainforest Connection Species Audio Detection			1 Bronze Medal

< 発表論文とその概要 >

1. 公益財団法人みずほ学術振興財団第 61 回懸賞論文【経済の部】「ESG 投資について考える」

[『エネルギー・気候変動と日本経済における 2050 年問題に向けた動的マクロ計量モデルに基づく ESG 投資の利用: 非線形パネル VAR-SPDE-LSTM モデル\(精度評価付き\)の開発』](#)

(要旨)

本稿は、エネルギー・気候変動と日本経済における 2050 年問題に向けた動学的応用一般均衡モデルを開発し、これに基づき、その解決手段として ESG(Environmental, Social, and, Corporate Governance)投資の利用を提案するものである。

ローデータを収集・整理した後、偏グレンジャー因果性に統計的有意性を確認する一方、共和分検定の結果、パネル VAR モデルによる動的直接相関係数を適当と判断し、その標準偏回帰係数に平均・分散を求めて幾何ブラウン運動を以て 2050 年までのシミュレーション行とともに、LSTM による精度評価を行った。

その結果、ESG 投資の利用をエネルギー・気候変動と日本経済における 2050 年問題に向けた解決手段として提案した。

1. 再生可能エネルギー普及策の推進
2. CCS(Carbon dioxide Capture and Storage)火力技術の開発・導入
3. シェール資源の輸入・産業利用
4. 気候変動対策に関する国際合意形成の推進における産学連携の資本

- [開発資料\(R による実行結果\)](#)

- [発表資料](#)

- キーワード:

- 共和分検定
- 単位根検定
- 偏グレンジャー因果性検定
- 非直交化インパルス応答関数
- パネル VAR(Vector Auto Regressive)モデル
- 幾何ブラウン運動
- 確率偏微分方程式
- LSTM(Long Short Term Memory)
- 多重共線性

2. 修士論文

[『エネルギー改革策が及ぼした環境・経済・社会的影響: 日独英仏国の実証分析と国際比較』](#)

- 日本のエネルギー改革策の持続可能性に関し、ドイツ・イギリス・フランスを先行事例とし、エネルギー政策に関連する政府統計を基に、各国に対数線形重回帰モデルを形成した。
- この結果から、日本のエネルギー改革策の持続可能性は必ずしも高くないと知見を得るとともに、再生可能エネルギーへの公共投資を第二次安倍内閣の新・経済政策における第三の矢とすることを提言した。

- 分析手法:

- 対数線形重回帰分析
- 主成分分析
- 分散分析
- 偏微分
- 重積分

Kaggle Code の詳細成果

1. 3 Silver Medals:
 - [“LightGBM Classifier & Logistic Regression + Report”](#)
 - [“Optimize LightGBM HyperParameter with Optuna & GPU”](#)
 - [“Optimized LightGBM with Optuna adding SAKT Model”](#)
2. 13 Bronze Medals:
 - [“RFCX: Residual Network with TPU Customized”](#)
 - [“Optimized Logit LightGBM Classifier & CNN Model”](#)
 - [“Predict Optimized Logit LightGBM Classifier & CNN”](#)
 - [“Riiid!: Optimized LightGBM with Optuna”](#)
 - [“LightGBM with the Inference & Empirical analysis”](#)
 - [“Optimize CatBoost HyperParameter with Optuna & GPU”](#)
 - [“LGBM on Lyft Tabular Data \[Inference\] + Tuning”](#)
 - [“Visualization Model with Customized Sharp Weighted”](#)
 - [“Research with Customized Sharp Weighted”](#)
 - [“Visualization & Empirical analysis of Submission”](#)
 - [“Introduction of LightGBM Tuner with Optuna”](#)
 - [“COVID-19 with H2OAutoML Baseline”](#)
 - [“【日本語】HuBMAP - Exploratory Data Analysis”](#)

<開発経験から学んだこと>

- 工程毎に顧客の要求項目・納期を明確化し、メール等によってエビデンスを取ってから、具体的に作業する重要性を再認識した。
- 各作業前に計画を立て、その手順を確認・共有することの重要性を再認した。
- 何事についても顧客や同僚らと齟齬を起こさないように報連相を欠かさず、迷ったら、確認を取ることを心掛けている。
- コーディングを行う際は、必ず設計を行い、ロジックを立ててから行うことで、デバッグ処理を短縮するよう心掛けている。

<スポーツ>

1. 筋力トレーニング
 - デッドリフト 400kg(リフティング・ストラップ付き)
 - スクワット 350kg
 - ナロー・ベンチプレス 180kg 6 回
 - チンニング(80kg 加重した状態) 8 回
 - ミリタリー・プレス 160kg
 - HIIT(High Intensity Interval Training) 10 セット以上
2. 水泳(4 泳法可能)
 - 平泳ぎ
 - クロール
 - 背泳ぎ
 - バタフライ

【扶養家族】 無 【配偶者】 無