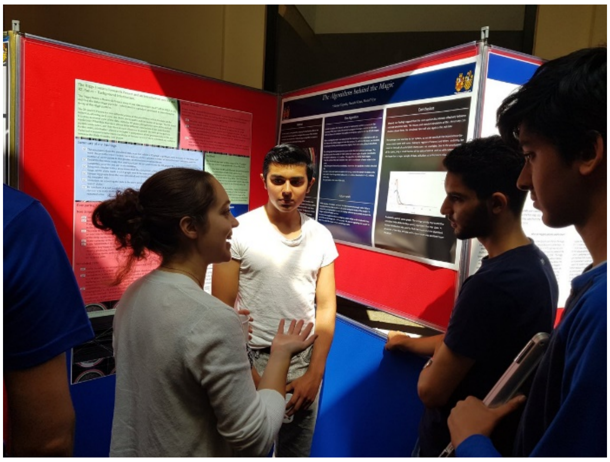
Proceedings of the First Higgs Hunters Schools' Conference

University of Oxford

June 2018



Photos from the Poster Session at the Higgs Hunters Schools' Conference at Oxford Physics Department

Editor's introduction

It's rare for students at high school to be allowed to do real research. So much of school science is about absorbing and understanding existing scientific knowledge that little space can be left for exploring the truly unknown.

But the core of the activity of science is not about absorbing a canon of information, it's about learning something totally new by addressing questions where the answer is not known by anyone.

The Institute for Research in Schools (IRIS) has been at the forefront helping UK school students perform independent research on exactly such open scientific questions. They are helping unlock the potential of school children to perform independent research – a potential that is greatly underestimated in the scientific community, and perhaps also in schools.

The Higgs Hunters schools partnership was established between the University of Oxford and IRIS to provide UK school students access to one of the flagship scientific projects of our time – the ATLAS experiment at the European Laboratory for Particle Physics (CERN).

CERN's 27km-long Large Hadron Collider smashes together protons at the highest achievable energy. The new particles created in the collisions are recorded by different components of the ATLAS detector. Some of the most interesting of those images were analysed by tens of thousands of citizen science volunteers working with the HiggsHunters.org project.

The Zooniverse citizen science web platform allowed the general public access to scrutinise images which might be hiding the tell-tale signs of the next breakthrough in particle physics.

ATL-OREACH-PROC-2019-002
29 May 2019

Much of the analysis at CERN is done by computers, but Zooniverse has shown that humans have an innate ability to spot the unusual or unexpected features.

The Higgs Hunters school students, whose work is summarised in what follows, were working on the second stage of the project. They were given full access to the citizen scientists' classifications of the images – records of features identified by the citizen scientists as being either (a) "weird" or (b) consistent with a new long-lived particle, outside the standard Model of particle physics.

The school students were asked to explore the data using whatever methods they saw fit, and to develop and address whichever questions they themselves thought to be most interesting.

The results are quite fascinating. The students have addressed a very broad range of scientific questions, ranging from "physics" questions about the nature of the collisions, through to the sociological questions about how the biases and experiences of the human citizen scientists might affect the results.

The methods used ranged from large-scale surveys of the images themselves, to analysis of the precision of citizen scientist clicks, and use of highly performant SQL databases and artificial intelligence techniques to store and classify the data.

In each case the students addressed open questions where nobody, either from the scientific community or elsewhere, knew what they would find.

They presented the results of their enquiries in a poster session to researchers at the University of Oxford physics department during a two-day schools conference in June 2018. Some of their findings are summarised in the proceedings which follow.

I hope you find them as fascinating as I do.

Alan Barr

Professor of Particle Physics, University of Oxford
Tutor in Physics, Merton College, Oxford

Identifying Weird Events

Elika Charlton, Kalle Cohen,
Alexander Stranescu,
Sula Harding-Cornish, Maxine
Chapman, Fern Goldsmith[1]
Camden School for Girls

Introduction:

Einstein's most famous equation, $E = mc^2$, shows that energy (E) is equivalent to mass (m) multiplied by the speed of light squared (c^2). This means that if energy is high enough it can actually be transferred into small mass. The world's biggest particle accelerator, the Large Hadron Collider, based at CERN in Geneva, accelerates protons to almost (99.999999%) the speed of light. At this incredibly fast speed the protons have the required tremendous amount of energy to be transferred into mass in the form of tiny particles, which are 'fundamental' as they cannot be broken down further, when they collide with each other.

