

Using SAS® Enterprise Miner for Categorization of Fitbit's Customer Complaints on Twitter

Sapna Bhoir, Jacky Arora, Dr. Goutam Chakraborty, Oklahoma State University

ABSTRACT

All companies are trying to be more customer-centric by implementing new measures to enhance the consumer experience. One such measure recently implemented by many companies is "Social Media Customer Service". According to J.D. Power, 67% of consumers have used a company's social media support page for customer service issue.^[5] It has been reported that consumers can expect a reply within a couple of minutes from the support team regarding the severity of the issue. Fitbit is one such company, which has grown quite popular in the recent days. It has encouraged its customers to generate a buzz in social media by expressing their reviews, discussing the new product launch and utility of Fitbit in customer's day-to-day life. However, on the flip side, Fitbit's Twitter support page is flooded with issues that consumers are facing while using their products. There is at least one tweet (@FitbitSupport) every minute by a user or by the support team responding to a user's complaint. The primary objective of this research is to categorize these complaints and figure out the major issues such as whether it's related to activity tracking, design, tech specs or application interactivity and so on. Since the tweets are model specific, we will compare if or not the issues are resolved between two generations of the product.

INTRODUCTION

Fitness trackers are getting popular and many of the fitness tracker users rely on Fitbit to track their day to day activities. In the current era of social media, consumers usually express their reviews and feedback of a certain product on different social media channels like Twitter and Facebook. Imagine if we could analyze this data to gain better insights on different products and certain issues primarily related to Fitbit. We can review each product and most common issues related to that product. We can view a summarization of reviews of a certain product. This might turn out to be of great help to the support team as well to know whether or not their new models are able to resolve the preexisting complaints.

DATA PREPARATION

The data preparation for our analysis was done using following steps:

- Extracting the data from Twitter.
- Importing the textual data (.csv file) in the SAS environment to create a SAS dataset.
- Text parsing to convert the unstructured text to spreadsheet (structured) type format for ease of analysis and identify linguistic terms.
- Text filtering to filter out the terms that have little to no information value. This will create more relevant topics in the analysis.

DATA EXTRACTION

The textual data (Fitbit customer comments) was collected from Twitter using web-scraping. This was done using TwitterSearch package in Python 2.7.10 . The web contents of the customer comments are parsed using HTML parser in Python. Irrelevant contents (@FitbitSupport replies to customer complaints) were removed using regular expressions in python.

The data collected was in an unstructured format as follows:

| 1 | username | date | retweets | favorites | text | geo | mentions | hashtags |
|----|------------|-----------------|----------|-----------|--|----------|---------------------|------------|
| 2 | ryanfa12 | 7/10/2016 19:20 | 0 | 8 | Don't buy from @FitbitSupport ... 2 Fitbits broke in 1.5 years, talk about | | @FitbitSupport | |
| 3 | MeierCarc | 7/10/2016 12:56 | 0 | 5 | PSA: fitbits are cheap & I haven't h | Woodbury | MN | @FitbitSup |
| 4 | madeline | 7/10/2016 1:19 | 0 | 9 | My @fitbit fell off for the second time tonight... @FitbitSupport | | @fitbit @FitbitSupp | |
| 5 | GanatraNi | 7/8/2016 12:28 | 2 | 18 | Thankyou @FitbitSupport @FitbitUK for sorting my watch out so quick! | | @FitbitSupport @Fit | |
| 6 | Nelson_K | 7/7/2016 22:33 | 4 | 17 | @FitbitSupport actually I ended up calling and they helped me out #m | | @FitbitSu #mahalo # | |
| 7 | FightingFr | 7/7/2016 14:19 | 4 | 43 | HUGE shoutout to the @fitbit customer service team. THAT is how cust | | @fitbit @ #Fitbit | |
| 8 | DCBuckey | 7/7/2016 13:40 | 0 | 2 | @Monte_Colorman @FitbitSupport Dynamite drop in Monte | | @Monte_Colorman | |
| 9 | Monte_Cc | 7/7/2016 13:39 | 0 | 8 | . @FitbitSupport A coupon off another purchase is worthless. If your pi | | @FitbitSupport | |
| 10 | Monte_Cc | 7/7/2016 13:38 | 1 | 5 | . @FitbitSupport Product should last more than 8 months. I've recomm | | @FitbitSupport | |

Figure 1: Twitter Data Collected (.csv)

The customer comments and username columns were selected from extracted data for text analysis.

| Variable Name | Level | Description |
|---------------|-------|---|
| User | ID | Identifier variable |
| Cust_comments | Text | Actual customer comments posted by the customer on Twitter. |

Table 1: Data Dictionary

| | User | Cust_comments |
|----|--------------|--|
| 1 | @LynnPace711 | @FitbitSupport the bluetooth sync has stopped working on my chargehr. Is there a way for me to fix this?? |
| 2 | @NigelHoney | Hey @asphotos is this what your @fitbit did? @FitbitSupport https://t.co/BLXblkgpUP |
| 3 | @susant36 | @FitbitSupport can't believe my fitbit_charge band snapped not a happy chappy not even 6 months old |
| 4 | @place_s | @FitbitSupport Magic! you did it. Thanks @fitbit. So happy to save that data #13.1 is a lot of miles for me #runhappy! |
| 5 | @Lancair360 | @fitbitsupport How do I return a Flex bought on 3/29 from Amazon for repair? It stopped working 2 days ago.No lights can't reset won't chg |
| 6 | @losavio | My @fitbit crapped out and it's not even worth moving around anymore. @FitbitSupport please help! |
| 7 | @kggonline | @FitbitSupport will do. thanks! so as i understand the bluetooth AND location requirements were implemented by android/google NOT fitbit? |
| 8 | @JJ_Web | @FitbitSupport my ChargeHR's piece over the charge port broke off when removing from charger & won't stay on anymore https://t.co/yDYi75gMvl |
| 9 | @place_s | @FitbitSupport when I contacted you last month abt this restarting app created loss of data. Hesitant to do this. Was told its server issue |
| 10 | @MoniqueA_S | On Wednesday night my chargehr stopped syncing with my phone. I've removed the device and tried to pair it... help? @FitbitSupport @fitbit |

Figure 2: Sample of Data Used for Text Analysis (.csv)

METHODOLOGY

First, the comments were imported from the time span of January 1, 2016 – July 18, 2016. These comments were extracted using multiple search queries and only the first comment of a specific user was extracted with the intent of extracting only the complaint and not the conversation. Fitbit’s reply was also removed as it was mostly just to confirm or for sympathy. Then the data was imported into SAS® Enterprise Miner™ 14.1 and defined below is the process flow.

