Continuous Multimodal Authentication

Nico Tripeny, Macintyre Sunde, Aaron Kriegman

The goal for our final project is to improve and expand our work from labs two, three, and five by building a continuous authentication system using gate and swiping biometrics. For swiping, we use more features to better understand which are most discriminating and hopefully get better results. For the gate biometric, we extend our authentication system to use more data in its analysis of users. Another focus of the project is to better understand ways to use continuous authentication, as the traditional single time authentication assessment will not be as effective. These three goals will combine to form a continuous authentication system that can be used both when a smartphone is being used and when it is not. The gate biometric can also combine with smartwatches when not being used.

Swiping

For the swiping biometric, our original work was rather limited. We used few features which lead to a rather inaccurate authentication system. For this project, we usefeatures similar to those in 'Benchmarking Touchscreen Biometrics for Mobile Authentication' [1]. Specifically, for the data on points we find the following features:

- slope of best fit lines
- difference in x values
- difference in y values
- ratio of the above differences
- distance from start point to end point

- the arclength of the swipe
- Speed of swipe
- pressure, area of press using the mean, standard deviation, and first, second, and third quartile
- Duration of swipe

For the initial data, we take windows of small amounts of swipes. For each swipe, we find these features then average them. Since many of the "swipes" are actually just single or double points, this window system allows to avoid this noise, similar to a rolling average. This allows us to still have many data points for each user to check each person against. This is helped by the fact the database has an extremely large amount of swiped for each person, so we felt checking against each one was unnecessary.

The performance of the swiping system was better than random but still variable. The chart below was created by sklearn's classification_report function. It outputs the performance of each user as well as the average performance across all users. The precision and recall were both around 0.77. It accepts around three times as many genuine as imposter attempts, and is correct about 75% of the time. So while this was not as high as we would have liked, the results are still much better than we achieved in lab 5. Deciding to extract more features helped a great deal in our systems reliability. It is worth noting that in testing, there are many more imposters than genuine attempts. We used 30 users, so there are around 30 times more imposters than genuine attempts. When we accounted for this in our testing by using the class imbalance algorithm. As a result, The precision and recall increased to around 0.96. This seems to imply that perhaps the algorithm is better than we originally thought.

One interesting aspect is that the precision and recall varied a great deal across the users. To check that this is not simply randomness, we run the process multiple times. This accounts for any one time bias in either the class imbalance solution or the feature selection favoring one user randomly. We users tend to keep relatively similar scores each time. For example, user 1 was typically around or above 0.8 for both precision and recall, while user 10 was rarely above 0.6 for either and was often below 0.5 for both. This seems to correspond to the idea that some peoples biometrics tend to be more unique than others. Because the scores varied so much, we can see that even if 0.7 precision was good enough for our system, what we have developed is still unacceptable for widespread use.

	accuracy	macro avg	weighted avg
precision	.77	.73	.78
recall	.77	.73	.78
f1-score	.77	.72	.77
support	.78	336	336

There are many steps we could take to improve the system. To start, we would want to try more classification algorithms to find the different scores. We are currently only using KNN and it is likely that other algorithms could improve our scores. Furthermore, it seems likely that trying other feature selection algorithms could give us better results as well.

In addition to algorithms, we made many choices throughout that we can experiment with. If there was more time, an obvious step would be to use different distance metrics on our data. Currently, we are only using Euclidean distance between the feature vectors, and it would be worth trying Manhattan distance, or another metric to tune our system. We also could find

ways to better understand the most discriminating features and scale those more for our benefit beyond feature selection. Another option would be to tune the window length used and see how that affects the accuracy. Having smaller windows would increase the data but also increase the noise. If we did lower the window, it would probably be best to more thoroughly preprocess our data to delete and singleton data points, where the swipe is in fact a tap. Lastly, we could also extract more features in an attempt to find other discriminating aspects of the swipe. One obvious feature that we do not use is the swipes acceleration. We also use no ratios between any two features. It seems possible that neither area nor pressure are useful features on their own, but ratio could be. Exploring this would undoubtedly help create a stronger system.

Gait Feature Extraction:

Gate feature extraction is a continuation of lab 2 where we take the x-data from an accelerometer to authenticate users by their walking pattern. In this project, we use x, y, z data from gyroscope and accelerometer to authenticate the user. This is successful when fitting the model to small numbers of users; we are unable to run large numbers of users because it takes too much time to compute. However, the results from the model are promising.

