An Evaluation of Factors Leading to Successful Research

ALEX COLEMAN
Produced in Steffen Guenzel's Fall 2015 ENC 1102

By examining different factors that affect the amount of success that a researcher can achieve through publishing research papers, I showed which factors lead to the most amount of success for researchers. Because of how awards are structured, I was able to demonstrate that at least one paper with a high citation count is needed to become famous as a researcher in most cases. Through this, I demonstrated that the most efficient use of time for a researcher would be devoting all of their efforts to one or two papers and ensuring that the papers were thorough, rather than publishing many trivial papers. Foundational papers showed to further improve the success rate of the researcher since more people will need to use the paper and cite it if it presents a new focus in the area of research. With this, I propose that the current standard for indices (the h-index) to measure researchers needs to be replaced with a new model, the c-index, which emphasizes more citations instead of number of papers.

Artificial Intelligence describes the academic field that studies how to create computer hard- and software capable of intelligent behavior. It has become a more popular research field recently now that researchers are publishing more papers as basis for further research. Each new work is dependent upon the work of other researchers, so this development of foundational papers leads to a lot of growth in the field. Greene (2001) stresses the importance of the context of an academic conversation. He demonstrates the importance of building on something else when creating new academic material. If new material builds on something else, then creating the material to build on will certainly lead to much more research in the field. This will be one of the techniques examined in this paper. Zachary (2015) observes that artificial intelligence has taken off in recent years and is showing exponential growth (para. 5). It is an ideal field for examining the effects of various research techniques.

The goal of this paper is to conclude what methods researchers can use to become more accredited researchers in the field of Artificial Intelligence. I will examine a variety of researchers to examine how effective different methods are at leading to successful research. In his study on predicting success of researchers, Rokach, Kalech, Blank, and Stern (2011) offered a rationale for using Artificial Intelligence (AI). They first suggest that narrowing down to an individual niche, like AI instead of a general community, like Computer Science, will make comparisons easier. Then they add that there is ample bibliographic data for AI publications, committees nominate enough AAAI Fellows to use as validation for the model, and the researchers write many different types of

publications like journals, conferences, books, and chapters (p. 1). The results of this paper should also be applicable to other fields of research, but looking at research through the field of Artificial Intelligence will be simpler.

The three main methods I will examine for successful research are publishing many papers over a brief period of time, devoting a lot of time to publishing a single paper, and publishing a paper on a subject that lays the foundation for new research. There can certainly be overlap between these methods though, so the ideal strategy may not just be one of them. My preliminary thesis is that researching something that is a foundation to new research will show the strongest correlation with high success. In addition, I think focusing on one paper primarily will also show more success. I also predict that an index which scores higher for individual papers of high quality rather than numerous low quality papers will be necessary. I will be using a number of citations and various indices as a measure of success. Collaboration is the most important part of research, because no researcher comes up with new ideas without building on something preexisting. Therefore, the number of times other papers cite a work demonstrates how much further research has come from that particular paper. This is an excellent measure of success since it shows the extent to which people decided to build on their idea using that paper. This will be the main methodology used in the paper. I also look at awards won by authors of papers and examine how that correlates with the number of times they were cited. Rokach et al. discuss the effective ways to predict researchers' success in their paper, so I will connect it with those. They examine the winners of the AAAI Fellowship awards over the past decade. These papers have a certain set of characteristics that distinguish them as successful works. I can apply my theoretical models to them to see what most accurately predicts success. Rokach et al.'s research is primarily the only research so far, so I will have the opportunity to expand or validate the research here. The majority of the evidence behind my research will come from examination of researchers and their published research papers in the field of Artificial Intelligence. Therefore, the results of this paper will only apply to Artificial Intelligence with certainty, but may work for other fields as well. Various researchers will provide the different evidence from which I can draw a reinforced conclusion.

Scientific Background

In the past, the best method of predicting success for a researcher involved looking at their citation counts of past works. Hirsche came out with his h-index in 2005, with Egghe developing his g-index in 2006. It was difficult to quantify which was better though at that point since there was no effective way to tell which score was more accurate. In 2010, Rokach et al. pushed the bubble within machine learning prediction accuracy. His method works by teaching the program how to categorize the papers simply by telling the program what answer it should have reached for some sample data. They sought to improve the prediction accuracy so that people can apply it to more real world applications like promotions. The purpose of Rokach et al.'s paper was simply to improve the effectiveness of predictive models since they are not perfect yet. They showed that their new model improves the precision and recall rates slightly. They revealed that researchers can also apply this new neural networks model to systems other than the AAAI awards. They gave the example of predicting which researchers (or workers) will get promotions. Their purpose therefore was also to expand the ways that these methods can be used. Rokach et al. (2011) offered that evaluating a researcher has many uses such as deciding whether to hire them, promote them, or grant them a competitive award (p. 1). They enumerated some ways outside of just predicting the AAAI awards for which the algorithm can be used. The new methods and applications of predicting fellowship recipients alone provided enough material to make this paper a keystone work. However, Rokach et al. did not conclude which type of index best measures researchers since this was not their primary focus. I will combine their findings to analyze what works best.