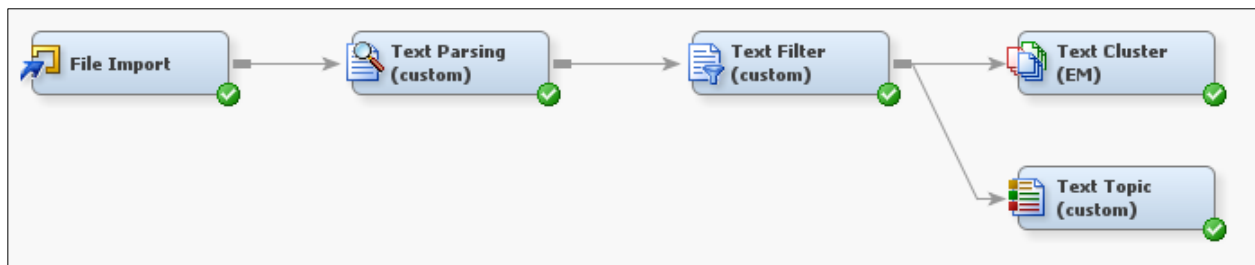


Figure 3: Process Flow

FILE IMPORT

The text data (.csv file) was imported into the SAS environment using File import node in SAS® Enterprise Miner™ 14.1. The delimiter was set to comma. It contains 33,355 customer comments.

| Property | Value |
|---------------------------|-------------------------------|
| General | |
| Node ID | FIMPORT |
| Imported Data | ... |
| Exported Data | ... |
| Notes | ... |
| Train | |
| Variables | ... |
| Import File | H:\OSU\Fitbit_data\tweets.csv |
| Maximum rows to import | 1000000 |
| Maximum columns to import | 10000 |
| Delimiter | , |
| Name Row | Yes |
| Number of rows to skip | 0 |
| Guessing Rows | 500 |
| File Location | Local |
| File Type | csv |
| Advanced Advisor | No |
| Rerun | No |

Figure 4: SAS File Import Node Property Panel Settings Text Parsing

After importing the text data, the text parsing node was attached to it. This node was used with some customized settings in the property panel to convert the unstructured data into a structured format for ease of analysis.

| Property | Value |
|--|--|
| Train | |
| Variables | ... |
| <input checked="" type="checkbox"/> Parse | |
| Parse Variable | Cust_comments |
| Language | English |
| <input checked="" type="checkbox"/> Detect | |
| Different Parts of Speech | No |
| Noun Groups | No |
| Multi-word Terms | SASHELP.ENG_MULTI |
| Find Entities | Standard |
| Custom Entities | |
| <input checked="" type="checkbox"/> Ignore | |
| Ignore Parts of Speech | Abbr' 'Aux' 'Conj' 'Det' 'Interj' 'Num' 'Par'... |
| Ignore Types of Entities | 'Address' 'Currency' 'Date' 'Internet' 'Lo'... |
| Ignore Types of Attributes | Num' Punct' |
| <input checked="" type="checkbox"/> Synonyms | |
| Stem Terms | Yes |
| Synonyms | SASHELP.ENGSYNMS |
| <input checked="" type="checkbox"/> Filter | |
| Start List | ... |
| Stop List | SASHELP.ENGSTOP |
| Select Languages | English |

Figure 5: SAS Text Parsing Node Property Panel Settings

The following properties were altered in the properties panel of the Text Parsing node:

- Detect “Different Parts of Speech” was set to “No”. This eliminated the repetitive terms with different parts of speech.
- “Find Entities” was set to “Standard”.
- Abbr (e.g., i.e., wiz.), prop (by, with, about, until) and num (2, 4, 8, 6) parts of speech are ignored. Mainly because of date variable in our data, so it was full of numbers.
- Address, currency, date, internet, location, measure, percent, person, phone, product, prop_misc, ssn, time, time_period, title, vehicle types of entities were ignored.
- Num and punctuation types of attributes were ignored.

The text parsing node generated the terms by document matrix, which helped to identify the most frequently occurring terms along with the number of documents it occurred in. It also provided the terms that rarely appear. Ideally, the most important terms for analysis are the moderately used terms.

