ソフトバンクインターン面接対策

・自己紹介

東京理科大学大学院工学研究科経営工学専攻修士1年小坪琢人です。学部時代に講義を通じて統計学に興味を持ち、統計解析や時系列解析を主とする研究室を選択しました。研究室にて、学会やKaggleなどで開催されるデータ分析コンペティションに参加する中で、データ分析や機械学習に興味を持ちました。本日はよろしくお願いします！

・取り組んできた内容について

日本オペレーションズ・リサーチ(OR)学会が開催するデータ解析コンペティションやKaggle等のコンペティションに取り組みました。データ解析に取り組む中でデータの前処理、基礎集計、統計モデルや機械学習手法の適用を行いました。また、githubの活用により膨大なデータを処理する際に作成した多数のコード・モデル等を管理しました。具体的に、SQLでは効率的なデータの前処理・基礎集計、Rでは基礎集計・可視化・解釈性を意識したモデル作り、Pythonでは勾配ブースティングなどの機械学習手法について学習しました。企業から提供されたデータを用いたデータ解析を通じて、データの概観・前処理・結果の解釈性が重要であることを学びました。OR学会が開催するコンペティションでは、分析課題も各チームで検討・設定しました。その中で、データの細部を理解し、データ内に存在する本質的でない問題を取り除いてからでなければ、真に必要とされている問題解決にはたどり着けないと感じました。美容院のPOS データを用いたデータ解析では、顧客の住所とお店の住所はわかるが、どのような経路で来店しているかわからないという問題に直面しました。最寄り駅や最短経路による推定は可能ですが、移動経路に関するデータが手に入れば、より正確な分析を行うことができると思いました。コンペティションを通じて、ビジネスでの活用につながる多くの学びがあると同時に、欲しいデータや知りたい情報が手に入れなければ、実行できる提案に制限があることを知りました。貴社にはモバイル端末の位置情報やYahooにおけるウェブアクセス情報など魅力的なデータが多く存在するので、分析の前段階で起こりうる問題を最小限に減らせると考えています。貴社でのデータ分析においても、これまでに学んだデータの前処理や理解を意識して、結果の精度だけでなく結果の解釈性に重きを置いた分析に取り組みたいです

・研究内容について

私は風向などの角度データに対する、多変量自己回帰モデルの構築、検証を行っています。円周上に存在する角度データに対して、一般的な自己回帰モデルを適用すると、予測値が定義域から外れた値となることがあります。そこで射影正規分布という円周上の確率分布に基づいた、円周上における時系列解析手法を研究しています。ベイズ推定の枠組みを利用し、各パラメータの事前分布と観測データを用いた尤度式を用いて、パラメータの事後分布を算出します。パラメータの事後分布は複雑で陽に表すことができないため、マルコフ連鎖モンテカルロ(MCMC)法によりパラメータの事後分布を擬似的に作成しました。得られたパラメータの事後分布の平均値(期待値)をパラメータの点推定値として代用し、角度データの予測値を推定しました。研究を選んだ理由としては株価等の時系列解析には様々な歴史があり、多くのアイデアが生み出されているのに対し、角度データに対する時系列解析は比較的行われておらず、新しいことを試せる発展性に興味を持ちました。角度データでは平均値や分散という基本的な統計量でさえ、通常の計算式を適用することができないので、文献等を調べながら新しい時系列解析方法を考えています。2018年6月に香港での学会にて、「Circular time series analysis based on the projected normal distribution」という題目で発表しました。

・自己PR

内容: 逆算して最短距離を考える力

原体験: 大学受験(目標達成に向けてどの教科で何点とるかまで考えた), GPAあげた(暗記一辺倒ではなく, 考えられる部分はぎりぎりまで考えた, なぜそうなるのだろうと問い続けることで, 金融工学で100点や統計解析で98点とか取った)

強みの活かし方:なぜなぜ?と自分に問えるということは, 自分が解釈できないものを生み出さないということ. データ分析という業務の中でそのような意識はかならず役に立つ

弱みとして現れたこと:細部にこだわると時間がかかる（前処理とか）,大局的な観点で物事を見るというのは意識的に取り組んでいる.(kaggleとかで全部EDAを丁寧にやるよりも、まずは一旦予測して結果を出してみるとか)/ 0or100みたいな考えに陥ることがあった(意識的にだらだらするようにした、半日とか時間を決めて)

・インターン/アルバイト

内容: ディップ株式会社, 家庭教師

理由1: 就職活動の準備として, データ分析を始めたばかりの企業を見てみたかった

* ファーストキャリアは学べる環境が良いと思った。効率性が悪い

理由2: 勉強を教えるのが好き, 人のモチベーションをどうやってあげるか考えたい

* 人にものを売るとか, クリックさせたいとかにつながる考え方

・インターンでやりたいこと

1. 位置情報データの活用例を知りたい or AI・データ活用の導入支援を知りたい
2. 来場者数のようなものは計測できるがそれは、”点”の情報でしかない.

どこから来ているのかという”線”の情報が加わり, どのくらいの頻度で来ているのかという”線の幅”の情報, どんな人が来ているのかという”線の色”というように複合的にどんどんいろんなことを求められる.

・なぜソフトバンクのインターン

1. アナログとデジタルをつなぐ -> 広告や検索などデータと接触する部分はデジタルでも, 実際に行動するとき(買いに行く, 食べに行く, 遊び行く等)はアナログ, そういったときにはモバイル端末の情報が不可欠で多く情報を持っている会社に入りたい
2. auとかdocomoとかと比べて違う点: Yahoo, たくさん投資している(例:pokemon goのおかげで通信障害が起きやすいスポットを判明, ホテルとか野菜工場とか自動車とか人々の生活を全て網羅できる), 自社にデータ分析を行う部署を抱えている

・ソフトバンクのニュース

通信障害: 改めてインフラの重要性, IT社会においてセキュリティ, 保持の重要性を認識しました。

MONET Technologies: ソフトバンクとトヨタ, オンデマンド型の配車サービスや企業向けのシャトルサービス, 異業種タッグ, MaaS（Mobility as a Service）

・ソフトバンクバリュー

「No.1」 「挑戦」 「逆算」 「スピード」 「執念」

・逆質問

ソフトバンクで求められる資質は？

インターンシップではどのような業務に取り組めるのか？

インターンシップまで期間があるのですが、準備しておいた方がよいこと?

分析業務を子会社化, 自社で抱える違い?

2/22の午後だけ休めないか?(別インターンでの発表会がある)