However, what actually mediates this transfer from energy to mass? What actually gives these fundamental particles their mass? In 1964, Peter Higgs theorised a particle that would mediate this transfer of energy to mass, hence it was called the Higgs Boson, and in 2012 it was detected at the Large Hadron Collider. This Higgs Boson has been observed to decay via five possible decay modes into fundamental particles. However, it has been theorised that it may decay in a sixth mode - first decaying into a smaller 'baby' Higgs Boson before that further decays into the smallest fundamental particles.

These smaller baby Higgs particles will have less mass and exist long enough to travel from when the Higgs Boson decays into it, to when it decays itself into smaller particles. This will result in an off-centre vertex on images captured in ATLAS, where the baby Higgs travelled from the centre of the decay of the Higgs boson, before decaying into the smaller particles which are detected. It is very difficult for computers to analyse images to spot something visual like an off-centre vertex, so 40,000 images of collisions from the ATLAS experiment at the Large Hadron Collider were uploaded for the general public (citizen scientists) to analyse and see if there were any off-centre vertices in any of these images. These images included real and simulated data to test the reliability of the general public's analysis or spot any patterns in how the general public interpreted the data. Each image could also be viewed from three different projections (angles) in the XY, XY Zoom and RZ. Members of the general public viewing these images can click on them and label points as 'weird', or if there's a vertex they can count how many tracks come from the vertex. In our group we each analysed images that had been labelled 'weird', especially the ones that had been labelled 'weird' by multiple users. This was to see if any of these images actually had a baby Higgs in them (which was incredibly unlikely), or at least to see what kind of features people labelled as 'weird' and if there were any patterns in people's labelling.

Aims:

To understand why citizen scientists labelled certain features as 'weird' and see if there was a pattern in what they labelled as 'weird'.

To analyse the effectiveness of citizen scientists by seeing if the images they labelled as 'weird' really did warrant further investigation or not.

[1] Teacher

Method:

1. Change how the data is sorted to group images of the same event and if they are 'weird' together.
2. Identify images that have been clicked on and labelled as 'weird' by at least more than one person and plot the coordinates of these clicks as a scatter graph on Excel.
3. Change the intervals on the axis of the graph so that both scales are from 0-1024, the vertical y axis is flipped and that the plot area is set to a square shape.
4. Set the background of the plot area as the image that was clicked on to compare the coordinates of the plotted clicks to what was on the image at that point.
5. Analyse what was interesting or unique about that point on that image that made someone want to click on it and label it as 'weird'.

Results & conclusion:

None of the clicks explored in this way were found to identify any off-centre vertices. Instead, the events had little in common. We concluded that this could be because the explanation of what the citizen scientists are told to click on when they first visit the website is brief. Everyone on our team visited the website and tried clicking on some images ourselves and initially, before we had practice classifying the images, we struggled with what exactly to look for and click on. It would be interesting to see if the objects identified as weird might become more interesting as the citizen scientists gain experience with the data.

The Reliability of Citizen Science Data

N. Bush, H. Jolly, and Tsun J
Watford Grammar School for Boys

For citizen research to be useful in a scientific context, the validity of the data must be proved. For the Zooniverse project, valid data constitutes users frequently clicking in the correct regions. By computing the mean distance of each click to the correct simulated feature, it is clear that the majority of users were consistently accurately at identifying features, and therefore the data has substantial scientific value.

Introduction

On the 4th of July 2012, the Higgs Boson was discovered at CERN's Large Hadron Collider (LHC). This event landmarked possibly one of the greatest advancements in particle physics because the Higgs Boson had answered the long-sought question of why certain fundamental particles have mass. The data from the LHC however, may suggest that the uncharged Higgs Boson could have decayed into charged 'Baby Higgs' particles. Supporting this phenomenon, the images from the ATLAS experiment display missing transverse momenta and off-centre decay vertices from the proton collisions. As a result, understanding the possible behaviour, trajectory and lifetime of these particles is important to uncover new properties of the Higgs and possibly extend the Standard Model [2].

