

Mining Consumer Complaints*

Extended Abstract[†]

Cary Sullivan

Ryan Wills

University of Colorado Boulder

University of Colorado Boulder

Cary.sullivan@colorado.edu

Ryan.wills@colorado.edu

PROBLEM STATEMENT / MOTIVATION

In 2016 alone, The Federal Trade Commission reported 3 million accounts of credit card fraud or identity theft related complaints. If companies can find trends in data and narrow down fraudulent activity, they will be able to find solutions as to why common areas may be targeting users. We strive to find complaints about fraudulent credit or debit card use and link these issues to specific Card Companies as well as finding trends in common zip codes. Fraud most commonly occurs online or over the phone with credit card information stolen.

In 2016, 15.4 million Americans were affected by credit card fraud (Weisbaum, nbcnews.com). The number significantly increases every year and will continue to do so as the card robbers have better technology and strategies at their fingertips. This collection of data will hopefully shed some light on what the common factors are.

If trends can be caught in certain areas and logged, data mining can help make conclusions about where threats are posed and how to avoid them. Certain stores, zip codes, card companies, are all evidence that can be tied together and analyzed for results. We hope to find patterns based on the company used, credit card issue, and area that will help consumers

and companies alike understand the types of credit card crimes that are common.

REFERENCE:

Steele, J. 2017. Credit card fraud and ID theft statistics, digital. <https://www.creditcards.com/credit-card-news/credit-card-security-id-theft-fraud-statistics-1276.php>

Weisbaum, Herb. Identity Fraud Hits Record Number of Americans in 2016, digital. <https://www.nbcnews.com/business/consumer/identity-fraud-hits-record-number-americans-2016-n715756>

1 LITERATURE SURVEY

Data in Action - Combatting Fraud: One company uses big-data analytics to find grey charges on user's credit cards and debit cards by drawing upon billing dispute data from the web, banks, and the CFPB's open consumer complaint database. 'Grey' charges are defined as lingering charges that a user previously signed up for a subscription or renewal service and may not be aware of the charge on their credit or debit card. While grey charges are not illegal, the user may not remember or have completely understood the terms presented and the charges can be misleading. Data.gov provides the highlight of a non-specific

company combating grey charges using our same database, but the specifics of their analysis have not been released yet. Therefore, their research and techniques will not influence our Data Mining much, but it is important to note that work in this field is being conducted.



Figure 1: The Economic Impact of Grey Charges on Debit and Credit Card Issuers. 2012. Found at <https://aitegroup.com/report/economic-impact-grey-charges-debit-and-credit-card-issuers>

Figure 1 clearly shows the Volume of Grey charges occurring in a single year – the front runner being free-to-paid subscriptions customers likely forget about. There is a fine line between fraud and grey charges and the described study works to understand the difference.

2 PROPOSED WORK

2.1 What Needed for Data Collection

It was fairly easy to collect our data. The Financial Services Consumer Complaint Database has been collecting this data for a long time and had a large data set that we were able to access and download. Since the data came from a .gov website, it was organized and formatted well and had plenty of data points. We also knew that since the data was collected by a government, then it was likely highly reputable and reliable. However, since it was such a large database, we knew that it would be more susceptible to error and empty data points, and would therefore

not be perfect and would need cleaned up, a prediction which proved to be true.

2.2 Preprocessing

In order for our data to be usable, it will be necessary to clean up the large number of data points we have. Our data was organized well and due to it being collected from a good source, was formatted nicely. Inconsistent quotations will need to be synchronized and made consistent throughout the data set. There are many cases of empty cells, inconsistent formatting on cells, and partial zip codes. By getting rid of incomplete entries or making them zeros or empty strings, we should be able to make our data consistent, considering the data is in an excel type format. There are also several attributes that don't serve our purpose for the intent of this project that will also be removed.

2.3 How Mining Consumer Complaints is Different from Previous Work

For what we would like to do, we will not be looking at the types of charges that were made on the credit cards. We will be looking at the complaint, how the complaint was handled, the credit card company used, etc., without going into depth on purchases made with the credit cards. We wanted to avoid the purchases and financial part of the data, and focus on the customer service response from the customer complaint.

3 DATA SET

We are using the Financial Services Consumer Complaint Database found at: <https://www.data.gov/consumer/>. Currently downloaded on Cary's Mac (469 MB). Within the data set there are 18 attributes to include: Date received, Product, Sub-product, Issue, Sub-issue, Consumer complaint narrative, Company public response, Company, State ZIP code, Tags, Consumer consent provided, Submitted via, Date sent to company, Company response to consumer, Timely response, Consumer disputed, and Complaint ID. These attributes and data accumulate to over 17 million data points.

