

それでは、大阪大学大学院情報科学研究科の山本航平が、LTE 環境における応答遅延特性の時系列モデリングによる分析に関する発表を始めさせていただきます。よろしくお願いします。

始めに、本研究の研究背景を説明します。従来の工場の運用管理では、作業員の巡回により工場内の機器の点検業務を行っていました。しかしながら、作業員の人員コストや目視確認による人的ミスの発生などが課題でした。産業用モニタリングシステムは、この点検業務を自動化します。これにより、人員コストの削減や人的ミスの低減に加え、リアルタイムなデータの利活用などの効果が期待できます。

システムの概要はこのようになっています。まず、工場に設置された機器の稼働情報をリアルタイムで収集します。収集したデータは、キャリア回線を通じてクラウドサーバに送信されます。そして、クラウドサーバ上でデータ処理を行い、運用管理担当者に提示します。

しかしながら、長期間の運用のなかで、パケットロスや通信の甚大な遅延などの障害は避けられません。このような障害が発生した場合には、工場内の機器の稼働状況を把握することが困難となります。それにより、工場の稼働停止や業務の遅れなどが引き起こされ、大きな損失が生じる可能性があります。

障害発生時には障害の迅速な発生検知や予測が求められます。しかし、障害の原因や内容、規模はさまざまです。このような中で、我々の研究グループでは、産業用モニタリングシステムにおける、さまざまな障害の発生検知とその発生箇所の特定に取り組んでいます。

この発表ではその様々な障害のうち、工場内の無線機器からクラウドサーバまでの通信路で発生する、通常とは異なる振る舞いの遅延を対象とします。そこで、通常の遅延の傾向を把握するために応答遅延を計測し分析を行いました。

以降は本研究で行った計測と分析について説明します。分析は二つに分かれます。一つ目は時系列モデリングによる分析です。ここでは、使用する時系列モデルとモデルの次数設定、回帰結果について述べさせていただきます。二つ目はクラスタリングによる分析です。ここでは、パラメータの前処理とクラスタ数の決定、クラスタリング結果について述べさせていただきます。

計測方法を説明します。計測環境は、実際の産業用モニタリングを模擬しました。モニタリングシステムにおける工場内の無線機器として Raspberry Pi を用いました。Raspberry Pi は大阪大学敷地内に設置しました。また、クラウドサーバとして一台の AWS サーバを用いました。Raspberry Pi と AWS サーバは、LTE 回線で接続されています。この環境で、Raspberry Pi 上で 15 秒毎に、時刻取得と ping を用いた応答遅延の計測を行いました。また、通信時間帯が応答遅延に与える影響を調べるため、計測は 3 時、7 時、12 時、17 時、20 時のそれぞれ 1 時間において行いました。計測期間は 4 週間としました。

次に分析方法を説明します。先の計測によりこのような時間変動をする応答遅延の実測値の区間データが計 122 個得られます。さらに以降の分析では、実測値の各区間データに対して、計測時点間の応答遅延の差分である変動値の区間データも対象とします。これらに対して二つの分析を行いました。一つ目は時系列モデルによる回帰分析です。これにより、各区間の実測値や変動値の変化の仕方を分析します。二つ目はモデルパラメータにもとづくクラスタリング分析です。これにより、応答遅延の曜日や時間帯に依存した傾向を分析します。

まず、時系列モデルによる回帰分析について説明します。時系列モデルには、古典的なモデルの一つである、ARMA-GARCH モデルを用います。実測値の区間データに対して、時刻 t における推定値は、定数項 c と過去の p 時点前までの実測値、 q 時点前までの誤差のそれぞれの重み付き和とノイズ項によって表されます。また、ノイズ項は平均 0、分散 h_t の独立同一な正規分布に従います。この時刻 t における分散は、定数項 ω と過去の r 時点前までの誤差、 s 時点前までのノイズ項が従う分散のそれぞれの重み付き和によって表されます。時系列モデルによる回帰では、適

切な次数 p, q, r, s のもとでパラメータ $a_i, b_i, c, \alpha_i, \beta_i, \omega$ を算出します。なお、実測値と変動値の時系列データはともに定常性を満たしていました。

