

New Twitter Analysis Maniax

[大改訂増補版]

rtweet, R, Excel VBA, KH Coder による

真に最強 (?) のツイッター分析

後藤和智

(後藤和智事務所 OffLine)

New Twitter Analysis Maniax

[大改訂増補版]

rtweet、R、Excel VBA、Access による
真に最強（？）のツイッター分析

著：後藤和智（後藤和智事務所 OffLine）

発行：2022年8月13日

（コミックマーケット 100）

注意

本書を著作権法の定める私的使用の範囲外で公開などを行うことを禁じます。また、本書の使用により生じた問題についての責任は負いかねます。

New Twitter Analysis Maniax

rtweet、R、Excel VBA、KH Coder による真に最強 (?) のツイッター分析

0.1 はじめに

お久しぶりです。最近は分析も創作も疎かになっている後藤和智です。思えば、2019年に開催されたコミックマーケット 97 から 2021年 12月下旬現在に至るまで、まとまって書いた文章は「現代ビジネス」の2本と「wezzy」の1本だけで、同人関係はブログ記事や電子書籍も含めて全くと言っていいほど書いていないんですよね……。本当にこの期間の怠惰を猛省したいです。本書も締め切り直前に新刊 RTA 状態で書いていますし。やっぱり同人作家には大型イベントと締め切りが必要なんだ……。

さて本書ですが、私は2019年の夏コミ（コミックマーケット 96）で、『Twitter Analysis Maniax : twitteR、Excel VBA、KH Coder による最強 (?) のツイッター分析』という同人誌を出しています。この同人誌のサブタイトルで「最強 (?)」というタイトルを付けていましたが、実を言いますと同書の執筆時点で、同書で用いた、フリーソフト「R」のパッケージ「twitteR」は、「最強」とは言えないということを知っていました。ただ、今更書き直すというのもできないし、またパッケージの存在を知ったのは執筆中で、まだ使い慣れていなかったの、言い逃れができる形で「最強 (?)」としていたのです（ちなみに本書のサブタイトルも「真に最強 (?)」としていますが、それは現時点で最強だと思うパッケージを紹介しておりますが、実はもっと「真に真に最強」のツールがあるかもしれないという理由です）。

本書で用いるパッケージは「rtweet」です。実は『Twitter Analysis Maniax』を書いていた時点で、twitteRは既に時代遅れ、これからはrtweetの時代という雰囲気は既にありました。また、Rによるテキストマイニングの解説書の定番である石田基広『Rによるテキストマイニング入門』（森北出版、初版2008年、第2版2017年）においては、2017年に発行された第2版において、Rによるツイッター分析のツールとしてtwitteRが紹介されていましたが、2020年3月に出た、同じ著者の『実践 Rによるテキストマイニング』（森北出版）では、twitteRは見るともなく、紹介されていたのはrtweetでした（恐らく同書が商業ベースの書籍での、本格的なrtweetの解説が書いてある書籍だと思います）。

後で詳しく述べますが、rtweetの、専門的な研究者ではないレベルの、twitteRとは違う重要な点は、第一に「APIの取得が必要ない」「twListToDFによるデータの整形が必要ない」、さらにWindowsユーザーにとってはこれも極めて重要である「文字コードの変換が必要ない」というものがあります。そのため、Rをインストールしてそしてパッケージをインストールすれば、すぐに分析を行うことができるという優れたものです。ただし、長期間のデータを取得したいなら、研究者向けのAPIを取得する必要がありますが、少なくとも同人誌を書いたりあるいは雑誌やニュースサイトなどに社会問題を取り扱った文章などを書くレベルにおいては「工夫でなんとかなる」レベルです。

他にもいくつか「必要ない」ものはありますが、ただ、rtweetで取得できるデータの関係上、私の説明できる範囲では活用するのが難しいところもありますので、そういう部分においては前著『Twitter Analysis Maniax』を参照していただくか、もしくは今後の加筆（改

版)にご期待ください……。

あと、2年振りのコミックマーケットということで、2016年に出したテキストマイニングの入門的解説書である『Text Mining Maniax』を、『Text Mining Maniax Advance』の一部の章を移植して増補版としてリリースしております。この2冊を持っていない方は是非とも読んでね。

0.2 大改訂増補版まえがき

というわけで本書は、2021年12月の「コミックマーケット99」で出した同人誌『New Twitter Analysis Maniax』の大改訂増補版になります。完全版商法って言わないでください。

正直、旧版は久々の同人誌ということもかなり大急ぎで書いた代物なので、いろいろと不足しているところも合ったのですが、今回は書き忘れたところも含めて大幅に追記しているので、ある程度は完璧に近いものになったかと思います。