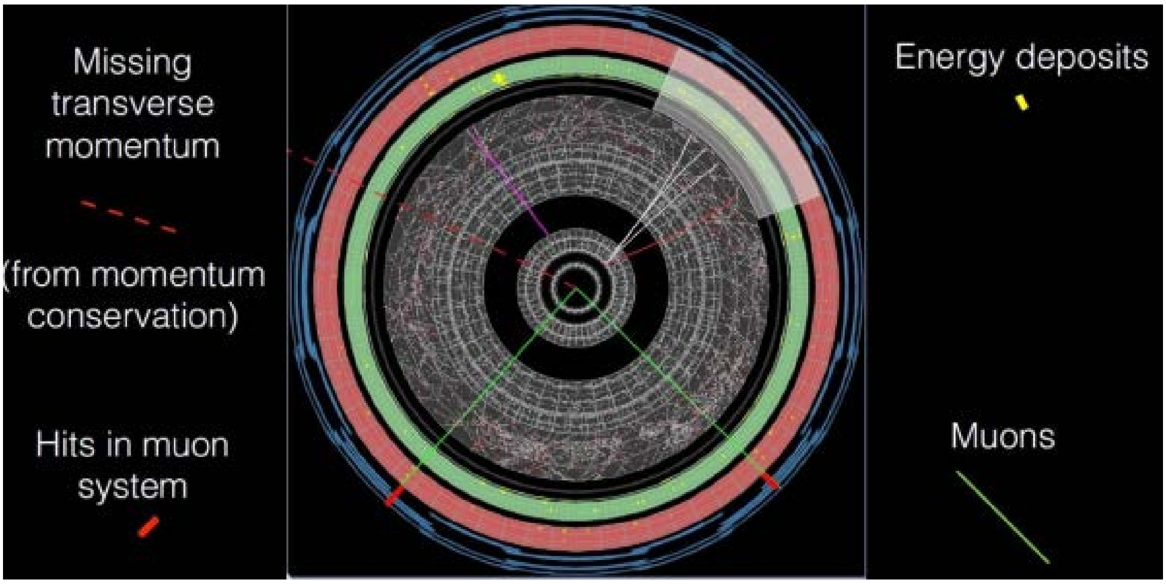


Figure 1 One of the images provided by Zooniverse as part of the simulated data

Together with people at Zooniverse ('the world's most popular platform for [citizen research]' [1]), ATLAS researchers formulated and compiled a set of simulated data which was examined by about

20,000 participants [3]. The participants were presented with a series of images to locate off - centre tracks. The Higgs Hunters Schools Project, organised by the Institute for Research in Schools (IRIS), allows us to test a particular hypothesis surrounding this data.

We understood that in order to formulate accurate conclusions surrounding the data from the ATLAS experiment, one must work with inherently reliable data. As a result of this understanding, we believe it's important to validate the reliability of the citizen scientists' classifications of the simulated data before conducting further experiments.

Method

By iterating over a list of user clicks that identify decay points, we calculated the pixel- distance from those points to the true coordinates. The histogram of pixel-distances is shown in *Figure 3*.

We iterated the data again to produce a value of mean and standard deviation to produce *Figure 2*.

Analysis

Our first graph *(Figure 2)* shows most users (indicated by dots) clustered near the origin, showing a low standard deviation and low mean distance. This means that on average most users were able to identify the legitimate decay vertices with excellent consistency. Producing the graph also allowed us to find a small cluster of anomalous users with a high mean distance and a low standard deviation, meaning that they were consistently incorrect at determining the true coordinates of decay.

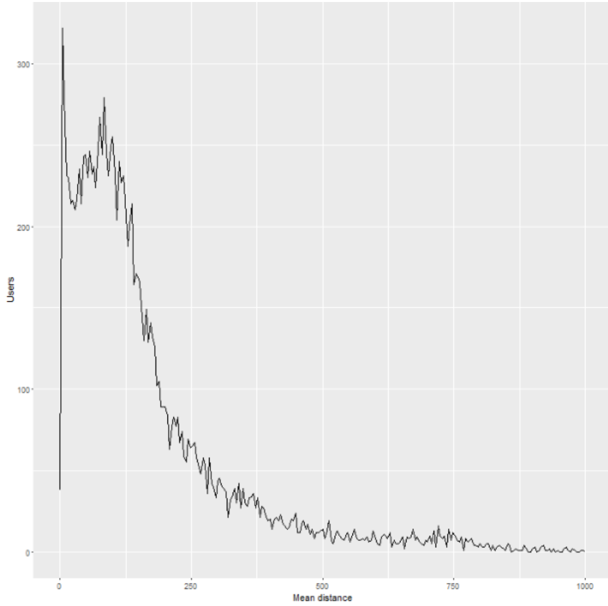


Figure 2 A Graph to Show *Mean Distance* (x-axis) against *Standard Deviation* (y- axis)