Methodology

The main factors I will consider when examining researchers are the frequency with which they publish papers in conjunction with how often the researcher ends up being cited. Ideally, the researcher would have the highest citation counts on each paper and publish them every day, but that is not feasible. Instead, I will examine what reasonable combination of devotion to one paper and diversity leads to the most success by looking at the total number of papers published by that author. To do this, I examine a few different researchers that will represent each of these categories (high frequency, low frequency, and presence of a foundation). I will then calculate the overall 'success' of these researchers. In order to do this, I will look at Rokach et al.'s findings with their model. I will use the results of their model for some papers to see how successful it predicts them to be. These will also go hand-in-hand with the various indices that are used to measure researchers. AAAI Fellows are determined based on a single paper or theory that revolutionized some aspect of the field, so having a high maximum citation count is very important to winning the fellowship. It follows that the maximum citation count (or any index that models this well) is a good indicator of success.

In addition, I will also analyze the results of different methods in the past and add in discussion on whether or not the conclusions of each were reached with valid evidence and without logical fallacies. The most recent published method is Rokach et al.'s technique of using machine learning in combination with citations to predict success, but it is a program and not a simple model. Currently, researchers rely heavily on Hirsch's model of the h-index. Hirsch (2005) proposed that a researcher could be given an index of h if they have published at least h papers with h or more citations (para. 1). This method does not work well for people with only an individual paper since their h-index cannot exceed one**.** Google Scholar offers another model, the i10-index. This model gives a score that is the number of works published with more than 10 citations. While this is a much better indicator for the amount a researcher has done, it still does not represent well a researcher who focuses on one paper. I will actually look at whether these models actually have any grounds to see if they dictate the ideal distribution of effort over papers. If not, I will propose a new model that better rates researchers to come to a new conclusion about the methods for successful research.

Empirical Data and Results

The researchers I am examining are Rokach, Agrawal, and Baughman. Rokach represents the idea of publishing many different papers in different areas of your field. He has published more than 300 papers in the field of Artificial Intelligence with a variety of foci. He published his first paper in 2001 and continued publishing papers for the next 15 years, reaching his peak around 2010 with one paper being cited 1400 times. This suggests that publishing many different papers may eventually lead to success. As Weingart (2012) concluded from looking at the data of many researchers, being cited over 100 times is generally considered far above average, so being cited over 1000 times is certainly a mark of success (para. 5). Rokach has an h-index of 37. This means he has published 37 papers with over 37 citations each. However, he has had a single paper with 1400 citations, which this h-index model does not represent. His i10-index is 95, but again this is not distinguishable from somebody with 95 papers with only 10 citations each.

Agrawal represents publishing a single foundational work (in addition to many other works in his case). Agrawal, Imieliński, and Swami (1993) begin their paper by stating that they will introduce the new problem of mining large amounts of data for correlation with some amount of specified confidence (p. 1). Since this work introduces a new and relevant problem, it became a foundational work. Foundational works, by nature, tend to have the highest citation counts and therefore lead to the most success. His two most popular works were each cited over 15,000 times

due to how large a foundation they built. His main work, "Mining association rules between sets of items in large databases," laid down the groundwork for data mining, a large field of its own now. He has an h-index of 91. While this number is better than Rokach's because of all the lower level papers he published, it would not be if he had only published his most popular paper. The issue with this as a measure of success is that it only takes a single paper with 15,000 citations to become a hugely famous researcher, but that only yields an h-index and i10-index of one.

Baughman serves as a model for the practice of spreading out focus over many papers and only building on others' works without laying the grounds for new research. In their introduction, Baughman, Chuang, Dixon, Benz, and Basilico (2014) state that their paper describes an analysis of DeepQA's trials (para. 1). This means that their paper will only be building on the work of others and not leaving room for anybody to expand. This is likely what led to the low citation count of one for that paper. Baughman has an h-index of only seven, simply because he does not have many papers with more than seven citations, even though he has one paper with 100 citations. His maximum citation count of 103 demonstrates that the practice of publishing lots of low-level papers is not effective. This effectively disproves the previous idea that Rokach may have achieved success eventually simply by publishing many papers (since this did not work for Baughman). Harald Steck is another example of spreading the focus. He only published three papers before 2003 and they all had less than five citations. In the past five years, he published 15 more papers and raised his maximum to 27. This could indicate that there is some merit in publishing many papers, but not definitely. Steck (2010) wrote in his highest citation paper that his new method of selecting data shows dramatic improvements over previous, sophisticated methods (para. 1). This indicates a revolution in this area, since an old sophisticated model can be outperformed by a new simplified technique. Therefore, it is likely that the only reason Steck reached this peak in his citation counts is that he worked on a foundational paper, not that he published many papers. Steck has an h-index of four, which seems to represent accurately his success. However, at low h-indices, the value is less important since distinguishing between researchers will likely have to take place with researchers that are more successful. Having an accurate index for Steck is not relevant, as long as it is low enough to exclude him from award selection.