ただ、rtweetを解説した本書によって『Twitter Analysis Maniax』が完全に用済みになるかという、そうでもありません。Rの外の操作、特にExcelやKH Coderの操作に関しては、同書だけで紹介しているものもあるので、是非ともそちらも読んでいただけると幸いです。

今回も旧版ほどではないものの新刊RTA状態で大急ぎで執筆しましたが、それでも手抜きをしたつもりはありませんし、一つの新しい技術書としてお楽しみください。

New Twitter Analysis Maniax

rtweet、R、Excel VBA、KH Coder による真に最強 (?) のツイッター分析

目次

- 0.1 はじめに 2
- 0.2 大改訂増補版まえがき 3

第 1 章 rtweet の導入.....6

- 1.1 はじめに 6
- 1.2 rtweet を始める 6
- 1.3 rtweet の代表的なコマンド 8
- 1.3 rtweet で取得できるデータ 10
- 1.4 必要なデータだけを取り出したい場合 11
- 1.5 時間について 13

第 2 章

絵文字を Excel VBA で使うための基礎知識..... 15

- 2.1 はじめに 15
- 2.2 絵文字の文字コード 15
- 2.3 絵文字を Excel VBA で扱うために 17
- 2.4 Excel で絵文字を表示してみる 17
- 2.5 では Excel VBA では? 18

第 3 章 リツイート調査 1：データの取得..... 19

- 3.1 はじめに 19
- 3.2 コマンドの作成 19
- 3.4 リツイートデータの集計と集計 28
- 3.5 データの絞り込み (Access は使いません!) 30

第 4 章 リツイート調査 2：データの分析 32

- 4.1 はじめに 32
- 4.2 多次元尺度構成法による分析 32
- 4.3 クラスタ分析 36

4.4 因子分析 39

第5章 プロフィールの変換と単語の集計.....42

5.1 はじめに 42

5.2 プロフィールの変換 42

5.3 KH Coder を用いた集計 47

5.4 決定木分析 49

New Twitter Analysis Maniax

rtweet、R、Excel VBA、KH Coder による真に最強 (?) のツイッター分析

第 1 章 rtweet の導入

1.1 はじめに

rtweet マン!

rtweet は Twitter API 取得が必要ありません!

さらに、twListToDF によるデータの整形も必要なし!

さらにさらに、なななんと! 検索時に文字コードの変換も必要ないのです!!

こんなに便利でいいんでしょうか!?

ムムッ!!!

さあ、今すぐ R と rtweet をインストールしてツイッター分析!!

はい、というわけで某クレジットカードの CM 風に紹介してみました。とりあえずこれだけでも rtweet が twitteR を駆逐してしまっている (主に私の中で) 理由がわかると思えます。

ここに採り上げたように、rtweet は前著『Twitter Analysis Maniax』で紹介したツイッター分析ツール「twitteR」と同じ、フリーの統計ソフト「R」のパッケージとして提供されているツイッター (Twitter) のデータ取得ツールですが、(専門的な説明になってしまって申し訳ないですが) twitteR で取得したデータは List 形式で提供されるのに対して rtweet はデータフレームとして提供されます。twitteR ではその過程で多くのデータが犠牲になってしまうので、取得したデータをそのまま使える rtweet は分析者としてはとても魅力的です。

rtweet を始めるには、R があればいいです。なお、R のインストールに関しては本書において既に行われているものと見なして進めていきますので、ご了承ください。

1.2 rtweet を始める

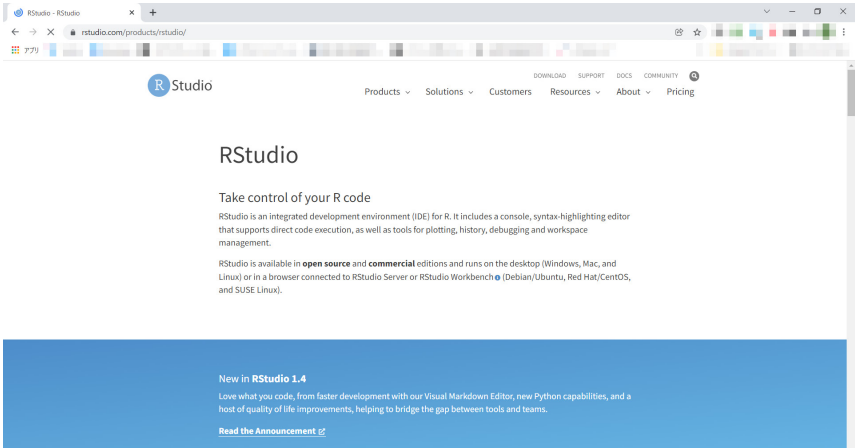
R をインストールしたところで、早速 rtweet を使しましょう。まず、R を起動させて、パッケージをインストールします。パッケージのインストールの方法はいろいろありますが、とりあえず rtweet だけインストールしたい場合はこのようにします (本当はできるだけ多くのパッケージをインストールした方がいいのですが……)。

```
> install.packages("rtweet")
```