| Term | Role | Attribute | Freq | # Docs | Keep | Parent/Child Status | Parent ID | Rank for Variable numdocs |
|---------------|------|-----------|------|--------|------|---------------------|-----------|---------------------------|
| not | ... | Alpha | 4028 | 3439N | | | 8985 | 1 |
| + be | ... | Alpha | 3892 | 3371N | | + | 8970 | 2 |
| + have | ... | Alpha | 3076 | 2625N | | + | 8958 | 3 |
| + do | ... | Alpha | 2616 | 2350N | | + | 8983 | 4 |
| fitbit | ... | Alpha | 2365 | 2234Y | | | 223 | 5 |
| + get | ... | Alpha | 1511 | 1438N | | + | 8984 | 6 |
| s | ... | Alpha | 1409 | 1316N | | | 8908 | 7 |
| + help | ... | Alpha | 1222 | 1205Y | | + | 2579 | 8 |
| chargehr | ... | Alpha | 1181 | 1171Y | | | 4446 | 9 |
| + sync | ... | Alpha | 1264 | 1154Y | | + | 3274 | 10 |
| + charge | ... | Alpha | 1258 | 1108Y | | + | 798 | 11 |
| + try | ... | Alpha | 1060 | 1012N | | + | 8918 | 12 |
| + work | ... | Alpha | 1067 | 1004Y | | + | 2816 | 13 |
| + thank | ... | Alpha | 935 | 924N | | + | 9068 | 14 |
| just | ... | Alpha | 932 | 909N | | | 9005 | 15 |
| now | ... | Alpha | 823 | 808N | | | 8987 | 16 |
| + no | ... | Alpha | 837 | 783N | | + | 9054 | 17 |
| + step | ... | Alpha | 839 | 766Y | | + | 4831 | 18 |
| + new | ... | Alpha | 729 | 711N | | + | 8907 | 19 |
| + day | ... | Alpha | 761 | 707Y | | + | 102 | 20 |
| + time | ... | Alpha | 729 | 675Y | | + | 559 | 21 |
| + customer | ... | Alpha | 638 | 611Y | | + | 4473 | 22 |
| + service | ... | Alpha | 582 | 567Y | | + | 3577 | 23 |
| still | ... | Alpha | 571 | 563N | | | 9009 | 24 |
| + replacem... | ... | Alpha | 571 | 557Y | | + | 2010 | 25 |
| + go | ... | Alpha | 544 | 531N | | + | 9001 | 26 |
| + month | ... | Alpha | 502 | 496Y | | + | 3090 | 27 |
| + issue | ... | Alpha | 475 | 462Y | | + | 3145 | 28 |
| + update | ... | Alpha | 486 | 455Y | | + | 3896 | 29 |
| + band | ... | Alpha | 480 | 448Y | | + | 6133 | 30 |
| + say | ... | Alpha | 452 | 424N | | + | 8990 | 31 |
| + much | ... | Alpha | 428 | 421N | | + | 9013 | 32 |
| + support | ... | Alpha | 430 | 420Y | | + | 2909 | 33 |
| battery | ... | Alpha | 444 | 418Y | | | 5007 | 34 |
| + break | ... | Alpha | 431 | 418Y | | + | 203 | 34 |

Figure 6: Text Parsing Output

The most frequently occurring terms were chargehr, sync, step, customer service, replacement, issue, update, band, battery, support, break etc. which makes sense, as we were analyzing Fitbit customer comments.

TEXT FILTER

After text parsing, the Text Filter node was added to reduce the number of terms used in the documents.

| Train | |
|-----------------------------|---------------------------|
| Variables | |
| Spelling | |
| Check Spelling | Yes |
| Dictionary | EMWS1.ENGDICT |
| Weightings | |
| Frequency Weighting | Default |
| Term Weight | Default |
| Term Filters | |
| Minimum Number of Documents | 15 |
| Maximum Number of Terms | |
| Import Synonyms | |
| Document Filters | |
| Search Expression | |
| Subset Documents | |
| Results | |
| Filter Viewer | |
| Spell-Checking Results | EMWS1.TextFilter3_spellDS |
| Exported Synonyms | TEMP.Cust_Syn_2 |
| Report | |
| Terms to View | Selected |
| Number of Terms to Display | 20000 |

Figure 7: SAS Text Filter Node Property Panel Settings

The following properties were altered in the properties panel of Text Parsing node:

- “Check Spelling” was set to “Yes”.
- English dictionary was used to identify and correct the spell check errors.
- “Minimum Number of Documents” was set to 15. This eliminated the terms that occurred less than 15 times in all the documents.

The “Check Spelling” option corrected wrong spellings of words.

- For example, “uninstalling” was corrected to “uninstalling”.

| | Parent # Docs | Term | # Docs | Parent | Role | Parent Role | Min Distance | Dictionary | Key | Parent ID |
|----|---------------|--------------|--------|--------------|------|-------------|--------------|------------|--------|-----------|
| 46 | 5.0 | uninstalling | 1.0 | uninstalling | | | 4.0 | N | 1358.0 | 1813.0 |
| 47 | 23.0 | steep | 1.0 | step | | | 12.0 | Y | 2461.0 | 1827.0 |
| 48 | 4.0 | router | 2.0 | route | | | 12.0 | Y | 1089.0 | 1837.0 |
| 49 | 15.0 | experiencia | 1.0 | experience | | | 13.0 | N | 1216.0 | 1904.0 |
| 50 | 15.0 | bluetooth | 1.0 | bluetooth | | | 6.0 | N | 1099.0 | 1934.0 |
| 51 | 29.0 | quik | 1.0 | quick | | | 12.0 | N | 1387.0 | 2107.0 |
| 52 | 19.0 | actives | 1.0 | active | | | 10.0 | Y | 277.0 | 2187.0 |
| 53 | 97.0 | emails | 6.0 | email | | | 6.0 | N | 2514.0 | 2352.0 |
| 54 | 97.0 | e-mail | 2.0 | email | | | 10.0 | N | 168.0 | 2352.0 |

Figure 8: Text Filtering Spell Check

Interactive Filter Viewer was used to create a smaller set of custom synonyms and drop some irrelevant terms. For example, terms like “Fitbithr”, “hr” and “hrcharge” were treated as synonyms and grouped together as “chargehr”.

| term | parent |
|------------|-----------|
| fitbithr | chargehr |
| hr | chargehr |
| hrcharge | chargehr |
| charger | charger |
| chargers | chargers |
| charge | charging |
| chargeable | charging |
| charged | charging |
| charges | charging |
| charging | charging |
| checked | check |
| half | half-bar |
| hanging | hang |
| hant | hang |
| lag | hang |
| lagging | hang |
| hardest | hard |
| hardly | hard |
| hardtail | hard |
| heal | heartbeat |
| heart | heartbeat |
| heart-rate | heartbeat |