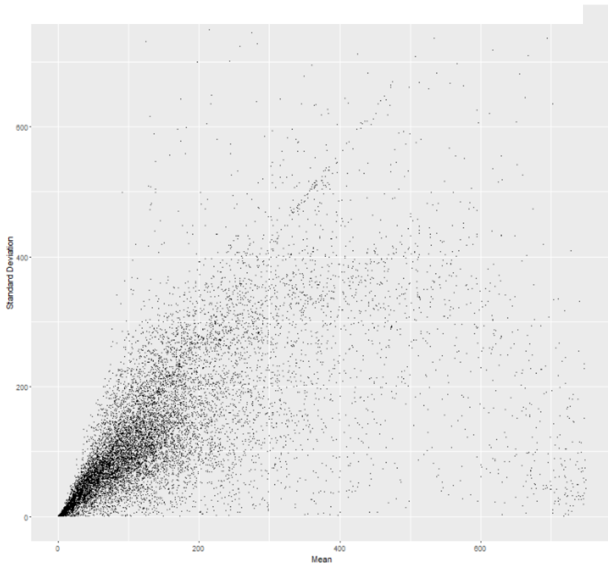


Figure 3 A Graph to Show *Mean Distance* (x-axis) against *User Count* (y- axis)

Before we attended the Higgs Hunters Competition to visit Oxford University, we believed that after recognising these users, the most appropriate action would be to omit their results for analysis in any future non-simulated data that Zooniverse wanted to produce. However, after speaking to a member from Zooniverse, we learned that these citizen scientists may correct their mistakes in the raw data. To our surprise, a faint linear dependence between *Mean and Standard Deviation* of approximately *y=2x* was found *(*Figure 2*)*.

Although we are unable to establish why the trend exists, we leave this as future work for the remaining duration of the Higgs Hunters Programme.

Our second graph *(Figure 3)* shows that a high density of users have a low mean distance away from true points of decay. This means that a lot of users are good at pinpointing decay points. As the pixel- distance from the decay point increases, the user count for those distances decreases.

Because the distribution of clicks in the simulated data approximately follow the normal distribution with the majority of clicks being less than 100 pixels distance from the true coordinates, we would also expect the raw data to follow the same distribution; knowing that the participants in simulated data and non-simulated data were selected from the same population (by finding similar Zooniverse ID in both tests).

Conclusion

We conclude that citizen science programs are reliable enough to extract reliable data for scientific development and research because the majority of the approximately 20,000 users were consistent at accurately identifying points of decay.

In the future, using this information and knowing that the same people participated in non-simulated data, we can determine true points of decay by selecting reliable users with small standard deviation. Due to the large volume of consistently accurate users, we plan to combine our averaged findings with the AI group at WBGS. Using their predictions to find points of decay, we could validate the AI predictions by matching them with our findings.

Acknowledgements

We would like to thank the Institute of Research in Schools for providing the data, image exercises and the opportunity to be part of the Higgs Hunters Conference at Oxford University. We would also like to thank Oxford University for accommodating us at Merton College and finally to the teachers that supervised the project at our school – Dr R Cerezo- Balsera and Dr C Cianci.

References

[1]- https://www.zooniverse.org/about

[2]- CERN Website- https://home.cern, New Scientist- Volume 238; No. 3182- 16th June 2018 https://www.newscientist.com/issue/3182/

[3]- http://www.researchinschools.org/higgs_hunters/home.html

Comparing click accuracy in different types of image

Jared Richard

Haberdashers Askes Boys School

Introduction

The Large Hadron Collider (LHC) at CERN, the world's largest particle accelerator, accelerates protons to 99.999999% of the speed of light. Einstein's energy – mass equivalence, $E = mc^2$, tells us that these high energies will be converted into a small mass in the form of many fundamental particles (particles that cannot be broke down further). The Higgs field causes these particles to slow and gives them mass relative to their interaction with this field. E.g. top quarks move slowly, interact greatly, hence they have a large mass. However, photons move quickly, hardly interacting with the Higgs field which is why they are massless. 2012 saw the detection of the Higgs Boson, a particle that is an excitation in the Higgs field and was believed it enforce it, completing the standard model. The Higgs Boson decays to fundamental particles in five modes already observed, however the Higgs boson may decay in another previously unseen way; first to two Baby Higgs which then themselves will decay to fundamental particles. These Baby bosons exist for longer and are lighter, so they travel further from the collision point before decaying, producing off-centre vertices. These displaced vertices act as proof for the Baby bosons, however the reliability of the analysis of their positions must be tested. 40,000 events (collisions) were uploaded for public analysis and each event had 3 projections (viewing angles): XY, XY Zoom and RZ. These images contain simulated and real data to help determine the patterns and reliability of public analysis. I will analyse data from proton-proton collisions at CERN to assess the reliability of the citizen scientists' analysis of the public data to identify the most accurate projection for future investigation. This will provide a conclusion on the most effective method for future analysis for the Baby Higgs.