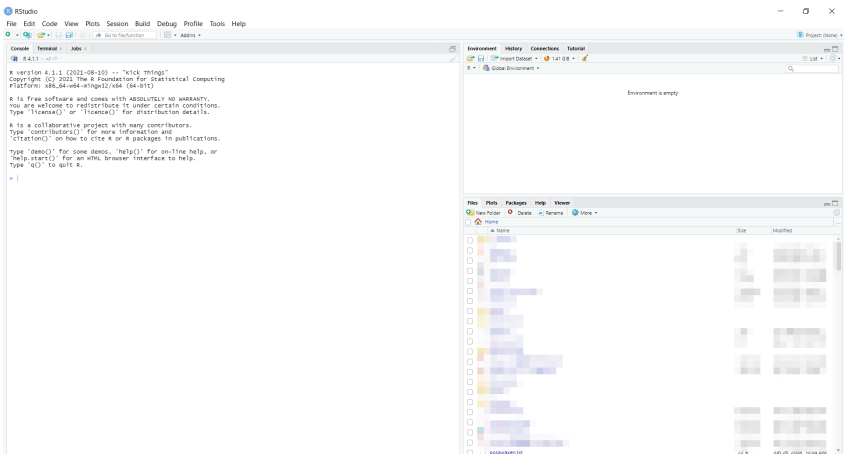
あと、R だけでなく、R の補助ツールである「RStudio」もインストールしておきたいです。というのも、本書では Excel VBA で多くのコマンドを出力するプログラムを作成するほか、それ以外にも (少なくとも私の行っている) 実際の分析では、テキストファイルに実行した

第1章 rtweet の導入

1.1 はじめに



RStudioの公式サイト



RStudioの画面

いプログラムを書いて、それをコンソールにコピー&ペーストをするということをよく行いますが、本家のRではとても時間がかかり、またRのソースファイル（拡張子が「.r」というテキストファイル）を読み込む方法だと、エラーがあるとそこで止まってしまうからです。特にリツイート調査（第2章参照）だと、検索に引っかけられないということも少なくないのですが、それがエラーと見なされてそこでコマンドの実行が止まってしまうのです。そういうことを避けるためにも、私はRStudioを使うことを推奨しています。

RStudioもフリーソフトです。インストールは同ソフトの公式サイト（<https://www.rstudio.com/products/rstudio/>）からインストールします。RStudioは1回インストールしておけば、あとはコンピュータにインストールされている最新のRを使ってくれるので、

New Twitter Analysis Maniax

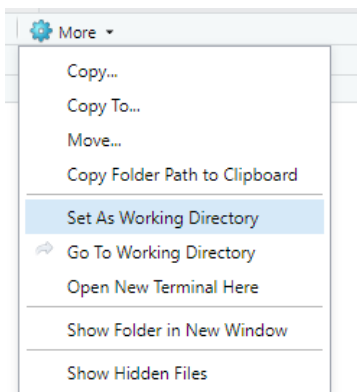
rtweet、R、Excel VBA、KH Coder による真に最強 (?) のツイッター分析

とても尊重します。ただし、念のため、新しいバージョンの R をインストールしたら、古いバージョンのものは削除しておきましょう。

なお、RStudio で作業を行う際は、あらかじめ作業ディレクトリを変更しておく必要があります。右下のファイル欄で作業したいディレクトリを閲覧している状態にして、「More」から「Set As Working Directory」を選択します。そうしないと、出力するデータは「ドキュメント」フォルダに延々と吐き出されることになります。

前節の冒頭でも述べたとおり、rtweet は、Twitter API (詳しくは検索してください) の取得が必要がありません。そのまま、コマンドを実行しましょう。なお、ここで取得できるデータは、概ね最新 9 日間のデータです。

rtweet を使えるようにするには、次のように入力します。



作業ディレクトリの変更

```
library(rtweet)
```

なお、CRAN (R のサーバー) にある rtweet の公式の解説書もあるので、そちらもご参照ください (<https://cran.r-project.org/web/packages/rtweet/rtweet.pdf>)。