次に次数の設定について説明します。本来、各区間データごとに適切な次数 (p, q, r, s) が存在します。しかし本研究では、モデルパラメータをもとにクラスタリングを行うため、全区間データで共通の次数を用いることとしました。これには、AIC をもとに各区間データごとに定まる最適な次数 p, q, r, s を、それぞれの最大値で設定します。 p, q は 0, 1, 2, r は 1, s は 0, 1 とした各組み合わせに対して共通次数を検討しました。その結果、実測値に対する共通の次数は (2, 2, 1, 1) となり、変動値に対する共通の次数は (2, 2, 1, 1) となりました。

このもとで行った回帰結果について述べさせていただきます。図は、3 月 2 日月曜日の 3:00～4:00 において得られた実測値に対する回帰結果です。以下の内容は他の区間データでも言えるため、この区間データを用いて説明させていただきます。図より、赤線で描かれた推定値は、青線で描かれた実測値の細かな変化を捉えられていないことがわかります。特に、最低応答遅延時間と考えられる 40ms や突発的に発生する 100ms 以上の大きな応答遅延を推定できていません。しかし一方で、ずれはあるものの平均的な振る舞いにおおよそ追従していることがわかります。つまり、時系列モデリングにより中期的な応答遅延の変化の仕方を捉えることが可能だと考えられます。

次に変動値に対する回帰結果です。先と同様に、一つの区間データの変動値を用いて説明させていただきます。図より、スパイク状の応答遅延の増加時には赤線の推定値と青線の変動値との差が大きくなっていることがわかります。これは、この突発的な応答遅延の増加は直前の応答遅延と無関係に発生するためだと考えられます。一方で、その直後の減少時には赤線の推定値と青線の変動値との差は小さいことがわかります。このように、実測値よりも変動値の方が推定しやすいと考えられます。表は実測値と変動値を用いた場合のそれぞれに対する、平均二乗誤差とノイズ項を加味した対数尤度の平均を示しています。表より確かに、僅かながらも実測値よりは変動値を用いた方がよい回帰精度を示すことがわかります。

また、次にモデルパラメータから応答遅延の変化の仕方を特徴づけ、比較することで、異なる振る舞いを読み取ることが可能であることを示します。例として、同じ時間帯、曜日の 3 月 2 日月曜日 12:00 – 13:00 と 3 月 9 日月曜日 12:00 – 13:00 のそれぞれの変動値に対するモデルのパラメータを表に示します。表より、3 月 9 日の方が a_1 と a_2 の絶対値が大きいことがわかります。 a_1, a_2 は、過去の変動値に対する重みです。したがって、3 月 9 日の応答遅延は過去の変動値の大きさが現在の変動値の大きさに与える影響が大きいと考えられ、この点で、3 月 2 日と 3 月 9 日は異なる振る舞いを示していたと思われます。

以上より、時系列モデリングにもとづいて異常を定め、検知を行う際には、以下のことが考えられます。まず、回帰精度のよさから時系列モデリングには変動値を用いるのがよいと思われます。また、異常検知手法としては、あらかじめ算出したパラメータによって推定される変動値と実測によって得られる変動値を比較し、その差が大きい状態が継続した場合を異常とする手法が考えられます。あるいは、リアルタイムで逐次的な回帰を行い、パラメータが大きく変化した場合を異常とする手法なども有効と考えます。

続きまして、モデルパラメータに基づくクラスタリング分析について説明します。まずはクラスタリングパラメータの前処理についてです。クラスタリングパラメータは、実測値または変動値の区間データに対するモデルパラメータを基とします。始めに各パラメータの分布が異なるため、平均 0、分散 1 となるように標準化を行いました。その後、主成分分析を行い次元を圧縮しました。次元圧縮の目的は、次元数の多さがクラスタの過度な細分化を引き起こすことを防ぐためです。また、主成分分析による累積寄与率は、実測値と変動値のいずれにおいても、第三主成分でおおよそ 80% となっていました。累積寄与率 80% とは、一般的に各データの特徴づけができてい

ラインとされています。したがって、実測値と変動値を用いる場合ともに第三主成分まででクラスタリングを行います。

次にクラスタ数の設定についてです。クラスタリングを行うためには、適切なクラスタ数を事前に定める必要があります。本研究においては、クラスタリング指標の一つである Pseudo F に改良を加えた Pseudo F with Min を用いて定量的に最適なクラスタ数を決定します。この指標の値は、クラスタ内分散に対するクラスタ間の距離の比として与えられます。したがって、クラスタ内が密で、クラスタ間が疎となる場合に大きな値となります。この図は、実測値と変動値を用いた場合のそれぞれに対し、クラスタ数を変化させながら Pseudo F with Min を求め、クラスタの良さを評価した結果です。図より、実測値の場合にはクラスタ数 8 で、変動値の場合にはクラスタ数 4 でそれぞれ評価値が最大になることがわかります。

