NaITE#14 開催レポート「メトリクス解析（データ解析）の初歩」開催レポート

2016 年 5 月 2 9 日

報告者：氏田　孝幸（NaITE 運営スタッフ）

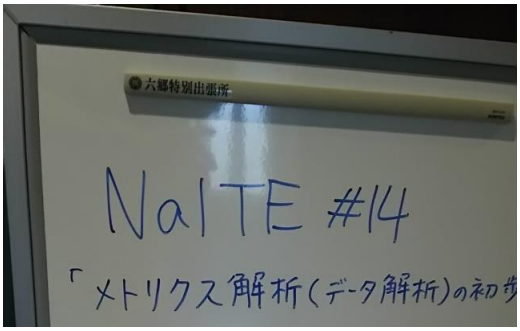
1 開催概要

2016 年 5 月 29 日　NaITE#14 を以下のように開催しました。概要は次の通りです。

開催日時	2016/05/29（日） 14:00 ～ 16:00		
開催場所	六郷地域力推進センター		
メインテーマ	メトリクス入門		
キーワード・タグ	メトリクス データ分析		
プログラム	14:00～14:10	オープニング「開会のご挨拶」	NaITE 運営スタッフ
	14:10～15:10	セッション「データ解析入門」	岡野 麻子氏
	15:10～15:20	休憩	
	15:20～16:00	セッション「データ解析 演習」	岡野 麻子氏
	16:00～16:10	クロージング「閉会の挨拶」	NaITE 運営スタッフ
受付サイト	http://nagasaki-it-engineers.connpass.com/event/31354/		

©NaITE　All rights reserved.
http://naite.swquality.jp/

2 勉強会の様子



今回はメトリクス解析入門をテーマに開催しました。

前回#13 に続いてほぼ満員となり，盛況な勉強会となりました。

今回が NaITE 初参加という方も多く，質疑応答が活発に行われていたことからもテーマへの関心の高さが伺えました。

2.1 オープニング

オープニングでは NaITE スタッフから NaITE についての紹介とプログラムの説明をいたしました。また，10 月 7 日（金）に長崎ブリックホール（長崎市）にて開催する長崎 QDG2016 について案内をいたしました。

また，7 月 30 日（土）に「Agile Japan 長崎サテライト with NaITE」をメルカつきまち（長崎市）にて決定した旨も告知いたしました。(本日参加受付ページオープン！)

その他現在活動中の SIG についての案内も実施いたしました。

2.2 セッション 1「データ解析入門」

セッション 1 は「ソフトウェアメトリクスについて」と題して，岡野氏に講義いただきました。

事前知識の無い人でも理解できるレベルと紹介があった通り，統計を使う「目的」や，統計で使われる用語(箱ひげ図，単回帰分析など)の説明などが，順を追って理解できるように構成されていました。

詳細な説明は発表資料に記載されていますので，ここでは一部を抜粋します。

- 四分位…データを昇順に並べて 4 等分したもの
- 中央値…平均がばらついている場合は，少数の大きな値に引きずられることがある。中央値ではそういった影響がほとんどない。
  - ※当日の発表では，中央値と平均値がよく使われる場面についての補足説明もされました。
- はずれ値… 箱ひげ図の両端から四分位範囲の 1.5 倍以上離れたデータ。
  （参考文献：データ指向のソフトウェア品質マネジメント）
- ばらつき…実力が一定であることが良いとされている。

©NaITE　All rights reserved.
http://naite.swquality.jp/

- 「ばらつき」が大きい場合は，ばらつきを小さくしていき(一定にする)，その後実力を上げていく流れがみられる。

- 回帰分析
  - 独立変数と従属変数の間の関係を推定するための統計的手法のこと。
  - 散布図に使われることが多い。

■ケーススタディ 1「品質指標を計画する」

品質指標を計画するということで，実際のデータ解析の流れを紹介していただきました。手順としては「①データ収集→②実力の把握→③指標値を決定」という流れになることが説明されました。

データ収集段階では，計画するプロジェクトと類似性の高い(※)プロジェクトのなかで「成功」の定義を決めて，データを集めていく必要があります。（類似性が高い： PM が同じ／参画メンバーが同じ／環境・場所が同じなど）

「成功」の定義は，組織ごとに変わってくるので注意が必要とのことでした。また，データの取得方法などが統一されていることも重要であるとされました。箱ひげ図を書くことで，どの辺りにばらつきが多いのか見極められるようになるそうです。