1.3 rtweet の代表的なコマンド

rtweet には本当に多くのコマンドが存在します。中には「何のためのコマンドだ」と思うものもありますが、とりあえず、研究者ではない、記者・ライターや同人作家の調査レベルで使うものとしては次のものが挙げられます。

`search_tweets(検索ワード, n=取得数)` : 特定のワードを含むツイートを検索し、取得する。やろうとすれば通常の API 検索の範囲内でいくらでも取得できる。多分。

`get_timeline(スクリーンネーム, n=取得数)` : 特定のアカウントのツイートを取得する。最大 3200。

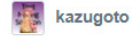
`get_favorites(スクリーンネーム, n=取得数)` : 特定のアカウントの「いいね」を取得する。最大 3200。

`get_retweets(ツイートの ID, n=取得数)` : 特定のツイートのリツイートを取得する。リツイートしたユーザーの ID だけを取得する `get_retseeters` もある。いずれも最大 200。

例えば、`search_tweets` で「錦鯉」というワードを検索して取得したいときは、次のようになります。

第1章 rtweet の導入

1.3 rtweet の代表的なコマンド



アプリケーションに戻ります。しばらくお待ちください。

アプリケーションを承認して……



こういう画面が出たら成功。データの取得が開始される

```
> dataset <- search_tweets("錦鯉", n=100)
Requesting token on behalf of user...
Waiting for authentication in browser...
Press Esc/Ctrl + C to abort
Authentication complete.
```

そしてブラウザが起動され、ツイッターに移行し、アプリケーションの承認画面が出ますので、そこでアプリケーションの連携を承認すると、文字だけの画面に遷移し、検索とデータの取得が始まります。なお、API 検索に必要なツイッターのトークンなどは、rtweet に内蔵されているものを使うため、このようになります。

さて、ここでは、検索ワードをそのまま「錦鯉」と入力していましたが、これが `twitterR` だと、文字コードを変換しなければならないので、`iconv("錦鯉", from="CP932", to="UTF-8")` というふうにしなければ鳴りませんでした。それが、rtweet ではそのまま入力できるので、なんとも便利になったものです。

rtweet の `search_tweets` によるデータ取得におけるもう一つの利点としては、一度に大量のデータを取得することができることです。Twitter API によるデータの取得は、1 回で 18000 個が限界なのですが、例えば、「アズールレーン」というハッシュタグで直近 30000 件のデータを取得したいときは、次のように入力します。

```
search_tweets("#アズールレーン", n=30000, retryonratelimit=TRUE)
```

この `retryonratelimit=TRUE` というのが、上限である 18000 を超えたデータを取得したいときのコマンドです。すると、次のような文章がでてきます。

```
> dataset3 <- search_tweets("#アズールレーン", n=30000, retryonratelimit=TRUE)
Downloading [=====] 100%
retry on rate limit...
waiting about 12 minutes...
Downloading [=====] 100%
```

New Twitter Analysis Maniax

rtweet、R、Excel VBA、KH Coder による真に最強 (?) のツイッター分析

```
> dataset3 <- search_tweets("#アズールレーン",n=30000,retryonratelimit=TRUE)
Downloading [====>-----] 13%
```

大量のデータを取得している

これは、まず、最初の「Downloading」が行われたあと、12分の休止を行い、そのあと再度さらに過去のツイートを取得するというものです。ちなみにDownloadingの後ろのゲージは進捗に応じてアニメーションします。これは、データを取得している途中の状態をコピーしたものです。

```
> dataset3 <- search_tweets("#アズールレーン",n=30000,retryonratelimit=TRUE)
Downloading [=>-----] 8%
```

他のコマンドに関しては、例えば私 (@kazugoto) のアカウントのツイートや「いいね」を取得したいときは、次のように入力します。

```
get_timeline("kazugoto",n=200)
get_favorites("kazugoto",n=200)
```

また、私のツイート、例えば (<https://twitter.com/kazugoto/status/1465676009252933636>) をリツイートしたユーザーを取得したいときは、次のように入力します。

```
get_retweets("1465676009252933636",n=100)
```

ツイートのURLの最後の数字をそのまま入力する必要があります。

1.4 rtweet で取得できるデータ

rtweet では、twitteR のように TwListToDF によるデータの整形の必要がないため、取得したデータをそのまま使用することができます。試しに私のツイートを取得し、中でも2021年12月に岩手県奥州市水沢にて開催された同人誌即売会「CRUSH162」の告知ツイートのデータを取得してみましょう。

```
> dataset <- get_timeline("kazugoto",n=200)
> dataset2 <- subset(dataset,status_id=="1469640903715614725")
```