Aims

The aims of my investigation are to understand the images and what constitutes as evidence for the Baby Higgs. Using this understanding to compare "click data sets" of citizen scientists with original images to investigate the reliability of the analysis from the citizen scientists. Following this comparison, to then identify common errors from public analysis, highlighting imperfections in CERN's data and how this affected the public analysis. As the common errors and imperfections have been identified, to then select data which I believe is reliable and use this to suggest existence of the Baby Higgs.

Method

- Use a systematic sampling technique to select every 10,000th simulated event, using its zooniverse ID.
- For each event selected, analyse the accuracy of the XY zoom and RZ projection.
- To do this: insert the raw data of the users' click coordinates for a chosen event into excel. (This data should include the coordinates of the clicks made by users and the coordinates of the true values of decays.)
- Select the X and Y coordinates of the clicks of all users for that particular event,

inserting this data into a scatter graph with scale axis 0-1024 and the vertical y axis flipped.

- Underlay the image projection as the background so that it lines up with the coordinates and the accuracy of the clicks can be compared.
- Identify any imperfections with the chosen event, take these into consideration when analysing the accuracy.
- Calculate the percentage error of each click and take an average percentage error in accuracy for each event.
- Take an average of the percentage error for the XY and RZ projection.
- Compare the percentage errors and infer which projection leads to the most accurate analysis (lowest percentage error).

Analysis:

XY zoom projection:

The XY zoom projection shows a magnified transverse cross section of the Atlas detector, essentially a face on view. This produces a more understandable view of the decays and therefore will likely provide more reliable public analysis.

$$\text{Average \% error from the selected sample of XY zoom projections}$$
$$= \frac{(47.8) + (2.36) + (3.63) + (0.37) + (72.5) + (61.1) + (38.7) + (16.8) + (1.59) + (74.1)}{10}$$
$$= 31.9\% \text{ error}$$

RZ zoom projection

The RZ projection shows a longitudinal cross section of the Atlas detector, a side on view. There is a greater error of uncertainty when images are transferred to the RZ projection and its distorted appearance will likely produce less accurate public analysis.

$$\text{Average \% error from the selected sample of RZ projections}$$
$$= \frac{(10.4) + (22.0) + (98.6) + (20.4) + (91.3) + (19.5) + (31.8) + (3.12) + (21.2) + (97.7)}{10}$$
$$= 41.6\% \text{ error}$$

The percentage error of coordinate analysis in the XY zoom projection is lower than the RZ projection (31.0 < 41.6). Therefore, there is sufficient evidence to accept the XZ projections as more accurate and reliable.

A comparison between the percentage error of the citizen scientist's analysis in the two projections from the selected sample shows that, on average, the percentage error in XY was lower. There are many factors that affect this, however a significant one is the visibility of XY and how it is generally more easily understandable from a citizen scientist's perspective. As mentioned by users on ATLAS online, the RZ projection is often found to be confusing and distorted, producing large uncertainties in coordinate transformation. This can lead to incorrect clicks which will alert LHC of "weird" decays when in fact the scientist was not able to correctly decipher what they were seeing.

Conclusions

There are several potential solutions, for example: offering extensive training would allow users to be more comfortable with handling RZ projections, therefore producing more accurate analysis, however this is costly and time consuming. Less people may be willing to participate if they must train for it. Another option is to use a computer algorithm to analyse any irregular decays, however the simulated and non-simulated projections had many imperfections, as highlighted by scientists, which would confuse the algorithm possibly causing a significant systematic error. Therefore, human analysis is arguably a more reliable option.

To conclude, because XY shows a greater accuracy, any decays identified as “weird” in the XY projection should be further analysed as they have a greater chance of providing substantial proof for Baby Bosons.

Infrastructure for large-scale and efficient processing of Higgs Hunter data

Charles Thomas
Shipston High School

Introduction

The Higgs Hunter project was a citizen science project created with the aim of classifying data from CERN in order to look for exotic decays of the Higgs Boson. The data from this project has been made available for processing and this document outlines the pipeline I developed in order to search for these decays as well as the relevant background information.