実力の把握，指標値の決定については，事前配布のサンプルデータ(Excel ファイル)を使用して実際に算出するところを実演いただきました。

■ケーススタディ 2「相関を見る」

メトリクス同士の関係を見る時には「平準化」を行うことが多く，特に開発規模やドメイン毎の特性などに引きずられることの多いメトリクスを使用する場合は，重みを考慮したり，開発規模で割ったりすることによって，異なる特性のものを同じ土俵に乗せて解析することができるようになるとのことでした。

異なる特性のものを解析する場合，そのメトリクス同士・データ同士の関係性，データの持つ意味を考えて「どういう回帰分析を行うのが最適なのか」を最初に考えることが必要だとされました。

データ同士の微分量を見ることで，どういう関係性にあるのかを把握できることもあると説明されました。

今回の講義では「メトリクス解析（データ解析）の初歩」ということで初歩的な内容を中心にかみ砕いて説明いただきましたが，普段聞きなれない単語が多いためか，理解に苦戦している参加者もおもられたようです。今回の勉強会を機に，習得に励んでみてはいかがでしょうか。

©NaITE　All rights reserved.
http://naite.swquality.jp/

◆質疑応答◆

Q.メトリクスは何のために使うものなのか。

A.今は品質指標をとって見ることに注目している。（メトリクスをとってあるという前提）。PSP（パーソナルソフトウェアプロセス）を実践すると，自分自身の作業時間，単位時間あたりの作業数などを測ることにより，苦手分野などが把握でき改善に繋げることができる。自分の日常もメトリクスと捉えて考えて生活していくと楽しいと思う。PSP に関しては，インストラクタなどもいると思うので，色々調べてみると良い。

2.3　セッション２「データ解析 演習」

セッション２は「ソフトウェアメトリクスの関係を導出しよう(単回帰分析演習)」と題して，岡野氏に演習とディスカッションを実施いただきました。

演習は東京と福岡の日々の気圧の差をまとめたサンプルデータ(Excel 形式)を用いての実施でした。(Excel に分析ツールを入れておくことが前提となります)

今回の演習では東京の気圧に応じて福岡の気圧がどう変わるかを見ることにしていました。Excel により分析出力した結果，重相関が約 80%であったため，相関があると判断することができるとのことでした。

また，メトリクスにおいては GQM という計測の枠組みが成り立つと説明されました。

- Goal →なんのために
- Question →どういうことを見たいのか
- Metric →何をアウトプットに

ここからは，チームごとのディスカッションを行いました。GQM の理解を最終目的とし，

- どういうメトリクスをプロジェクトでとっていますか？
- また，それでどういうことを把握したいと思いますか？

を話し合うというものでした。

議論では様々なことが話題になりました。以下にいくつか紹介します。

©NaITE　All rights reserved.
http://naite.swquality.jp/

○A チーム

取っているメトリクス：EV，バグ数・工数

どういうことを把握したいか：実際にプロジェクトで何が測れる，測れないかが気になる。

○B チーム

取っているメトリクス：レビュー指摘数，レビュー時間，あとは A チームと同じ。

どういうことを把握したいか：ドキュメントの数はあまり測らない。最適な量が見えないからだと思う。

※件/page，件/kloc の変換が出来ないと，差が大きかった場合に件/page の妥当性が疑わしくなってしまう。

○C-1 チーム

取っているメトリクス：バグ件数とか，統計としてとっているが 1 プロジェクトで閉じていて他に展開できていない。

どういうことを把握したいか：組織として貯めこまれていないという問題がある。

○C-2 チーム

メトリクスを取っていない。

これから取っていく上でどうするかという話だが，収集する情報が個人情報なので活用が難しい。特に広告等になると，ビッグデータ等の方向に進んで行ったりする。（大手企業がビッグデータとして情報を握ってしまう）

Web ページなどではアクセスした人の年代等が溜まっていくので，活用できることもあるのではないか。雑誌など紙媒体だと想像が難しい。

○D チーム

これから分析をしていきたい。バグの分類はしているが，そこから先の分析ができていない。

©NaITE　All rights reserved.
http://naite.swquality.jp/

テスターの能力のカテゴライズをしてみたいと思っている。（スキルを計測してレーダーチャートを作るなど）そこから活用していきたい。

スキルをメトリクスにすると，判定する人によってばらつきが出てしまう。そういったところを揃えるのも，ばらつきを減らすことに繋がる。

○E チーム

取っているメトリクス：工数，バグ出現率。プロジェクトの予算など。

どういうことを把握したいか：発表を聞いて，業務でメトリクスを測っているつもりでいたが，GQM への意識が薄かったためか，ただの数字遊びになっていたかもしれないと感じた。改めて，GQM を意識してメトリクスと向き合ってみたいと(現場を巻き込んでみたい)と思った。