これを出力したデータが表1になります。なお、ツイートが誰かのツイートに対するリプライの場合は、reply_to_status_id にリプライ対象のツイートが、reply_to_screen_name にそのアカウントのスクリーンネームが入ります。また、リツイートの場合は、「retweet_」で始まる箇所に、リツイートしたツイートのあらゆるデータが入ります。これはリツイートを分析する際に非常に有用ですので、どんどん使っていきましょう。

さらに、プロフィールに関するデータも持ってきてしまいます。まず name は名前 (スクリーンネームとは別の、プロフィールに表示されている名前)、location はプロフィールに

次のページではこのツイートのデータを見せます

使われている住所、description はプロフィールにある説明文です。その他フォロワー数や「いいね」の数、さらにはアカウントを作成した日時まで取得できます。

なお、先の引用文で使ったコマンド `subset` は、データテーブルから特定の条件を満たす行のみを抽出するコマンドです。

```
subset(データフレーム, 条件)
```

また、取得したツイートにおいて、ハッシュタグが複数使われているときは、hashtag の中身がベクトル形式になります。しかし、こうなると、CSV データに出力する際に支障が生じます。この活用法については、本書の続編ないし改版をご期待ください……。

1.5 必要なデータだけを取り出したい場合

そんなわけで、rtweet においては、このようにおよそ 90 にわたるデータを取り出すことができるわけですが、その中には（とりあえず記者やライター、同人作家レベルにおいては）無意味・不要なデータも多く、また有用なデータにしても、すべてを使うということは絶対ないでしょう。

必要なデータだけを取り出したい場合のコマンドは次の通りです。例えば dataset という文字に検索結果を格納していて、そこからツイートの ID と内容（本文）、ツイートされた時間と手段（パソコンかスマートフォンアプリか、それとも別の何かか）だけを持ってきた

New Twitter Analysis Maniax

rtweet、R、Excel VBA、KH Coder による真に最強 (?) のツイッター分析表 1

<https://twitter.com/kazugoto/status/1469640903715614725> のデータ

| | |
|-------------------------|---|
| user_id | 86502537 |
| status_id | 1469640903715614725 |
| created_at | 2021-12-11 12:11:10 |
| screen_name | kazugoto |
| text | 直前のお知らせになってしまいましたて申し訳ありません……！明日、岩手県奥州市水沢・Zホールにて開催される「CRUSH162」に、後藤和智事務所OFFlineはブロック14にて出展します。2年以上ぶりの岩手県です！よろしくお願いたします！ #CRUSH奥州 https://t.co/s3syw78clk |
| source | Twitter Web App |
| display_text_width | 127 |
| reply_to_status_id | NA |
| reply_to_user_id | NA |
| reply_to_screen_name | NA |
| is_quote | FALSE |
| is_retweet | FALSE |
| favorite_count | 1 |
| retweet_count | 3 |
| quote_count | NA |
| reply_count | NA |
| hashtags | CRUSH奥州 |
| symbols | NA |
| urls_url | NA |
| urls_tco | NA |
| urls_expanded_url | NA |
| media_url | http://pbs.twimg.com/media/FGU2YfrUcAIi0d.jpg |
| media_tco | https://t.co/s3syw78clk |
| media_expanded_url | https://twitter.com/kazugoto/status/1469640903715614725/photo/1 |
| media_type | photo |
| ext_media_url | c("http://pbs.twimg.com/media/FGU2YfrUcAIi0d.jpg", "http://pbs.twimg.com/media/FGU2YfrUUAU1-gRk.jpg", "http://pbs.twimg.com/media/FGU2YfrVcAMxxi1.jpg", "http://pbs.twimg.com/media/FGU2YfrVgAEqe0l.jpg") |
| ext_media_tco | c("https://t.co/s3syw78clk", "https://t.co/s3syw78clk", "https://t.co/s3syw78clk", "https://t.co/s3syw78clk") |
| ext_media_expanded_url | c("https://twitter.com/kazugoto/status/1469640903715614725/photo/1", "https://twitter.com/kazugoto/status/1469640903715614725/photo/1", "https://twitter.com/kazugoto/status/1469640903715614725/photo/1", "https://twitter.com/kazugoto/status/1469640903715614725/photo/1") |
| ext_media_type | NA |
| mentions_user_id | NA |
| mentions_screen_name | NA |
| lang | ja |
| quoted_status_id | NA |
| quoted_text | NA |
| quoted_created_at | NA |
| quoted_source | NA |
| quoted_favorite_count | NA |
| quoted_retweet_count | NA |
| quoted_user_id | NA |
| quoted_screen_name | NA |
| quoted_name | NA |
| quoted_followers_count | NA |
| quoted_friends_count | NA |
| quoted_statuses_count | NA |
| quoted_location | NA |
| quoted_description | NA |
| quoted_verified | NA |
| retweet_status_id | NA |
| retweet_text | NA |
| retweet_created_at | NA |
| retweet_source | NA |
| retweet_favorite_count | NA |
| retweet_retweet_count | NA |
| retweet_user_id | NA |
| retweet_screen_name | NA |
| retweet_name | NA |
| retweet_followers_count | NA |
| retweet_friends_count | NA |
| retweet_statuses_count | NA |
| retweet_location | NA |