4 EVALUATION METHODS

Our results will be evaluated with cross validation. If we are able to use WEKA, there are built in evaluation tools to streamline the process. If we use Python, there are supporting libraries and methods to do the same.

5 TOOLS

We will be using Python and tools and libraries within such as matplotlib, pandas, numpy, etc. Also, we hope to get familiar with WEKA in order to take advantage of its evaluation tools.

6 MILESTONES

6.1 Milestones Completed

- All data cleaned up and groomed – 19 March.

Cleaning took a very long time with millions of data points. We decided to fill the empty cells with a null value since deleting said columns or averaging would skew the data too much. Below in the "Results so far" category is a side-by-side comparison of what the empty cells look like with an "empty" filling them.

- Simple scatter plots and data correlation – 2 April.

We are behind on this milestone due to the unexpected effort for cleaning, but currently working on it.

6.2 Milestones to do

- Using methods such as clustering and sequential patterns based on previous findings – 13 April.
- Analysis of all results – 19 April.
- Refactoring – 23 April.
- Final analysis and conclusion – 26 April.

RESULTS SO FAR:

We've done a lot of data cleaning, filling empty cells with "empty" to allow for uniformity. This will

make it easier once we finish out analysis and apply our code to the data.

ORIGINAL:

Date received	Product	Sub-product	Issue	Sub-issue	Consumer comment	Company response	Comments
3/12/14	Mortgage	Other mortgage	Loan modification, collection, foreclosure				M&M
10/1/16	Credit reporting		Incorrect information	Account status	I have outstanding	Company has	TRAC
10/17/16	Consumer Lc	Vehicle loan	Managing the loan or lease		I purchased a new car on		CITI
6/8/14	Credit card		Bankruptcy				AMER
9/13/14	Debt collection	Credit card	Communication	Frequent or repeated calls			CITI
11/13/13	Mortgage	Conventional	Loan servicing, payments, escrow account				U.S.
6/15/15	Credit reporting		Credit report	Inadequate	on my	Company has	Exper
11/13/15	Mortgage	Other mortgage	Loan modification, collection, foreclosure			Company has	Aldre
10/21/14	Mortgage	Conventional	Loan modification, collection, foreclosure				OCV
4/14/17	Mortgage	Other mortgage	Loan modification, collection, foreclosure			Company has	Shel
3/30/12	Student loan	Non-federal	Repaying your loan				Stud
2/3/16	Debt collection	Other (i.e. private)	Disclosure violation	Not given	en	This company refuses to p	The
1/7/15	Credit reporting		Incorrect information	Account status			Exper
3/15/13	Credit card		Closing/Cancelling account				FIRS
7/18/16	Credit reporting		Incorrect information	Account status			EQU
2/17/16	Debt collection	Credit card	Improper collection	Talked to a tl	This	Company has	SQU
11/7/14	Mortgage	Conventional	Loan modification, collection, foreclosure				U.S.
4/17/15	Mortgage	FHA mortgage	Application, originator, mortgage broker				WE
3/9/16	Credit reporting		Incorrect information	Information is not mine		Company has	Exper
2/4/15	Debt collection	Mortgage	Cont'd attempt	Debt was paid			PHH
3/17/17	Bank account	Checking account	Making/receiving payments, sending money			Company has	UNI
3/8/12	Mortgage	Other mortgage	Loan servicing, payments, escrow account				Dite
3/27/13	Mortgage	Conventional	Loan servicing, payments, escrow account				NAT
2/25/14	Debt collection	Other (i.e. private)	Cont'd attempt	Debt is not mine			Nav
11/18/16	Mortgage	Conventional	Settlement process and communication		Started the refinance of home		AMER
7/16/15	Mortgage	Conventional	Application, originator, mortgage		In XXXX, I and my ex-husband		HSB
11/3/15	Debt collection	Medical	Disclosure violation	Right to dispute notice not		Company has	Reve
8/9/16	Credit reporting		Credit report	Problem with	I have disputed several accounts		EQU

UPDATED:

Date received	Product	Sub-product	Issue	Sub-issue	Consumer comment	Conclusion
3/12/14	Mortgage	Other mortgage	Loan modification	empty	empty	empty
10/1/16	Credit report	empty	Incorrect information	Account status	I have outstanding	Con
10/17/16	Consumer Lc	Vehicle loan	Managing the	empty	I purchased a	emj
6/8/14	Credit card	empty	Bankruptcy	empty	empty	emj
9/13/14	Debt collection	Credit card	Communication	Frequent or	empty	emj
11/13/13	Mortgage	Conventional	Loan servicing	empty	empty	emj
6/16/15	Debt collection	Medical	Improper collection	Contacted error	empty	Con
6/15/15	Credit report	empty	Credit report	Inadequate	An account	Con
11/13/15	Mortgage	Other mortgage	Loan modification	empty	empty	Con
10/21/14	Mortgage	Conventional	Loan modification	empty	empty	emj
4/14/17	Mortgage	Other mortgage	Loan modification	empty	empty	Con
3/30/12	Student loan	Non-federal	Repaying you	empty	empty	emj
2/3/16	Debt collection	Other (i.e. private)	Disclosure violation	Not given enough	This company	emj
1/7/15	Credit report	empty	Incorrect information	Account status	empty	emj
3/15/13	Credit card	empty	Closing/Cancellation	empty	empty	emj
7/18/16	Credit report	empty	Incorrect information	Account status	empty	emj
2/17/16	Debt collection	Credit card	Improper collection	Talked to a third	This	Con
11/7/14	Mortgage	Conventional	Loan modification	empty	empty	emj
4/17/15	Mortgage	FHA mortgage	Application, information	empty	empty	emj
3/9/16	Credit report	empty	Incorrect information	Information	empty	Con
2/4/15	Debt collection	Mortgage	Cont'd attempt	Debt was paid	empty	emj
3/17/17	Bank account	Checking account	Making/receiving	empty	empty	Con
3/8/12	Mortgage	Other mortgage	Loan servicing	empty	empty	emj
3/27/13	Mortgage	Conventional	Loan servicing	empty	empty	emj
2/25/14	Debt collection	Other (i.e. private)	Cont'd attempt	Debt is not resolved	Started the re	emj
11/18/16	Mortgage	Conventional	Settlement process	empty	Started the re	emj
7/16/15	Mortgage	Conventional	Application, information	empty	In XXXX, I am	emj

Competitive Research Programme (CRP Award No. NRF-CRP 10-2012-03).

REFERENCES

- [1] Patricia S. Abril and Robert Plant. 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan. 2007), 36–44. DOI: <http://dx.doi.org/10.1145/1188913.1188915>
- [2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. 2002. Wireless Sensor Networks: A Survey. *Comm. ACM* 38, 4 (2002), 393–422.
- [3] David A. Anisi. 2003. *Optimal Motion Control of a Ground Vehicle*. Master's thesis. Royal Institute of Technology (KTH), Stockholm, Sweden.
- [4] P. Bahl, R. Chancr, and J. Dungeon. 2004. SSCH: Slotted Seeded Channel Hopping for Capacity Improvement in IEEE 802.11 Ad-Hoc Wireless Networks. In *Proceeding of the 10th International Conference on Mobile Computing and Networking* (MobiCom'04). ACM, New York, NY, 112–117.
- [5] Kenneth L. Clarkson. 1985. *Algorithms for Closest-Point Problems (Computational Geometry)*. Ph.D. Dissertation. Stanford University, Palo Alto, CA. UMI Order Number: AAT 8506171.
- [6] Jacques Cohen (Ed.). 1996. Special Issue: Digital Libraries. *Commun. ACM* 39, 11 (Nov. 1996).
- [7] Bruce P. Douglass. 1998. Statecharts in use: structured analysis and object-orientation. In *Lectures on Embedded Systems*, Grzegorz Rozenberg and Frits W. Vaandrager (Eds.). Lecture Notes in Computer Science, Vol. 1494. Springer-Verlag, London, 368–394. DOI: <http://dx.doi.org/10.1007/3-540-65193-429>
- [8] Ian Editor (Ed.). 2008. *The title of book two* (2nd. ed.). University of Chicago Press, Chicago, Chapter 100. DOI: <http://dx.doi.org/10.1007/3-540-09237-4>

A.1 Literature Survey

A.2 Proposed Work

A.2.1 What Needed for Data Collection

A.2.2 Preprocessing

Component Structures

Magnetization.

A.2.3 How Mining Consumer Complaints is different from previous work

A.2.4 Ground-State Magnetization Determination and DMM Micromagnetic Simulations

A.3 Data Set

A.4 Evaluation Methods

A.5 Tools

A.6 Milestones

A.7 Results So Far

ACKNOWLEDGMENTS

This work was partially supported by the MIUR-PRIN 2010–11 Project 2010ECA8P3 “DyNanoMag” and by the National Research Foundation, Prime Minister's office, Singapore under its