2.4　セッション３「最後に…」

セッション３は「最後に…」と題して，岡野氏からアドバイスやメッセージをいただきました。

- 箱ひげ図だけで判断しないこと。
- あくまで母集団のなかの標本であることを意識する。
- データの偏りを考えて分析をするならば，正規分布をする母集団なのか否かを考慮する。
- 標本数が少ない場合は，t 分布を適用することが大事である。

また，以下について補足の説明をいただきました。

- 大数の法則
  - 大標本では，観察された標本平均を母集団の真の平均とみなしても良いという常識を数学的に証明したもの。
- 中心極限定理
  - 確率分布の形は， n が大のときは，だいたい正規分布と考えてよい
- 二項分布
  - コインを投げるときなどの確率分布に適用される。

©NaITE　All rights reserved.
http://naite.swquality.jp/

3 質疑応答

Q.配布された Excel の２枚目と１枚目のシートの使い方が分からない。

A. １枚目の値を参考に，２枚目の値を作る。その結果として四分位の値が自動的に作られているところを見ていく。

Q.Excel で数値を出したとしても，その数字をどう解釈したら良いのか分からない。人によって見方が違うこともあると思う。

A.いろいろなデータを見て，傾向を自分でつかむようにする。生産量とか，業務によって違うものがある。

生産性などある程度お決まりのデータは，IPA でデータ白書として公開されているので，こちらを参考にすると良い。IPA が年２回程度，セミナーを開催している。ここで見方や活用方法を知ることができる。

Q.普段の生活から考えた方が良いということだが，実際そこからどういう風にデータをとっているのか想像が出来ない。普段の生活から取るときの考え方は例えばどういったものなのか。

A.例えば，家庭では「実際にどれだけの時間で宿題ができるか」など，時間を測っている。集中力の有無や，問題の傾向によって時間に差が出る。受験のときも含め，かかった時間の多い・少ないから箱ひげ図を書いて，目標校の問題の出題傾向的に合格可能性はどうかなどを見ることがある。どれくらい外で遊んだかとご飯の食べ具合がどうかなどを普段から考えている。

Q.箱ひげで書くのに適したものは？

A.工数は箱ひげで出したりする。フェーズごとにどのくらいの割合がかかっているかとか，比率にするとわかりやすいことがある。何かを何かで割ったもの（比率）は傾向を示せるので使いやすい。そのままの値を使ってしまうと，対象のプロジェクトの特性自体に左右されてしまうので，平準化して比率として扱った方が良い。

Q.ばらつきを扱う手段として箱ひげ図とヒストグラムのどちらを使うかの判断は？

A.箱ひげ図は範囲を知りたい場合に用いる。ヒストグラムは，どの辺りが多いか把握したい場合に使う。それぞれ使う目的が違うので，それぞれ使い分けてほしい。

©NaITE　All rights reserved.
http://naite.swquality.jp/

Q.箱ひげ図の第１と第３四分位で囲まれた範囲は２５～７５%の範囲なので，全体の５０%ということになる。ということは，指標範囲と設定すると，2 回に 1 回はずれるということになるが，どう考えるか？

A. この質問は，実際に CMMI レベル４のアセスメントでも聞かれることがある。なかなか回答が難しいが，通説のような答えをご存じの方は，ぜひアドバイスがほしい。

Q.指標の取り方。各プロジェクトで全く違う指標でやっていたので意味の無い値を撮り続けていたことがあった。何か良い方法はないか。上司を説得するのに困っている。

A.例えば，IPA のデータを使うことがある。

同じ項目で測っても，全然違う結果が出ることがある。（例えば Web 系と組み込み系の違いなど）

ドメインを意識して切って，ドメイン別に事業を分けて比較できるようにすることが重要。

全ドメインを比較するときは補正をかけるようにする。ひとつのデータも見るドメインによって，見る重みも変わるので，重みづけが大事になる。重みの妥当性がどうかという話になると，過去の論文で言及されていることがあるので，そちらを参照するとよい。

4 参考情報・リンク等

当日の発表資料等は以下をご参照ください。

- http://www.slideshare.net/asakoyanuki/naite14

5 次回の予定等

次回は 6/25「えくす・でぃ・でぃ・ぴぃ概論＆入門ワークショップ」をテーマとして開催の予定です。

派生開発カンファレンス 2016 にて行なわれたワークショップセッション「えくす・でぃ・でぃ・ぴぃ概論＆入門ワークショップ」を，八木将計様（日立製作所）に再演していただきます。ぜひ参加をご検討ください。

以上

©NaITE　All rights reserved.
http://naite.swquality.jp/