Specifically, the precision and recall both had values around 0.91. This implies that the system is much more likely to accept a genuine attempt over an imposter, and more likely to accept rather than reject a genuine attempt. While the numbers presumably decrease as more imposters are added in, these numbers show promise for a biometric system. It is also much more successful than our first attempt in labs 2 and 3, which makes sense as we are using more aspects of the gate, and fitting the model using the sklearn environment.

	User10	User2	User3	User4	User5	User6	User7	User8	accuracy	macro avg	weighted avg
precision	1	0.66	1	1	1	0.75	1	1	0.91	0.92	0.93
recall	1	1	1	1	0.66	1	1	0.83	0.91	0.93	0.91
f1-score	1	0.8	1	1	0.8	0.85	1	0.90	0.91	0.92	0.91
support	2	2	2	3	3	3	2	6	0.91	23	23

The structure of Gait Feature Extraction consists of a Params class to store essential and mod-able inputs, a custom distance class using the package fastdtw, and the main function. The code works by deserializing the preprocessed data, partitioning it into training and testing segments, fitting a model, and finally training and testing to return the chart above. Most of this is done within the sklearn environment.

The most challenging part of gait authentication is understanding how to use sklearn and fastdtw effectively in order to save time coding. These packages are effective, but the they also run too slowly, making it difficult to test the system across many users.

There are many directions we can take future work. The first obvious next step would be to run on a large set of users. As said earlier, the length of time that the analysis took made this unreasonable for the time we had, but it would undoubtedly give us more insight into the successes and shortcomings of our model. Another interesting question to answer is what features are most useful for the system. Gate does not have feature extraction in the way swipe does, but as we use multiple sensors, it is possible that one axis is much better at discriminating between users than the other two. This would be worth exploring in future research to increase

the reliability of our program. It may also be worth attempting to do more traditional feature extraction and combining that with our current methods.

Another question one could ask for both swipe and gate is what users are harder to secure. If we used a larger sample set, we may be able to see what characteristics of gate or swipe make a person's false accept rate higher, and general biometric worse.

Concluding Remarks:

While neither system is perfect, both have promising aspects and are large improvements on our prior work. There are also many steps we could take going forward to continue this project. The major goal we hoped to achieve at the beginning of this project was an idea for a continuous authentication system using both gate and swipe biometrics. Sadly, we were unable to fully explore the results and analyze everything we found. This prevented us from understanding how one could really implement these systems or generally what use they could have. It is clear from our data that neither would serve as a one time authentication system. They are both far too inaccurate for that setup and are nowhere near other biometric systems for that. However, both do hold promise as a continuous system.

To explore this more, we need to better understand effective ways to implement a continuous system. One question, for example, is if we feed the system ten genuine attempts, how often will it correctly categorize at least 7. We would compare this to if we feed 10 imposter attempts, how many times does it incorrectly accept 7. These numbers are arbitrary, but they do explain the ideas of continuous authentication. Unlike entry only authentication, it is ok for a continuous system to occasionally accept an imposter or reject a genuine user, as long as it does

not do this more than a certain threshold. So perhaps our most important path for ongoing research would be better exploring questions like these, both theoretically and prac. practically

Bibliography

[1] https://arxiv.org/pdf/1501.01199.pdf Benchmarking Touchscreen Biometrics for Mobile Authentication

[2] I. C. Stylios, O. Thanou, I. Androulidakis, and E. Zaitseva, "A review of continuous authentication using behavioral biometrics," in ACM International Conference Proceeding Series, 2016, vol. 25-27-September-2016, pp. 72–79.

[3] Amith K. Belman, Li Wang, Sundaraja S. Iyengar, Pawel Sniatala, Robert Wright, Robert Dora, Jacob Baldwin, Zhanpeng Jin, Vir V. Phoha, "SU-AIS BB-MAS (Syracuse University and Assured Information Security - Behavioral Biometrics Multi-device and multi-Activity data from Same users) Dataset ", IEEE Dataport, 2019. [Online]. Available: http://dx.doi.org/10.21227/rpaz-0h66. Accessed: Apr. 28, 2020.