また、クラスタリングには、階層クラスタリング手法の一つであるウォード法を用い、距離関数はユークリッド距離を用います。このもとで行ったクラスタリング結果について述べさせていただきます。まずは、実測値を用いた場合のクラスタリング結果についてです。図は、時間帯や曜日で区切ったクラスタごとの積み上げ棒グラフになります。いずれのクラスタにおいても属する区間データの時間帯や曜日に偏りがあることがわかります。まずは、クラスタ 1 とクラスタ 3 に注目します。これらのクラスタには、移動通信トラヒックが増加し、ネット利用の行為者率が最大となる 20 時台と移動通信トラヒックが多い 12 時台が多く属していました。さらに、曜日では、一日を通じてトラヒック量が多い土日が多く属していました。以上のことから、利用者の多い状況での応答遅延には似た傾向があると考えられます。

次に、クラスタ 4 に注目します。クラスタ 4 には、通信トラヒックが少ない 3 時台の区間データがクラスタの要素数に対して比較的多く含まれていました。また、一日を通じた通信トラヒックが少ない月曜日を除く平日の区間データも比較的多く属していました。このことから、利用者の少ない状況での応答遅延に表れやすい傾向があり、結果としてクラスタ 4 に属する区間データに偏りが生じたと考えられます。

次に、クラスタ 5 とクラスタ 7 に注目します。クラスタ 5 には、一日を通じたトラヒックの変化が激しい水曜日と金曜日の区間データが多く属していました。一方でクラスタ 7 には、一日を通じてトラヒック量の変化が緩やかな土日の区間データが含まれていませんでした。このことから、通信トラヒックの変化が激しい応答遅延には似た傾向があると考えられます。

最後に、変動値を用いた場合のクラスタリング結果について述べさせていただきます。図より、多くの区間データがクラスタ 1 とクラスタ 2 に属していることがわかります。しかし、曜日や時間帯に応じた顕著な特徴は見られません。これは、変動値の系列は実測値の大きさによらず 0 ms を中心に変化することから、似た傾向を示しやすいためだと思われます。しかし一方で、要素数が少ないクラスタ 3 とクラスタ 4 に属する区間データは、通常時と何らか異なる振る舞いをしていたのではないかと考えられます。

以上のクラスタリング結果から、応答遅延の変動の仕方に曜日や時間帯による移動通信トラヒックの特徴に応じた傾向が存在する可能性が考えられます。

したがって、クラスタリングにもとづいて異常を定め、検知する際には、以下のことが考えられます。まず、クラスタリングには傾向を捉えやすい標準化後の実測値のパラメータの主成分を用いるのがよいと思われます。また、検知手法としては、予備計測に基づいたクラスタリングによりクラスタの代表点を求めておき、その代表点に対応するモデルパラメータを用いて回帰を行うことで推定される実測値と運用環境における実測値が継続して大きく異なる場合を異常とする手法が考えられます。または、逐次的に時系列解析とクラスタリングを行い、区間データが曜日や時間帯による傾向を反映したクラスタとは異なるクラスタに収容され続けた場合を異常とする手法も有効と考

えます。

本研究では、産業用モニタリングシステムでの障害の検知に向けて、時系列モデリングやクラスタリングによって応答遅延の特性分析に取り組みました。その結果、時系列モデリングの実測値では中期的な応答遅延変動を捉えられることを明らかにしました。また、変動値に対しては実測値の場合に比べるとわずかながらも平均的に精度良く回帰を行えることを明らかにしました。また、標準化後の実測値のパラメータの主成分にもとづいてクラスタリングを行うことにより、応答遅延の変化の仕方に曜日や時間帯による移動通信トラヒックの変化に応じた傾向があることを示しました。

今後の課題としては、平時と異なるネットワーク利用状況での応答遅延の分析や異常検知手法の設計と検証、また、他の検知手法と組み合わせた精度向上などがあげられます。

以上で私の発表を終えさせていただきます。ご清聴ありがとうございました。