Figure 9: Exported Custom Synonyms SAS Dataset

The text filter output indicated the terms that were kept for text analysis. As we can see in Figure 10, these terms were kept as they were related to Fitbit in some way or the other.

| Term | Role | Attribute | Status | Weight | Imported Frequency | Freq | Number of Imported Documents | # Docs |
|---------------|------|-----------|--------|--------|--------------------|------|------------------------------|--------|
| + chargehr | ... | Alpha | Keep | 0.218 | 3251 | 3328 | 3228 | 3304 |
| + sync | ... | Alpha | Keep | 0.221 | 3581 | 3596 | 3275 | 3287 |
| + replacement | ... | Alpha | Keep | 0.242 | 1665 | 2673 | 1635 | 2599 |
| + charge | ... | Alpha | Keep | 0.246 | 3535 | 2814 | 3153 | 2546 |
| + step | ... | Alpha | Keep | 0.263 | 2350 | 2356 | 2134 | 2140 |
| + strap | ... | Alpha | Keep | 0.269 | 573 | 2130 | 539 | 1980 |
| + cust | ... | Alpha | Keep | 0.273 | 39 | 2010 | 39 | 1898 |
| + service | ... | Alpha | Keep | 0.287 | 1686 | 1687 | 1630 | 1631 |
| + update | ... | Alpha | Keep | 0.311 | 1408 | 1410 | 1294 | 1296 |
| + last | ... | Alpha | Keep | 0.315 | 1277 | 1277 | 1226 | 1226 |
| battery | ... | Alpha | Keep | 0.318 | 1263 | 1263 | 1193 | 1193 |
| + back | ... | Alpha | Keep | 0.318 | 1212 | 1214 | 1180 | 1182 |
| + good | ... | Alpha | Keep | 0.325 | 1136 | 1137 | 1094 | 1095 |
| + rest | ... | Alpha | Keep | 0.329 | 63 | 1102 | 60 | 1059 |
| + device | ... | Alpha | Keep | 0.331 | 1084 | 1086 | 1037 | 1039 |
| + today | ... | Alpha | Keep | 0.332 | 1022 | 1026 | 1012 | 1016 |
| + buy | ... | Alpha | Keep | 0.333 | 1018 | 1023 | 1000 | 1005 |
| + problem | ... | Alpha | Keep | 0.336 | 977 | 992 | 958 | 973 |
| + show | ... | Alpha | Keep | 0.342 | 988 | 991 | 927 | 930 |
| + flex | ... | Alpha | Keep | 0.341 | 946 | 946 | 924 | 924 |

Figure 10: Text Filter Output

CONCEPT LINKS

Using Interactive Filter View from the Text Filter Node properties panel, we observed the concept links indicating the strength of association of a few terms. We mostly concentrated on the different models of Fitbit to examine the most common issues associated with them.

The term in the center was linked by its associated terms. The thickness of line defines the strength of association, with a wider line indicating a stronger association between those terms. Each link node can also be expanded to view its sub linked nodes to have a better idea of the relationship between the two major nodes.

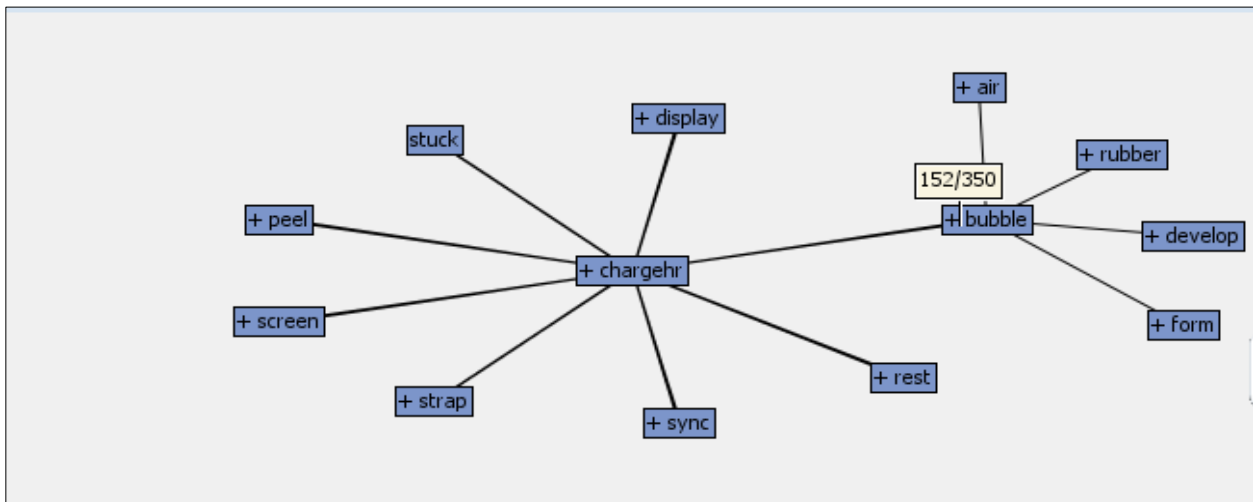


Figure 11: Concept Link for "Charge HR"

The concept link in Figure 11 was for the Charge HR. The other terms (issues) related to this model were +strap (mostly concerned with break issues), +bubble (air bubble developing in the rubber band), +peel (Rubber strap peeling off), and stuck (band stuck on the loading screen).

