Big Data Analytics in Biometric Identity Management

Robert W. Gasiewicz
Indiana University
711 N. Park Avenue
Bloomington, IN 47408
rgasiewi@iu.edu

ABSTRACT

The United States Government, through its collection and use of biometric data, has leveraged big data in order to protect its citizens and keep our country safe. The speed and accuracy with which this biometric data can be effectively matched to an identity can mean the difference between life and death, as well as the integrity of our institutions. This paper predominantly focuses on how the United States Government, collects, stores, and uses big data to facilitate solving crimes and to enhance national security.

KEYWORDS

i523, HID316, Big Data, Biometrics, Fingerprinting, 2-Print, 10-Print, Matchers, Matching Algorithms, DHS, Homeland Security, Border Security, National Security, Immigration, Terrorism, FBI, AFIS

1 INTRODUCTION

Across the spectrum, big data is rapidly changing the way we do business, the way we live, and the way governments around the world do everything they can to keep us safe in the face of an increasingly dangerous world. Long before the advent of big data, fingerprints were used as a means of forensic identification, but it wasn't until technology had progressed to the point to which these prints could be converted and stored in digital format, organized, and then matched against other stored data and even other databases, that this data truly became useful on the large scale that it is today.

Biometrics technology is changing rapidly, and with it, both the size and scope of data being collected. From 2-print to 10-print, iris to facial recognition, the demand for both data intensive processes and rapid matching have grown exponentially, and understanding how the United States Government uses biometrics is a case study in big data if there ever was one.

2 HISTORY OF FINGERPRINTING: THE ANALOG ERA

In 1858, a man by the name of Sir William James Herschel began using fingerprints as a means of identification [4] near Calcutta, India. This started as a means of not solving crimes, but preventing them; Sir William's aim was to thwart attempts at forging signatures - something that had begun to occur at epidemic proportions. Herschel also used fingerprinting to prevent the collection of pension benefits by relatives after the pensioner had deceased.

It wasn't until 1886 that Scottish surgeon, Dr. Henry Faulds, proposed the concept of using fingerprints to identify criminals to London's Metropolitan Police [3]. Incredibly, they dismissed his proposal.

By 1906, the concept of identifying criminals using fingerprints had made its way to the United States, first in New York City and then elsewhere throughout the country. In 1924, the United States Congress created the Identification Division of the Federal Bureau of Investigation (FBI) and 22 years later, they had processed over 100 million fingerprint cards. By 1971, this number had more than doubled [5].

3 BIOMETRICS ENTERS THE DIGITAL AGE

Before the 1960s and 1970s, fingerprints were stored on cards and expert examiners studied fingerprint features, or minutiae, such as ridges, enclosures, and bifurcations. Fingerprints were then filed according to the Henry classification system [1]. Processing was slow, taking weeks or even months and everything had to be done at one central processing facility. Big Data was perfect solution to this problem.

By the dawn of the 1980s, the completely analog system transitioned toward a more digital platform by storing filing codes on early computer systems. It wasn't until 1986 that the Automated Fingerprint Identification System was released commercially to agencies across the United States Government.

4 AUTOMATED FINGERPRINT IDENTIFICATION SYSTEM (AFIS)

In July of 1999, the AFIS or IAFIS system became a fully automated, nationalized computer system intended for enhanced and rapidly expedited matching capabilities. The AFIS system is not only a criminal and civilian database for fingerprints, photographs, as well as military and civilian data, it is also a matching system, providing either positive or negative identification of prints submitted against its cache of stored records. In addition to biometric identification, AFIS also serves as a means of biographic identification based on pieces of data such as name, date of birth, tatoos, various ID numbers, and other relevant personally identifiable information (PII).

As Simon A. Cole explains in his 2002 book, Suspect Identities: A History of Fingerprinting and Criminal Identification [1], AFIS can work in four of the following ways:

1) 2-print (left and right index finger) and 10-print (all ten of a person's digits) taken from a crime scene, body, or border checkpoint and can be checked against a database of other fingerprints

2) A single latent, or partial trace print can also be checked against a database of other fingerprints

3) A complete 2-print or 10-print image can be checked against other stored latent prints

4) So-called ”unsolved” prints, both latent and complete 2-print and 10-print images can be stored in the database and checked against any new subsequent additions.

Today AFIS is the largest biometric database in the world.