| | |
|------------------------|--|
| retweet_description | NA |
| retweet_verified | NA |
| place_url | NA |
| place_name | NA |
| place_full_name | NA |
| place_type | NA |
| country | NA |
| country_code | NA |
| geo_coords | c(NA, NA) |
| coords_coords | c(NA, NA) |
| bbox_coords | c(NA, NA, NA, NA, NA, NA, NA) |
| status_url | https://twitter.com/kazugoto/status/1469640903715614725 |
| name | 後藤和智@冬コミ2日目N28a/技術書典オンライン予定 <U+0001F3F3><U+FE0F><U+200D><U+0001F308><U+0001F3F3><U+FE0F><U+200D><U+26A7><U+FE0F> |
| location | 宮城県仙台市 |
| description | 学術系個人サークル「後藤和智事務所OffLine」で若者論を統計的に研究したり、学術解説書を書いたりしています。お仕事依頼などはkgoto1984@nifty.comまで。アイコンの立ち絵はdairi様から。BOOTH: https://t.co/oeUNJPQnTW 反性差別、反人種差別、反トランスジェンダー差別 |
| url | https://t.co/5tLfh3oYzU |
| protected | FALSE |
| followers_count | 5456 |
| friends_count | 992 |
| listed_count | 403 |
| statuses_count | 140961 |
| favourites_count | 14452 |
| account_created_at | 2009-10-31 10:31:39 |
| verified | FALSE |
| profile_url | https://t.co/5tLfh3oYzU |
| profile_expanded_url | http://kazugoto.hatenablog.com/ |
| account_lang | NA |
| profile_banner_url | https://pbs.twimg.com/profile_banners/86502537/1355145135 |
| profile_background_url | http://abs.twimg.com/images/themes/theme1/bg.png |
| profile_image_url | http://pbs.twimg.com/profile_images/80351977004130304/zwlmhGhM_normal.jpg |

い場合は、次のようになります。

```
dataset <- dataset[, c("status_id", "text", "created_at", "source")]
```

注意してほしいのは、(Rの解説書では何回か言っていますが) Rはコマンドにおいて大文字と小文字を区別するという事です。そしてrtweetにおいては、大文字が使われている行は現状において存在しないので、大文字は入力しないようにしましょう。

1.6 時間について

最後になりますが、created_at (ツイートが投稿された時間) や account_created_at (アカウントが作成された時間)、retweet_created_at (リツイートしたツイートが投稿された時間) などといった、時間を示す項目について補足をします。

これらの項目は、行を操作する過程で数字になってしまうことがあります。例えば、先の表で示したツイートが投稿された時間は「2021-12-11 12:11:10」ですが、コマンドを使って行列を操作していると、「1639224670」という数字になることがあります。

rtweetに限らず、Rの時間のデータは、Excelとは違い、UNIXの時間で示されます。これは、1970年1月1日を0として、1秒ごとに1を足していく数値です。他方で、Excelの時間(シリアル値)は、Windows版の場合1900年1月1日を1として、1日ごとに1を足していく数値になっています。なお、UNIX時間とExcelのシリアル値の変換は、次のように行うようです。

New Twitter Analysis Maniax

rtweet、R、Excel VBA、KH Coder による真に最強 (?) のツイッター分析

(UNIX 時間 + 32400) ÷ 86400 + 25569

(税理士法人あさひ会計・旭ブレインズのウェブサイトより「UNIX Time から EXCEL のシリアル値 (日付・時間) に変換する方法」)