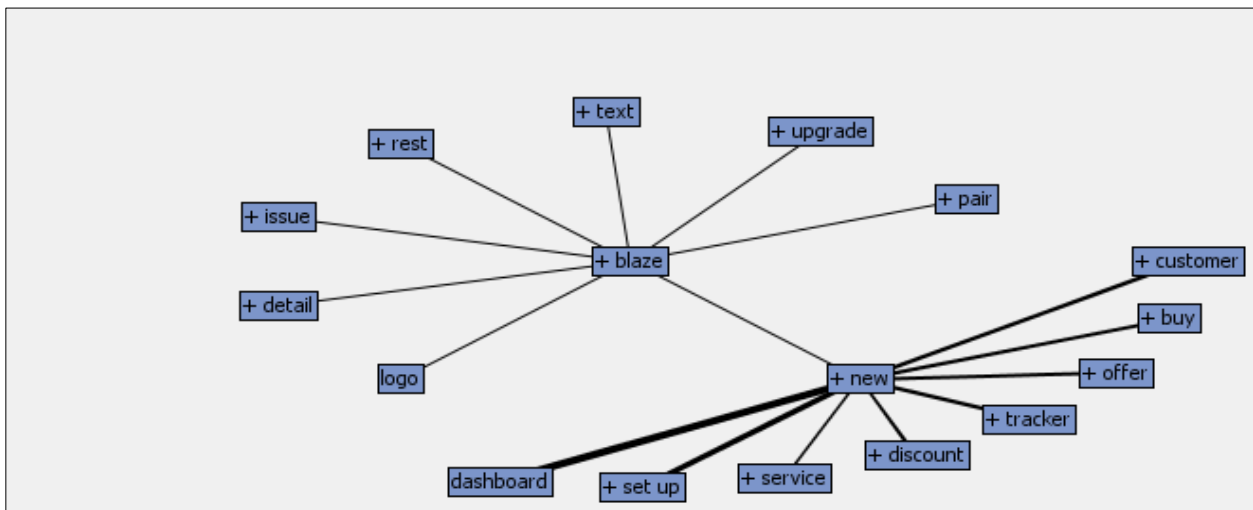


Figure 12: Concept Link for "Fitbit Blaze"

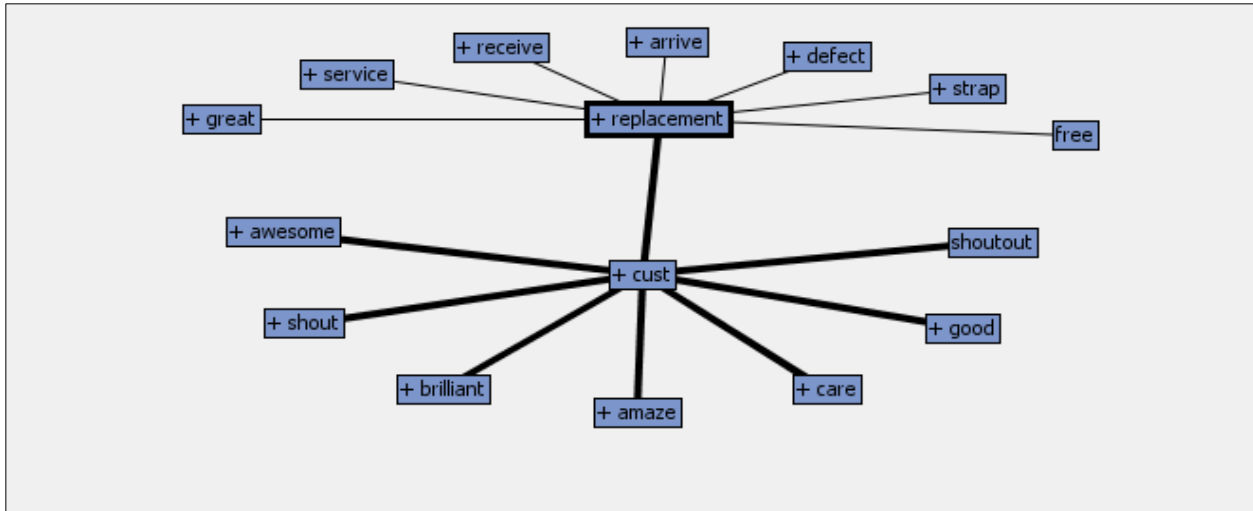


Figure 15: Concept Link for “Fitbit Support Customer Service”

Figure 15 shows the terms (feedback) related to Fitbit Support customer service, which comprised of mixed feelings from customers. Positive feedback was indicated in terms such as awesome, brilliant, and amazing. Negative feedback was indicated in terms such as short and bad. Terms related to replacement arrived, defect, free, great, and strap.

TEXT CLUSTERING

After filtering irrelevant terms and grouping similar terms using the Interactive Filter Viewer in Text Filter node, the Text Cluster node was used to group similar comments based on their terms and categorized those comments. The clustering algorithm used was Expectation-Maximization, as the clusters observed in starting were random and some words occurrences were not predictable or irrelevant to the context. Most suitable settings were “Number of Clusters: 5” and “Number of Descriptive Terms: 10” as the five cluster solution seemed to be well separated from each other.

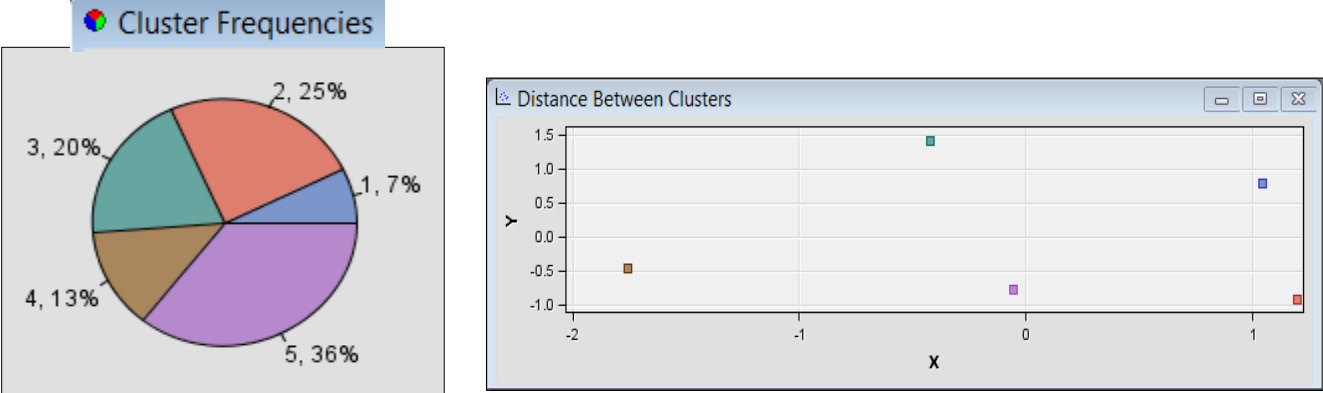


Figure 16: Text Clustering Node Output

The descriptive terms in the five cluster solution and their meaningful terms were as follows:

| Cluster ID | Descriptive Terms | Percentage | Meaningful Category |
|------------|---|------------|--|
| 1 | +work +sync +update +restart +connect +stuck +device +step +phone +reset | 7% | Sync and Update Issues |
| 2 | +blaze +logo +loose-fitting +new +issue +buy +screen +tracker +activity +exercise | 25% | Fitibit Blaze related issues and purchase advice |
| 3 | +good +cust service +great +know +problem +love +awesome +amazing +excellent +care | 20% | Feedback on customer service |
| 4 | +charging +battery +die +full +low +dead +light +plug +drained +port | 13% | Battery Issues |
| 5 | +chargehr +strap +bubble +broke +replacement +cust service +battery +charging +update | 36% | Charge HR related issues |

Table 2: Descriptive Clusters Associated with Each Cluster and Documents Frequency

TEXT TOPIC

After connecting the Text Filter node in SAS® Enterprise Miner™, the Text Topic node is attached. It enabled us to combine the terms into relevant topics for further analysis.

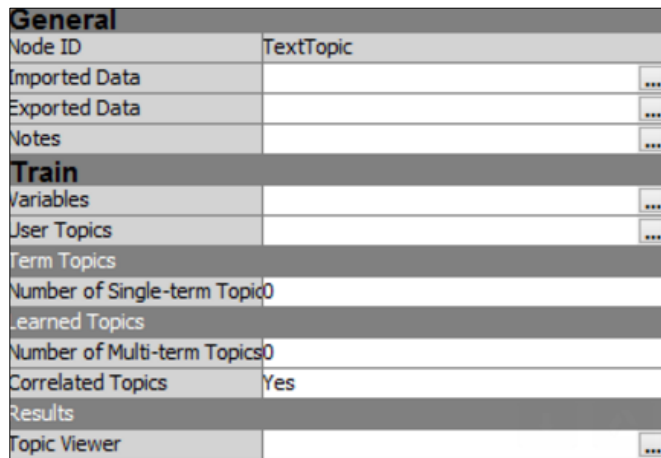


Figure 17: SAS Text Topic Node Property Panel Settings

The following properties were altered in the properties panel of Text Parsing node:

- “Number of Multi-term topics” was set to “0”.
- “Correlated Topics” was set to Yes.
- As we were interested in comparing various Fitbit models, we created user topics. These user topics were defined after analyzing the results of text cluster node and concept links obtained in the text filter node.

| Topic Id | Topic Terms | Explanation |
|----------|--|----------------------|
| 1 | +pink +setup +update +phone +restart +sync +push +alta | Alta_Issues |
| 2 | +logo +upgrade +pair +text +issue +detail +new +loose-fitting +minute +blaze | Blaze_Issues |
| 3 | +bubble +strap +break +old +replacement +air +progressbar +update +fail +warranty +half-bar +chargehr +display +green +flash | ChargeHR_Issues |
| 4 | +cust service +good +great +awesome +excellent +step +count +happy +care +amazing | Customer_Service |
| 5 | +strap +Fitbit_charge +portion +battery +month +die +side +less +rubber +cheap +strong +disappoint +button +display +face | Fitbit_charge_Issues |
| 6 | +rubber +bulge +wrist +break +split +main +develop +peel +strap | Strap_Issue |

Table 3: User Topics

Because we wanted to focus on the various Fitbit models, the weight of a few terms was changed. The screen shot below indicates few of those terms along with their relevant topics and weight.

| Topic | Term | Role | Weight |
|-----------------|-------------|------|--------|
| ChargeHR_Issues | Bubble | | 1 |
| ChargeHR_Issues | Strap | | 1 |
| ChargeHR_Issues | break | | 1 |
| ChargeHR_Issues | old | | 1 |
| ChargeHR_Issues | replacement | | 1 |
| ChargeHR_Issues | air | | 1 |
| ChargeHR_Issues | progressbar | | 1 |
| ChargeHR_Issues | update | | 1 |
| ChargeHR_Issues | Fail | | 1 |
| ChargeHR_Issues | warranty | | 1 |
| ChargeHR_Issues | Half-bar | | 1 |
| ChargeHR_Issues | chargehr | | 1 |
| ChargeHR_Issues | display | | 1 |
| ChargeHR_Issues | green | | 1 |
| ChargeHR_Issues | flash | | 1 |
| Blaze_Issues | logo | | 1.7 |
| Blaze_Issues | upgrade | | 1 |
| Blaze_Issues | pair | | 1 |
| Blaze_Issues | text | | 1 |

Figure 18: User Topics with Terms and Weights for Text Topic Node

After running, Text Topic node, we obtained the following results:

| Topic ID | Category | Topic | # Docs |
|----------|----------|----------------------|--------|
| 1 | User | Alta_Issues | 4396 |
| 2 | User | Blaze_Issues | 1919 |
| 3 | User | ChargeHR_Issues | 9564 |
| 4 | User | Customer_Service | 5657 |
| 5 | User | Fitbit_charge_Issues | 5782 |
| 6 | User | Strap_Issue | 2593 |