5 INITIAL ACHIEVEMENTS OF DIGITIZATION

With AFIS, the original intent of digitizing several hundred million fingerprint cards was to make it easier to do a job that was already being performed manually. As outlined above, it met two requirements: identify fingerprints and serve as a central reporting system on criminal history for the United States Government.

As time went on, AFIS began to earn additional credibility in other areas as well. It not only helped to improve the collection and identification process with regard to latent fingerprints, but it also forced the standardization process by which all fingerprints are collected, stored, and matched against. These standards are known as uniform biometric standards and were essential in enabling various government agencies to share data they collect.

In addition to saving the government and the environment an enormous amount of ink and paper by doing away with fingerprint cards, AFIS has also helped to expedite the pace at which criminals are able to be identified as well as how quickly cases are able to be adjudicated. Lastly, an additional immediately recognized benefit of digitization of fingerprint records has been the rapid improvement of digital image quality needed to more accurately match fingerprints.

6 BIOMETRICS AND BIG DATA

The ever-present question in the world of burgeoning big data is always: ”how is this useful?” Often large swaths of data are collected as a part of standard business processes, or, in this case, as a part of criminal investigations and only later are new uses found for the data that’s been gathered. As technology evolves new possibilities emerge and stewards of the data find new ways in which it can be used.

There are times, however, in which there are catalysts in addition to the steady march of technological advancement that force us to change the way we look not only our data, but at the world around us. After September 11th, 2001, the United States Congress passed the ”Homeland Security Information Act” which with the understanding that information systems for collecting biometric and biographical data were already in existence, must be efficient and should not be duplicated throughout the federal, state, and local governments. The U.S Department of Homeland Security was created in 2002, consolidating many disparate agencies under one roof and one new cabinet level position, reporting directly to the President of the United States.

Subsequent to this, it was incumbent upon the United States Department of Justice (DOJ) to use any means necessary to protect the United States from being subjected to any additional acts of terrorism. To accomplish this the DOJ would need to have other United States Government agencies working together to share information, but foreign law enforcement agencies as well.

7 ENHANCED BIOMETRIC DATA COLLECTION

Biometric Big Data got even bigger in 2003 when the recently formed U.S. Department of Homeland Security created the United States Visitor and Immigrant Status Indicator Technology (US-VISIT) program. In order to meet the ever-increasing demands to preserve and secure our national security, additional measures and enhanced collection at border crossings and at airports was undertaken. Prior to US-VISIT, as had been observed for hundreds of years, paper travel documents and biographical information could be easily forged, various systems were scattered across the U.S. Government and were not well-coordinated, and partner countries did not abide by the same sets of guidelines.

With the creation of the US-VISIT program, the digitization of both biometric and biographic details of individuals coming in and out of the U.S. ensured that these details could not be easily forged or altered. Specifically, the use of fingerprints, and moreover the ability to match them against the largest biometric database in the world in around 10 seconds, prevents untold hundreds of thousands of attempts by dangerous criminals and terrorists from obtaining visas or gaining entrance to the U.S.

By working closely with other agencies across the U.S. Department of Homeland Security, US-VISIT has the same access to crucial fingerprint data as:

1) Immigration and Customs Enforcement (ICE)
2) Customs and Border Protection (CBP)
3) FBI
4) Department of State (DOS)
5) U.S. Citizenship and Immigration Services (USCIS)
6) U.S. Coast Guard (USCG)
7) Department of Justice (DOJ), State, and Local Law Enforcement
8) Department of Defense (DOD) and Intelligence Community

This level of cooperation was solidified even further on October 25, 2005 with U.S. Presidential Executive Order 13388 [2]:

To the maximum extent consistent with applicable law, agencies shall, in the design and use of information systems and in the dissemination of information among agencies:

(a) give the highest priority to

(i) the detection, prevention, disruption, preemption, and mitigation of the effects of terrorist activities against the territory, people, and interests of the United States of America; (ii) the interchange of terrorism information among agencies; (iii) the interchange of terrorism information between agencies and appropriate authorities of state, local, and tribal governments, and between agencies and appropriate private sector entities; and (iv) the protection of the ability of agencies to acquire additional such information; and

(b) protect the freedom, information privacy, and other legal rights of Americans in the conduct of activities implementing subsection (a).