<http://asahi.gr.jp/wp01/1668>

第2章 絵文字を Excel VBA で 使うための基礎知識

2.1 はじめに

本書では、恐らくネット上にもあまりノウハウのない、文字コードが Shift-JIS の環境下における絵文字の使い方について解説します。

本書や弊サークルの統計解析の解説書では、R 言語の他 Excel VBA も多用します。本書にもサブタイトルに「Excel VBA」が入っていますし。

しかし、Shift-JIS では絵文字は表示できません。絵文字を表示するためには、Unicode という文字コードを使う必要があります。Excel では Unicode による表示に対応しております (Office 365 使用、2022 年 8 月現在)、また R のような UNIX 系のソフトでは文字コードは Unicode (UTF-8) になります。また、R からコマンド write.csv などデータを出力した際にも、Unicode で出力されますが、Windows (少なくとも Windows 10) の標準的な文字コードは Shift-JIS ですから、csv ファイルをそのまま Excel で開いた際には、文字化けを起こしてしまいます。本章では、元々 Unicode (UTF-8) で使われている絵文字を、文字コードが Shift-JIS である Excel VBA で使うための基礎知識を解説します。

2.2 絵文字の文字コード

図 2.1 は、本書で紹介する手法であるリツイート分析において取得したツイッターアカウントのプロフィール (前章における description) を、R のコンソールに表示したものです。印刷では少し潰れて見えづらいかもしれませんが、虹や時計、18 禁マークなどの記号のほか、ハングルも表示されており、絵文字や英語以外の外国語の文字などは、日本語だけに対応した Shift-JIS では表示されません。そもそもツイッターはアメリカで生まれたウェブサービスで、世界中に展開していますから、国際的な規格である Unicode を用いるのは必然と言えるでしょう。

Unicode は、それぞれの文字について「U+XXXX」という形式を割り当てております。Unicode の文字一覧のサイ

```

dat4_output$description[1:100]
[1] "抱いている"
[2] "ブラック、"
[3] "いきました"
[4] "..."
[5] "..."
[6] "..."
[7] "..."
[8] "..."
[9] "..."
[10] "..."
[11] "..."
[12] "..."
[13] "..."
[14] "..."
[15] "..."
[16] "..."
[17] "..."
[18] "..."
[19] "..."
[20] "..."
[21] "..."
[22] "..."
[23] "..."
[24] "..."
[25] "..."
[26] "..."
[27] "..."
[28] "..."
[29] "..."
[30] "..."
[31] "..."
[32] "..."
[33] "..."

```

図 2.1 取得したプロフィールを R で表示。絵文字も表示されている

New Twitter Analysis Maniax

rtweet、R、Excel VBA、KH Coder による真に最強 (?) のツイッター分析

トとして、例えば「Unicode 文字一覧表」(https://tools.m-bsys.com/ex/unicode_table.php) がありますが、これで私の名字の1文字目である「後」を検索すると、「U+5F8C」に割り当てられていることがわかります。ちなみにこの「5F8C」は16進数です。

| 凡例 | | 詳細テーブル | | | | | | | | | | | | | | | |
|---------|--------|--------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ラテン文字 | 記号 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
| 平仮名・片仮名 | 日本語記号 | U+5F0x | 并 | 弁 | 弁 | 弁 | 弁 | 弁 | 弁 | 弁 | 弁 | 弁 | 弋 | 弋 | 弋 | 弋 | 弋 |
| 漢字 | ハングル | U+5F1x | 武 | 武 | 武 | 武 | 武 | 武 | 武 | 武 | 武 | 武 | 武 | 武 | 武 | 武 | 武 |
| その他の文字 | サロゲート用 | U+5F2x | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 |
| 私用領域 | 未定義 | U+5F3x | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 | 未 |
| Zoom | | U+5F4x | 數 | 數 | 數 | 數 | 數 | 數 | 數 | 數 | 數 | 數 | 數 | 數 | 數 | 數 | 數 |
| 後 | | U+5F5x | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 | 丑 |
| 後 | | U+5F6x | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 | 禮 |
| 後 | | U+5F7x | 影 | 影 | 影 | 影 | 影 | 影 | 影 | 影 | 影 | 影 | 影 | 影 | 影 | 影 | 影 |
| 後 | | U+5F8x | 往 | 往 | 往 | 往 | 往 | 往 | 往 | 往 | 往 | 往 | 往 | 往 | 往 | 往 | 往 |
| 後 | | U+5F9x | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 |
| 後 | | U+5FAx | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 |
| 後 | | U+5FBx | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 |
| 後 | | U+5FCx | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 | 徐 |
| 後 | | U+5FDx | 志 | 志 | 志 | 志 | 志 | 志 | 志 | 志 | 志 | 志 | 志 | 志 | 志 | 志 | 志 |
| 後 | | U+5FEx | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 | 忠 |
| 後 | | U+5FFx | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 | 仲 |

