Analyzing Websites with Machine Learning

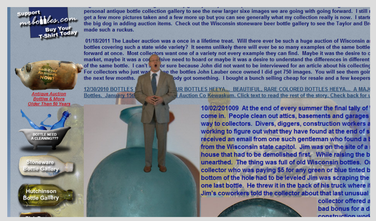
Adnaan Sachidanandan[1], Andrea Angquist[2]
Henry M. Gunn High School[1], Pivotal Labs[2]

INTRODUCTION

Every website has a specific purpose. In order to get more views and attention from audience, website creators modify the website by adding a color scheme, an overall design, different images, fonts, etc. However, many designers do not know exactly how to improve on their website. This study is intended to provide a solution for them. The study involves collecting data on the design components of the top websites in the internet, and using research data to create a machine learning program that uses the data and analyzes the design of websites and predicts the website's appeal to target audience.

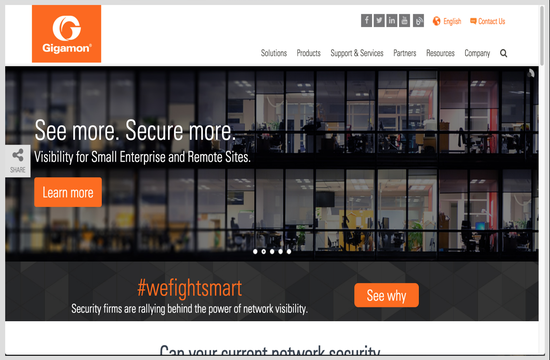
BACKGROUND AND SIGNIFICANCE

- It is intuitive that websites that look better have better reputation and will be visited more on the Internet.
- In addition, studies have found that design elements, such as font, colors, images, etc. have an effect on people's perceptions of websites.
- Neural Networks are models in machine learning that act like a human by learning from data and predicting outcomes based off of them. They are effective models that are capable of solving complex problems accurately and efficiently.
- Many people do not know whether their website will turn out to be good or bad when it is released.
- Neural Networks can be used to predict the future amount of views a website gets, given some training data.
- If a neural network successfully does this, designers can use it to see what is wrong with their websites and modify them, so they never release bad websites into the internet.



Prevent This Unorganized, Random Mixture of Color and Text

Encourage This Organized, Well-Planned Layout with a Color Scheme and Organized Test

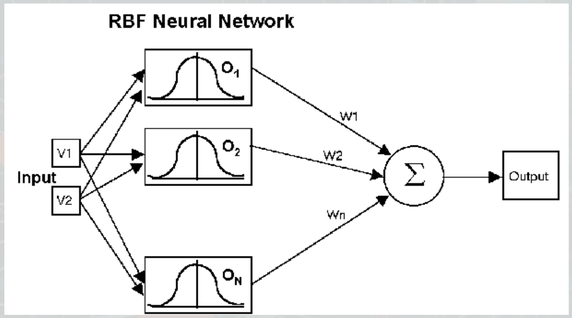


THE PROGRAM

1. The Neural Network

a. The neural network will be a radial basis function network.
   i. Good at handling simple data and using it to learn and change weights in the network to improve its accuracy
   ii. Used commonly iin predictive analytics. What the program is intending to do is predict the future of a website, so this is the perfect type of neural network

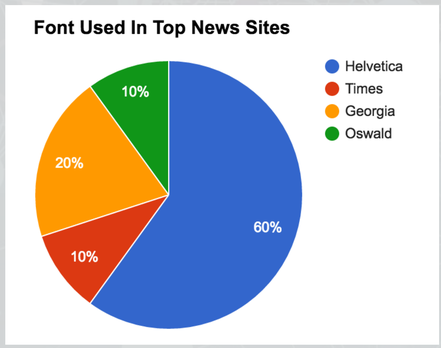
b. The neural network has a simple layout:



Source: DTREG RBF Neural Networks https://www.dtreg.com/solution/view/25

1. The Training Data

b. The training data will be taken from a list of the top most visited websites and their design components.
c. Types of websites will be divided into categories, such as shopping, company, blog, etc.
d. Each category of sites will have its own set of the top 50 most visited sites, and for each site, the font type, font color, and background color will be noted down in the data.
e. Example of one design element of sites in one category:



3. Training the Neural Network

a. The neural network will have a training function that will use the data collected as training data.
b. The neural network will determine the error that its current weights have and modify them to minimize that error.

RESULTS AND CONCLUSION

Results:

The Neural Network did not seem to predict the outcome of a website accurately. When tested with actual website data, the results varied greatly each time the program was run, and the results were highly variable between sites of high and low standards.

```
For CNN
[ 0.39907271]
```

*Program Output for Input Data from CNN

```
For NBC
[ 0.81487016]
```

*Program Output for Input Data from NBC

Conclusion:

In conclusion, a radial basis function network cannot accurately predict the outcome of a website being released into the Internet. Its structure results in little accuracy and highly variable predictions with similar and differing inputs. A radial basis function network is extremely basic and simple, so that could have led to the results seen. As a result, a more complex neural network with a different structure would have likely produced better and more accurate networks.

Expanding:

In the future, I hope to try this same study with a different, more complex neural network that can take much more data and produce accurate results that can truly be used in the real world. In addition, I hope to expand my neural network to be able to provide insight and analysis on various websites.

ACKNOWLEDGEMENTS / REFERENCES

Special thanks to Ms. Merchant and Andrea Angquist for helping make this project possible.

Works Cited:

Cho, Sook-Hyun, and Se-Joon Hong. *BLOG USER SATISFACTION: GENDER DIFFERENCES IN PREFERENCES AND PERCEPTION OF VISUAL DESIGN*. Korea University, 1 Sept. 2013. *Ebsco*, eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?sid=0d4a83e5-866f-4f94-81ed-f023b4ed06e5%40sessionmgr104&vid=0&hid=126. Accessed 25 Sept. 2016.

Hendricks, Drew. "Understanding the Full Impact of Web Design on SEO, Branding, and More." *Forbes*, Forbes.com LLC™ , 11 Feb. 2015, www.forbes.com/sites/ drewhendricks/2015/02/11/ understanding-the-full-impact-of-web-design-on-seo-branding-and-more/#3c8221662027. Accessed 15 Sept. 2016.

Hojjati, Nafiseh, and Balakrishnan Muniandy. *The Effects of Font Type and Spacing of Text for Online Readability and Performance*. Universiti Sains Malaysia, Apr. 2014. *EBSCO eBook Collection*, eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?sid=fc605d4e-5eb8-4b21-9d26-cabab5dfaac6%40sessionmgr103&vid=0&hid=126. Accessed 25 Sept. 2016.

Hsu, Yu-chang. *Better educational website interface design: the implications from gender-specific preferences in graduate students*. Wiley-Blackwell,Mar. 2006. *EBSCO eBook Collection*, eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?vid=14&sid=81722129-0af3-4569-9bf6-332f182d07df%40sessionmgr120&hid=126. Accessed 25 Sept. 2016.

Shah, Nehal. "Cross Cultural Considerations for User Interface Design." *Cross Cultural Considerations for User Interface Design*, Apr. 2013. *Human Factors*, www.humanfactors.com/newsletters/ cross-cultural-considerations-for-user-interface-design.asp. Accessed 25 Sept. 2016.