Many of the particles within the standard model have mass and the explanation of why they have mass comes in the form of the Higgs Field. The Higgs Field was predicted in the 1970s but was recently confirmed in 2012 with the discovery of the Higgs Boson at the Large Hadron Collider (LHC) at CERN, Switzerland. The Higgs Field pervades the entire universe and in the same way the photon carries the electromagnetic field the Higgs Boson carries the Higgs Field and is responsible for providing other particles with mass.

The Higgs Boson itself is believed to be unstable and therefore it should decay into other particles. These decays are officially known as 'exotic decays of the Higgs Boson' but are often called baby Higgs or Higglets. The LHC has continued to look for these Higgs decays, however, the problem is that there is a very large amount of data to sort through. While one solution to this problem is to use computers to search for patterns in this data it has been suggested that humans are more accurate when it comes to looking for odd events in this data. Therefore, some of the data from CERN was published online through a project called Zooniverse.

Through the Zooniverse platform users were shown either images either constructed from data from an actual collision or data from a simulation of what a Higgs decay may look like. Users were then asked to click on interesting events which may be include muons or evidence of missing momentum. This data was then made accessible, via the Institute for Research in Schools, for analysis.

Analysis

My analysis of this data can be split broadly into two sections. The first section dealt with analysing how users dealt with the simulated data and how accurate they were. The second data looked at using the data from accurate users to look for potential exotic decays.

Computing Infrastructure

In order to enable my analysis I relied on several computing tools, the most major of which was Microsoft SQL Server 2017. This database was chosen for both its high performance as well as its array of additional features its particular dialect of SQL (known as T-SQL) which makes handling and processing the data significantly easier. The other main tool I used in my analysis was the Python programming language alongside the scikit-learn library, which allowed me to quickly write the code necessary for my analysis as well as giving me access to a wide variety of performant data science tools.

Process

The first step of my analysis was to load both the non-simulated and simulated data into Microsoft

SQL server. This was straight forward although there were two caveats. Firstly, there are was no way of uniquely identifying rows in the data set so I had to add an integer primary key. The other caveats was that having previously explored the data in Excel was that some of the data had been altered automatically by Excel which initially raised some questions during the import.

Once the data was loaded into SQL Server, the next step was to work out how accurate users were when clicking on the simulated data. Each piece of simulated data contains the true positions of the decay which was simulated therefore to assess whether a user was accurate or not on a particular image I placed a radius around each of the true decay positions and if a user's click was within this radius then they were classed as being accurate on that image. A total score for their accuracy was then calculated as number of clicks classed as accurate over the total amount of their clicks on simulated data. This is then converted into a percentage.

In order to ensure that the percentage accuracy score is valid I excluded users who had clicked on less than 10 pieces of simulated data. This because user's were shown varying amounts of simulated data therefore it is possible that a user was shown only one piece of simulated data which they could have clicked randomly on. Their random click could be near the true decay which could lead to them being classed as being perfectly accurate when in fact their clicks are random.

Once users are ranked by their accuracy I then take all the users who are more than 90% of the time and process the images created from actual data that these user's clicked on. In these images I look for clusters of clicks because where multiple accurate users are clicking should be interesting events which could even be evidence of long-lived "Higglets".

In order to look for clusters of clicks I used the Mean Shift Clustering algorithm as implemented in scikit-learn. The main advantage of the Mean Shift Clustering algorithm is that it does not look for a particular number of clusters in data, rather it estimates how many clusters there are based on the data. This is in comparison to the more common k-means clustering algorithm which looks for specific number of clusters (denoted by the letter k) which has to be specified by the user. In this use case this is a major advantage because each image may have a different number of clusters and one wants to avoid looking at each image by hand as this would defeat the purpose of the automation.

The trade-off for using the mean shift clustering algorithm and its ability to determine the number of clusters is that it has quadratic complexity (often written as $O(n^2)$) - this means that if on 4 data points it takes 10 seconds to run, on 8 data points it takes 100 seconds. Therefore it may not be suitable for clustering a large volume of clicks, however, on this data set the number of clicks per image is small therefore this is not a concern, particularly as each image can be processed in parallel.

A possible future alternative to the Mean Shift clustering algorithm is the DBSCAN clustering algorithm which also determines the number of clusters in a data set but additionally removes noise. Removing noise means that in does not include outliers when forming the clusters. This could be useful in this context because some users may click of the screen or click randomly. Under the mean shift clustering algorithm this clicks shift the centre of the cluster, however, the DBSCAN clustering algorithm would exclude these points therefore potentially giving more accurate locations of clusters.