図 2.2 Unicode 文字一覧表で「後」を表示

ただ、絵文字は、この16進数4桁では収まらない領域に配置されているものが非常に多いです。確かに、4桁の16進数が割り当てられている絵文字もあります。例えば、絵文字で示される感嘆符 (!) は、U+2757 に、また赤いハートは U+2764 に割り当てられています。しかし、例えば青いハートは、U+1F499 という領域に割り当てられており、「U+」が5桁に拡張された場所が割り当てられていることがわかります。なお専門的なことには筆者は詳しくないので省略しますので、興味のある読者の皆様はご自分で調べてくださいますよう、何卒よろしくお願ひいたします。

さて、実を言うと、この5桁に拡張された領域でも、絵文字を全て表示することはできません。それだけ絵文字は多数あり、なおかつ必要に応じてどんどん追加されていくのですから、大したものです。例えば、私がツイッターのプロフィールに使っているレインボーフラッグとトランスジェンダーフラッグは、次のような Unicode で示されます（なお私は LGBTQ+ ないしトランスジェンダーの当事者ではないシスジェンダー男性 (he/him) ですが、差別への反対を示すため掲げております）。

レインボーフラッグ: U+1F3F3 U+FE0F U+200D U+1F308
トランスジェンダーフラッグ: U+1F3F3 U+FE0F U+200D U+26A7 U+FE0F

なんと、前者は4個、後者は5個の文字が使われているのです。しかも中にはU+の数字が4桁のものと5桁のものが混在しております。こうなると、RのコンソールやExcelの画面などではまともに表示されないことが多いです。

そもそも、各国の国旗（国連やEU含む）を示す絵文字もまた、2個の文字の合成として表されます。例えば日本なら「U+1F1EF U+1F1F5」、韓国なら「U+1F1F0 U+1F1F7」、アメリカなら「U+1F1FA U+1F1F8」です。しかし、これらの国旗をExcelで表示しようとすると、それぞれ「JP」「KR」「US」としか表示されません（しかも小さい）。従って、Excelではまともに絵文字を扱うことができないのです。

第2章 絵文字を Excel VBA で使うための基礎知識

2.3 絵文字を Excel VBA で扱うために

2.3 絵文字を Excel VBA で扱うために

絵文字を Excel VBA で扱うためには、下記の2段階を踏む必要があります。

1. U+ 以下の16進数を10進数に変換する。
2. 複数の Unicode 文字が使われている場合、それらの文字を合成する。

とはいえ、Excel VBA においては、そのままでは16進数を用いることができません。Excel には、16進数を10進数に変換する関数である HEX2DEC がありますので、それを用います。方法としては、Excel VBA に実装されていないが Excel には実装されている関数を使うための接頭語である「WorksheetFunction」を使って「WorksheetFunction.Hex2Dec」を使うか、あるいは一旦シートに HEX2DEC を用いたセルを書き出して使うかという方法があります。それを基にして Unicode 文字を探索します。

2段階目においては、Excel でも Excel VBA でも、記号「&」を使えば文字を合わせることができる（例えば、あるセルに「="後"&"藤"」という式を入力すると「後藤」と表示されます）ので、方法としては単純です。

2.4 Excel で絵文字を表示してみる

ここでは、コードから Excel で Unicode 文字を表示する方法を試してみたいと思います。

前々節で示したとおり、「後」という文字は、U+5F8C という場所に割り当てられています。この、「U+」以降の文字を抽出するためには、私は MID 関数を使います。MID 関数とは、次のようなものです。

```
=MID(文字列, 開始位置, 文字数)
```

例えば、般若心経の冒頭を使って、適当なセルに「=MID("観自在菩薩行深般若波羅蜜多時",4,2)」と入力すると、「菩薩」という文字列が返されるはずですが、セル C2 に U+5F8C と入力して、次に F2 に次のように入力します。

```
=MID(C2, 3, 100)
```

最後の文字が大きすぎるのは単なる私の趣味です。「U+」の次、すなわち3文字目以降がすべて表示されれば何でもいいのです。そうすると、F2 に「5F8C」と表示されますので、今度はこの16進数で表示された数字を10進数に変換するために、I2 に次のように入力します。

```
=HEX2DEC(F2)
```

これで I2 に 24460 という数字（10進数）が出るはずですが、そして、関数 UNICHAR（形式は「=UNICHAR(10進数の数字)」）を使って、L2 に「=UNICHAR(I2)」と入力すると、