Figure 19: Text Topic Results

In the Text Topic Viewer, we were able to see the customer comments related to the assigned user-defined topics. The output below displays the customer comments related to topic “ChargeHR_Issues”

| Topic | Category | Term Cutoff | Document Cutoff | Number of Terms | # Docs |
|----------------------|----------|-------------|-----------------|-----------------|--------|
| Alta_Issues | User | 0.001 | 0.001 | 4 | 4396 |
| Blaze_Issues | User | 0.001 | 0.001 | 7 | 1919 |
| ChargeHR_Issues | User | 0.001 | 0.001 | 14 | 9564 |
| Customer_Service | User | 0.001 | 0.001 | 8 | 5657 |
| Fitbit_charge_Issues | User | 0.001 | 0.001 | 10 | 5782 |
| Strap_Issue | User | 0.001 | 0.001 | 8 | 2593 |

| Topic Weight | + | Term | Role | # Docs | Freq |
|--------------|---|-------------|------|--------|------|
| 1 | + | chargehr | | 3487 | 3536 |
| 1 | + | replacement | | 2599 | 2673 |
| 1 | + | strap | | 2041 | 2196 |
| 1 | + | update | | 1296 | 1410 |
| 1 | + | old | | 791 | 795 |
| 1 | + | warranty | | 540 | 552 |
| 1 | + | display | | 466 | 480 |

| Topic Weight | Cust_comments |
|--------------|---|
| 2.366 | Hey @FitbitSupport my chargehr is just under 6 months old and an air bubble has formed on the strap just below the display. #nothappy |
| 2.192 | @FitbitSupport Hello My chargeHR is less than a year old & band started peeling around display & air bubble. Help |
| 1.909 | @FitbitSupport Hi I've noticed an air bubble forming under the band of my HR. How do I go about getting a replacement? It's under warrantee |
| 1.895 | @FitbitSupport Hi my chargehr has an air bubble in the band. Is this covered under the warranty? pic.twitter.com/GS9gRW9ok |
| 1.817 | @FitbitSupport Brand new chargehr display doesn't work. Tried charging and restarting and nothing shows on screen but green light flashes. |
| 1.809 | @FitbitSupport hi! My Fitbit ChgrgHR (bought in August15) has an air bubble under the band. Is this under warranty? pic.twitter.com/BI2GhP6oy |
| 1.738 | @FitbitSupport what's the warranty on my Fitbit chargehr? I got it in June and the band is bubbling. pic.twitter.com/0DHmhhwRv0 |
| 1.704 | @FitbitSupport Hi. A bubble is appearing on my chargehr (2 months old). How would I go about getting it replaced? pic.twitter.com/wRfomYe18 |

Figure 20: Topic Viewer Output

RULE-BASED MODEL FOR CATEGORIZATION

After creation of clusters, a sample of 18,000 customer comments was generated from the result of clustered data with each document assigned to a specific category.

Cust_comments variable was set as “Text” and TextCluster_cluster_ as “Target”. Splitting criteria were set to 70% for training and 30% for validation. All the properties of Text Parsing and Text Filter node were exactly the same as they were used for text clustering.

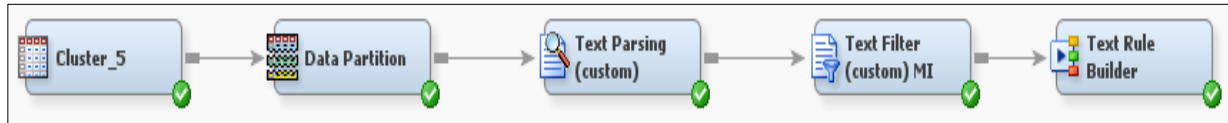


Figure 21: Modeling Diagram for Generating Text Rules for Categorization

We preferred using selection criteria as the misclassification rate and the Text Rule Builder node was run switching between three modes: Low, Medium and High setting for generalization error, purity of rules and exhaustiveness.

Minimum misclassification rate for the validation data was found to be 27.3% when keeping these settings to low. This indicated the rule-based model was able to categorize around 73% of the actual data correctly.

| Fit Statistics | | | | | |
|----------------|----------------|---------------------------------|-------|------------|----------|
| Target | Fit Statistics | Statistics Label | Train | Validation | |
| category | _ASE_ | Average Squared Error | | 0.036008 | 0.037292 |
| category | _DIV_ | Divisor for ASE | | 116720 | 50055 |
| category | _MAX_ | Maximum Absolute Error | | 0.814634 | 0.793836 |
| category | _NOBS_ | Sum of Frequencies | | 23344 | 10011 |
| category | _RASE_ | Root Average Squared Error | | 0.189758 | 0.193112 |
| category | _SSE_ | Sum of Squared Errors | | 4202.866 | 1866.665 |
| category | _DISF_ | Frequency of Classified Cases | | 23344 | 10011 |
| category | _MISC_ | Misclassification Rate | | 0.246402 | 0.273399 |
| category | _WRONG_ | Number of Wrong Classifications | | 5752 | 2737 |

Figure 22: Fit Statistics for Text Rule Builder Node

To understand what terms were used to categorize the reviews into the various categories, we looked at the rules that governed them.

| Rules Obtained | | | | | | |
|----------------|--------|--|----------------------|-----------|----------------------------|-----------------|
| Target Value | Rule # | Rule | True Positive/ Total | Precision | Valid True Positive/ Total | Valid Precision |
| 5 | 45 | old & ~sync & ~last & ~customer & ~good & ~charging & ~back & ~bubble & ~update & ~battery & ~replacement & ~device & ~brand | 212/215 | 98.48% | 111/121 | 94.75% |
| 5 | 46 | log & ~step & ~sync & ~update & ~back & ~show & ~customer | 172/177 | 98.55% | 79/83 | 95.02% |
| 5 | 47 | turned on & ~back & ~sync & ~customer & ~battery & ~step & ~update & ~rest & ~charging | 155/159 | 98.47% | 62/63 | 95.66% |
| 5 | 48 | screen & ~rest & ~show & ~chargehr & ~update & ~step & ~device & ~back & ~battery & ~brand | 216/225 | 98.39% | 93/97 | 95.95% |

Figure 23: Classification Rules Obtained

In Figure 23, the rules to determine the ChargeHR related issues (target value: 5) contained terms like chargehr, bubble, battery, replacement, sync, update, charging etc. with a precision of 98.48%.