While this process has now been implemented it has yet to run at full scale so as of time of writing there are no results to report, however, there are several future improvements which could be made to both allow the process to scale over the whole data set as well as produce more specific interesting events to investigate further.

In order to scale the process to run over the full data set it would be necessary to process the clustering of the clicks over multiple machines. In order to this I would use a piece of software known as Docker in combination with a piece of software called Kubernetes to run multiple instances of the clustering program over multiple machines over a network. Additionally, this would require the use of a data store such as Redis to distribute the images to each process.

The process itself could be expanded to also look for clusters of clicks in three dimensions. This is possible because for each collision there are two projections: XY and RZ. These two projections occur in different planes therefore theoretically can be combined to search for clusters in three dimensions.

Another improvement could be to look for trends which affect a user's accuracy. For example, are users more accurate on a specific projection? This could then be used to filter down which images generated from actual data are prioritised.

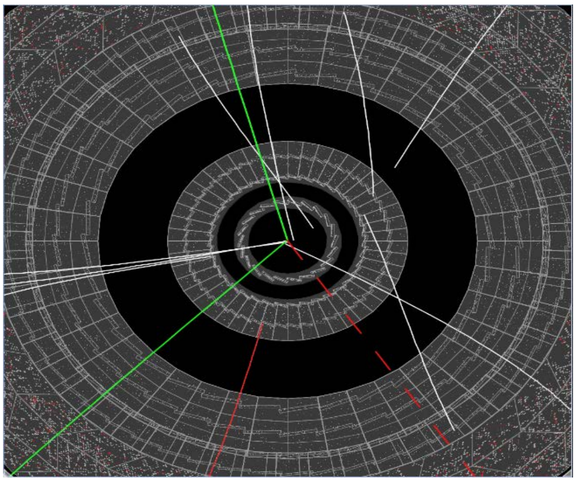
Conclusion

In summary, this document outlines the general pipeline for processing the images from the citizen science project Higgs Hunters and how one can use clustering algorithms in order to find evidence of potential decays of the Higgs Boson.

Neural Networks for particle discovery

H. Khan H, A. Tripathi, M Vyas
Watford Grammar School for Boys

The ATLAS experiment produces gigabytes of data every minute. Currently analysis of some of this data is performed manually by the citizen scientist community. Our research scrutinizes the reliability of the citizen scientists' observations and explores methods to optimise the analysis process through automation. By implementing machine learning we have found that convolutional neural networks are able to accurately analyse these data sets in a fraction of the time taken for their human counterparts.



Introduction

The Higgs boson was theorised by Professor Higgs in 1964 to assign ma**ss** to subatomic particles. In 2012, a particle with the required description was discovered at CERN's Large Hadron Collider, after almost a 50-year hunt. The Higgs Boson has a very short lifetime, but may decay into longer-lived baby Higgs particles.

A preselected sample of 100,000 events from the ATLAS experiment was placed on the Zooniverse platform, allowing non expert citizen scientists to help analyse the vast quantities of data produced. Citizen scientists would then place clicks in areas of the image where they observed interesting phenomena (Figure 3).

Measuring the reliability of clicks and ranking the users

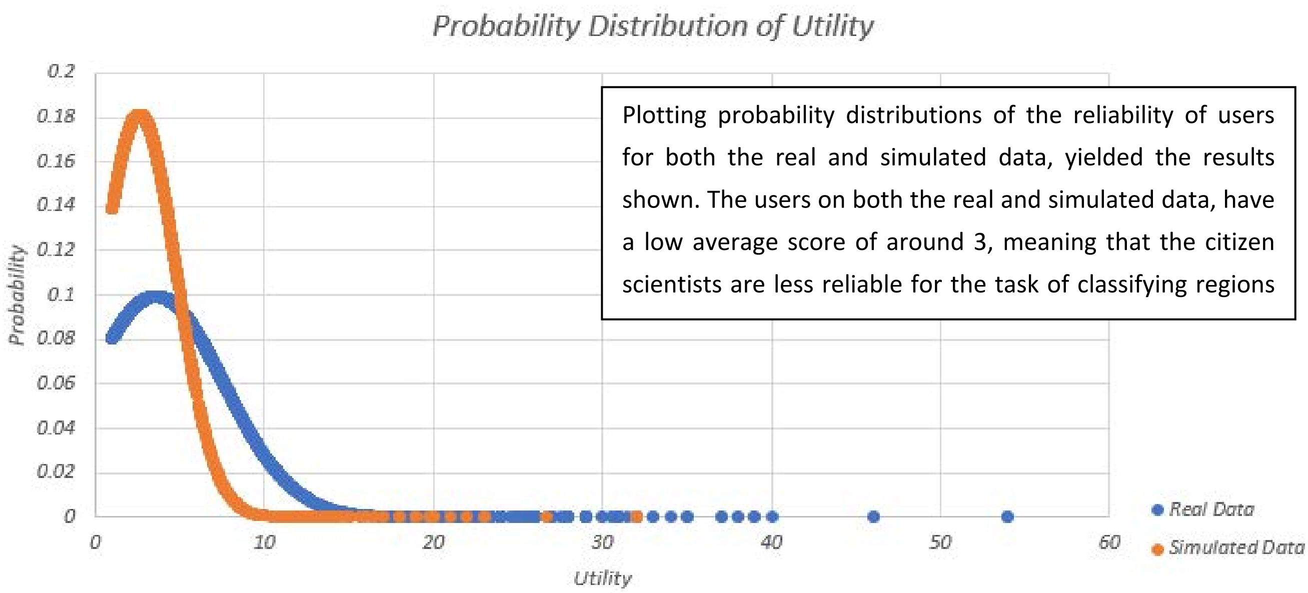
We made the assumption that the location of a click is independent of the location of other clicks made by the same user. We therefore deduced that the location of a click was depending on two factors: the reliability of the user, and the locations of patterns in the image. As the citizen scientists involved were instructed to click near specific patterns, we can define the reliability of