This E.O spelled out the sweeping changes that the U.S. Department of Homeland Security had already made to the way data was collected, processed, standardized, and matched against.

8 THE FUTURE OF BIOMETRICS AND BIG DATA

The future of biometrics and big data is bright. In the past decade, the U.S. Government has moved from 2-print to 10-print, with plans to begin using iris and facial recognition, as well as gait, to identify and neutralize threats. The move from 2-print to 10-print alone represented five fold increase in data storage needs. Storing detailed images of a person's eyes, their face, and the way they walk will require even more data storage capacity and the raw computing power to analyze it. Such advances are necessary to keep us safe in an increasingly dangerous world.

REFERENCES

[1] Simon A. Cole. 2002. *Suspect Identities: A History of Fingerprinting and Criminal Identification.* Academic Trade. (book).

[2] Information Sharing Environment. [n. d.]. Executive Order 13388. ([n. d.]). Retrieved October 4th, 2017 from https://www.ise.gov/resources/document-library/executive-order-13388-further-strengthening-sharing-terrorism-information-protect-an

[3] Henry Faulds. 1880. *On the skin-furrows of the hand.* Oxford University Press. https://doi.org/10.1038/022605a0 (book).

[4] William J. Herschel. 1916. *The Origin of Finger-printing.* Number ISBN 978-1-104-66225-7 in Fundamental Algorithms. Oxford University Press. (book).

[5] U.S. Marshals Service Website. [n. d.]. Fingerprint History. ([n. d.]). Retrieved October 3rd, 2017 from https://www.usmarshals.gov/usmsforkids/fingerprint_history.htm

9 BIBTEX ISSUES

- Warning–empty address in Cole2002
- Warning–empty year in ISE2017
- Warning–empty address in Faulds1880
- Warning–empty address in Herschel1916
- Warning–empty year in USMarshals2017
- (There were 5 warnings)

10 ISSUES

- no title
- you do not use “quotes” properly
- you need to *emphasize* and not “quote”
- this is cool
- Have you written the report in the speciffied format?
- Have you included an acknowledgement section?
- Have you included the paper in the submission system (In our class it is git)?
- Have you speciffied proper identiffication in the submission system. THis is typically a form or ASCII text that needs to be ffilled out (In our case it is a README.md ffile that includes a homework ID, names of the authors, and e-mails)?
- Have you included all images in native and PDF format in the submission system?
- Have you added the bibliography ffile that you managed (In our case jabref to make it simple for you)?
- In case you used word have you also provided the jabref?
- In case of a class and if you do a multi-author paper, have you added an appendix describing who did what in the paper?
- Have you spellchecked the paper?
- Are you usingaandtheproperly?
- Have you made sure you do not plagiarize?
- Is the title properly capitalized?
- Have you not used phrases such as shown in the Figure below, but instead used as shown in Figure 3 when referring to the 3rd ffigure?
- Have you capitalized fiFigure 3fi, fiTable 1fi, ... ?
- Have you removed any ffigure that is not referred explicitly in the text (As shown in Figure ..)
- Are the ffigure captions bellow the ffigures and not on top. (Do not include the titles of the ffigures in the ffigure itself but instead use the caption or that information?
- When using tables have you put the table caption on top?
- Make the ffigures large enough so we can read the details. If needed make the ffigure over two columns?
- Do not worry about the ffigure placement if they are at a different location than you think. Figures are allowed to ffloat. If you want you can place all ffigures at the end of the report?
- Are all ffigures and tables at the end?
- In case you copied a ffigure from another paper you need to ask for copyright permission. IN case of a class paper youmustinclude a reference to the original in the caption.
- Do not use the word fiIfi instead use we even if you are the sole author?
- Do not use the phrase fiIn this paper/report we showfi instead use fiWe showfi. It is not important if this is a paper or a report and does not need to be mentioned.
- Do not artifficially inffate your paper if you are bellow the page limit and have nothing to say anymore.
- If your paper limit is 12 pages but you want to hand in 120 pages, please check ffirst ;-)
- Donotusethecharacters & # % _ put a bakslash berfore them
- If you want to say and do not use & but use the word and.
- Latex uses double single open quotes and double single closed quotes for quotes. Have you made sure you replaced them?
- Pasting and copying from the Web often results in non ascii characters to be used in your text, please remove them and replace accordingly.