We scored the data with 2,500 observations to check if observations were correctly classified in the corresponding categories using score node in SAS® Enterprise Miner™.

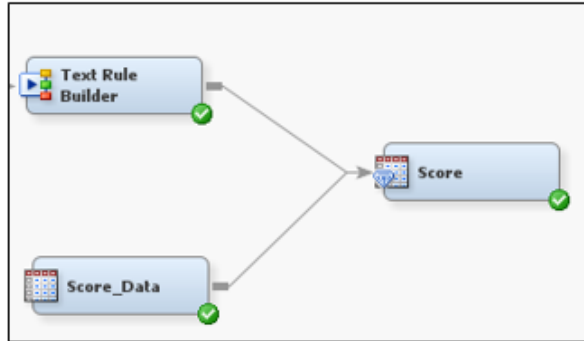


Figure 24: Modeling Diagram to Score New Data

| Cluster ID | Meaningful Category |
|------------|--|
| 1 | Sync and Update Issues |
| 2 | Fitibit Blaze related issues and purchase advice |
| 3 | Feedback on customer service |
| 4 | Battery Issues |
| 5 | Charge HR related issues |

Table 4: Cluster Categories

Figure 25 indicates the output after running the “Score” node.

| EMWS1.Score2_SCORE | | |
|--------------------|--|----------------|
| _document_ | Cust_comments | Into: category |
| 1.0 | @FitbitSupport my chargehr band is bubbling it's 6 months old. How can I get repair pls? | 5 |
| 2.0 | @FitbitSupport HR not syncing. Tried the why my fbit not working page. No luck. What a waste of money months of not syncing despite efforts. | 2 |
| 3.0 | super impressed by @FitbitSupport ... fantastic customer service! | 3 |
| 4.0 | @FitbitSupport yes that's what I have done. I get the Fitbit logo then nothing | 2 |
| 5.0 | @FitbitSupport could you please help I got a Fitbit chargehr at Christmas and is falling apart already! I love it but it's coming apart. | 5 |
| 6.0 | @FitbitSupport - Have to go to + sign now to track excercise then log.Cycling not offered as choice in search - Why not ? | 5 |
| 7.0 | @FitbitSupport After the last update the app stopped pulling all my data from my device. If I skip a day of uploading it will say 0 steps. | 2 |
| 8.0 | @FitbitSupport when I plug Alta in says it's 3/4 charged but it's working unless plugged in help please! Charged it completely yesterday. | 4 |
| 9.0 | @FitbitSupport Many thanks. | 2 |
| 10.0 | Hi @FitbitSupport I have a chargehr and noticed today that a bubble has formed on the strap. Can you help? :(| 3 |

Figure 25: Scored Data Output

CONCLUSION

With the validation accuracy of around 73%, the text model performance was fairly reasonable given the unstructured format of the data in the real world. Customer comments play an important role in giving a fair idea of the issues that customers face. For example, many customers complained about the Fitbit model ChargeHR having issues such as air bubbles in the strap or the strap peeling. Consumers also seem to be highly satisfied with the Fitbit customer service.

This research can help people who want value for their money to decide which Fitbit model they want to purchase. On the flip side, this could also be useful for Fitbit Inc. to identify issues in specific models and resolve them. This study can also help them to track how their newly released products are performing and what are customer review on the new products. Twitter is a good source to get customer insights.

REFERENCES

1. Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS® by Goutam Chakraborty, Murali Pagolu, Satish Garla.
2. SAS Institute Inc. 2014. Getting Started with SAS® Text Miner™ 13.2. Cary, NC: SAS Institute Inc.
3. <https://dev.twitter.com/rest/reference/get/search/tweets>
4. <http://twittersearch.readthedocs.io/en/v0.78.2/TwitterSearchOrder.html#advanced-usage>
5. <http://www.jdpower.com/press-releases/2013-social-media-benchmark-study>

ACKNOWLEDGEMENT

We thank SCSUG 2016 conference committee for giving us an opportunity to present our work. We also thank Dr. Goutam Chakraborty for his continuous support and guidance.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Sapna Bhoir

Oklahoma State University

Email: sapna.bhoir@okstate.edu

Sapna Bhoir is a graduate student at Oklahoma State University's Spears School of Business, pursuing a Master's degree in Business Analytics. She has a bachelor's degree in Electronics Engineering from Mumbai University. She has five years of work experience as Senior System Engineer for Infosys Limited, India. She is currently working as Marketing Analyst Intern with OGE Energy Corp., Oklahoma City since May 2016. She is a SAS® Certified Base Programmer, SAS® Certified Advanced Programmer and SAS® Certified Statistical Business Analyst.

Jacky Arora

Oklahoma State University

Email: jacky.arora@okstate.edu

Jacky Arora is a graduate student currently pursuing Masters in Business Analytics at the Spears School of Business, Oklahoma State University. He is currently working as a Marketing Analyst Intern with the Plymouth Rock Assurance, NJ, since May 2016. He is advanced SAS® Certified programmer and SAS® Certified Statistical Business Analyst.

Dr. Goutam Chakraborty

Oklahoma State University

Email: goutam.chakraborty@okstate.edu

Dr. Goutam Chakraborty is Ralph A. and Peggy A. Brenneman professor of marketing and founder of SAS and OSU data mining certificate and SAS and OSU marketing analytics certificate at Oklahoma State University. He has published many journals such as Journal of Interactive Marketing, Journal of Advertising Research, Journal of Advertising, Journal of Business Research, etc. He has over 25 Years of experience in using SAS® for data analysis. He is also a Business Knowledge Series instructor for SAS®.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.