Figure 3 A sample image from the project

the clicks, by their proximity to such patterns. However, the real dataset provided does not include the information of the location of the off centre vertices. Due to our assumption, there is a greater likelihood that clusters of clicks form around patterns.

The algorithm measures reliability of the clicks by using the number of clicks within the same angular region. We found that citizen scientists frequently misplaced clicks along lines in the images, which would falsify the results of a distance-based approach, as lines extending from the centre could be falsely classified as related, while points on the same line would be not.

For every XY image, our algorithm converts every click into polar coordinates, then counts the number of clicks within a region of ±0.1 radians compared to the real coordinates. The number of clicks within this region is equal to the reliability weight of the click. For every RZ image, our algorithm measures the Euclidean distance between clicks and counts the number of clicks within +-0.1 units of the click. In order to rank users, the average reliability of their clicks is used. The users are sorted from the most to the least reliable. This algorithm was run on approximately 67000 images in the datasets provided.



Plotting probability distributions of the reliability of users for both the real and simulated data, yielded the results shown. The users on both the real and simulated data, have a low average score of around 3, meaning that the citizen scientists are less reliable for the task of classifying regions

Creating the artificial citizen scientist

In order to classify events more accurately, it was decided to create an artificial citizen scientist using deep learning techniques. A convolutional neural network model allowed us to classify regions in the images as either lines, displaced vertices, empty space or the centre of the image. The Python programming language was used to implement both algorithm and neural network. We also used the Keras and Tensorflow libraries. Deep learning (a form of machine learning) requires large quantities of input data in order to train the algorithm, while our team only possessed 24 event images. Furthermore the images contained background noise in the form of white circles (Figure 3). The quantity of data was increased by splitting every image into 36 sections, providing 864 smaller images, while the background noise was removed by removing specific colours. These images were then manually labelled to ensure accuracy. The convolutional neural network was trained on 80%

of the dataset and achieved a 93% accuracy when classifying the remaining 20%. Access to more data, will allow for the creation of a system which can place clicks in more appropriate regions.

Conclusion

The process of manually classifying images is not only time consuming but also inaccurate in some cases which leads to inconsistency. To limit the disadvantages mentioned above we have automated this process by implementing machine learning in the form of a convolutional neural network. The Neural Network has been trained to be accurate in 93% of cases hence this process lends itself well to automation.

With access to more data, we will be able to train our system to an even higher degree of accuracy and ability. This would entail the use of autoencoders, a type of neural network that reconstructs 'normal' images, and flags images that it can't reconstruct as 'strange'. Hence allowing the citizen scientists to focus their expertise on the analysis of 'strange' images. The use of autoencoders would also allow for the representation of images in high dimensional space, with similar images being displayed close together.

Acknowledgments

We would first like to thank The Institute for Research in Schools and Oxford University for providing the necessary data required to fulfil our objective. We would also like to thank Merton College Oxford and the Oxford physics department for hosting the two-day conference where we were enabled to present our findings and receive inspiration for further research. Dr Cerezo and Dr Cianci played a vital role in the development this project, providing technical or moral support when needed.