

システム運用データを活用したデータ分析手法検討

SSE01-12 菅野未来 miki.sugano@ctc-g.co.jp

自社の課題

自社の強みとしてマルチベンダー対応を長年行ってきており、顧客対応を行っている中で様々システムデータやナレッジデータ等のデータが多く蓄積されている。それらのデータを活用し既存サービスへ付加価値をつけたい。

リソースデータの選定と目的

システムリソースを利用して将来の利用予測を行ったり、複数システムリソースを組み合わせて傾向を分析することで障害予兆を検知する仕組みを検討。また、ナレッジデータと新規アラートデータを分析しトラブルシューティングのレコメンドを行う手法を検討。

データ別分析手法

リソースの将来予測

機械学習アルゴリズムの『線形回帰』を用いて、システムリソースデータ(例えばメモリ使用率等)を分析することで、将来の利用予測が可能であると考える。キャパシティ分析のサポートとなることができれば短期・中期・長期でのリソース使用が把握できるようになると考える。

処理性能による障害予兆予測

障害は様々存在するが、今回は性能処理における予測にフォーカス。この分析にはクラスタリングとベイジアンネットワークによる分析が有効である可能性がある。特にベイジアンネットワークはデータ間の因果関係から事象の発生確率を算出できる手法であるため、異常な状態の発生確率を導き出すことで、確率が高い状態になった場合に検知することができる。性能が劣化する前に劣化する傾向をつかめれば障害予兆として利用できる可能性がある。

最適ナレッジのレコメンド

ナレッジに記録されているアラート情報と新しく発生したアラート情報を利用して、最適な過去ナレッジをレコメンドする手法を検討。アラート情報が自然言語のような文章で出力されることが前提だがWord2vecという手法を活用することで最適なナレッジをレコメンドできる可能性がある。

共通課題

データと目的別に分析手法を調査したが、いずれの場合にも精度を出すためには適切な学習データを使用している必要がある。特に異常を検知するためには、正常性のデータのみを学習させる必要があり、異常時のデータが混在していると検知漏れにつながる可能性がある。なるべく多くのデータを利用して学習させることが望ましいがデータが多いと学習に時間がかかるためこれらの課題をどう解決するかが重要。

感想

今回は主に分析データと目的に応じて分析手法の調査と調査から見えてきた課題を考えるまでにとどまったが、他にもさまざまな手法や課題が浮上すると思われるためシステムとして実装する際にはより検討を深める必要があると考える。
自社の組織構成として企画と開発が分かれており、私個人は企画という立場から、目的と実装案をいくつか提示することを目標としているため、今回技術検証までは行っていないが、今回の調査したで得られた結果をもとに開発側と情報連携し、検証を進めていきたい。