In contrast to these models, I propose a new index, the c-index, in which a value of c is given to prioritize individual highly cited papers over many lowly cited papers.

The h-index model is not an effective predictor of which researchers will win awards and achieve success. The h-index model does not distinguish someone with 37 papers of exactly 37 citations from someone who published one paper with 37 citations and 36 papers with over a thousand citations. It prioritizes the number of papers since they are more limiting than citations within a paper at higher levels of success. Therefore, it does not handle outliers like a researcher publishing a single paper with 1400 citations and no others over 50. It is more accurate at predicting comparative success of lower-level researchers. Egghe (2006) proposes another much better model for my purposes, the g-index system, that works by counting the top g papers with at least g^2 citations in total (p. 3). He states that there is an issue with the h-index where it does not handle outliers of highly cited papers, and addresses this somewhat well with his g-index (p. 2). He gives Rokach a g-index of around 65 since his top 65 papers sum to about 4200. However, it is also possible that somebody could reach this index with 65 papers of only 65 citations each.

In contrast to these models, I propose a new index, the c-index, in which a value of c is given to prioritize individual highly cited papers over many lowly cited papers. One such index would

give a value of c if the average of the top $\frac{100}{\sqrt{c}}$ (rounded up) works' citations were at least c. This would give Rokach a c-index of over 200. It is impossible for anyone to beat this index without having some papers with at least 200 citations on a single paper. This is much better than the old models where it only takes a large number of low-citation papers to inflate the index. The effect of taking the square root of c is to slow the decrease in number of papers so a reasonable amount of papers can be reached for both low and high citation counts. However, it still takes a decent number of papers to solidify a researcher's g-value at low citation counts. A c-index of four can be reached with just 25 papers with four citations, while a c-index of 100 will require 10 papers averaging over 100 citations. By dividing by a function of c, fewer papers are needed to solidify higher c-indices since they represent more on their own.

Another better alternative would be to square the individual citation counts before summing them. In this case, the c-index could be the sum of the squares of citation counts for a researcher. For example, having three papers each cited 10 times leads to $10^2 + 10^2 + 10^2 = 300$. In comparison, having 30 papers cited once each (leading to the same total of 30 citations) only leads to an index of 30. Having one paper cited 30 times trumps all of these, since $30^2 = 900$. This c-index would give Rokach a value of over 2 million from his top work alone. This would take 100 works with 100 citations each to match, which is a reasonable tradeoff. Furthermore, to reach his c-index using only papers cited 10 times would take 200,000 papers. The idea here is that it is only feasible to match his c-index by also having highly cited papers, or by having a very large amount of other papers. The rationale for this alternative is that many optimizations already used in mathematics (like least squares regression), square the individual residuals before summing them up in order to make one large change have more of an effect than many small changes. However, the actual formula that is used to prioritize individual highly cited papers over many lowly cited papers is not very important.

Regardless of the formula that accounts for it, the greater importance is that some type of shift is necessary to change the focus from high number of papers to high individual citation counts. The reasoning behind this shift is that a higher number of citations on fewer papers, rather than simply a 1:1 ratio of citations to papers, has been shown to lead to more success. For example, AAAI Fellows are determined based on a single paper or theory that revolutionized some aspect of the field. Publishing many low citation papers will not earn a researcher a fellowship award. So in order to give higher scores to those who received fellowships, the proposed c-index works much better.

Summary and Conclusions

I have demonstrated that the most productive approach to research is publishing a few papers and reaching high citation counts on these papers, while simultaneously focusing on foundational concepts. While this is the most efficient use of time, it may be necessary to publish some papers with lower citation counts in order to build credibility first and practice before working on the magnum opus. High citation counts on a few papers results in the most success, because it is much more important to have few works with one that becomes famous than to simply publish hundreds of low-citation papers. Committees select fellowship and award recipients based on the effect of an individual paper. Therefore, devoting time to only a few papers will lead to more success in an equal amount of time as many papers with a short amount of time for each.

Because of this, I also proposed alternatives for a new index for scoring researchers' effect called the c-index. This new index matches my conclusion about what type of research leads to more success by scoring researchers with high citations (and possibly fewer papers) higher than researchers with large numbers of low-citation papers. Currently, Google Scholar only shows the number of citations, the h-index, and the i10-index. The ACM only lists number of citations and average citations. None of these effectively capture the greater effect of a high citation count on an

individual paper, rather than an equal total citation count spread out over papers. While I have not concluded whether the g-index (which does handle the necessity to prioritize individual high-citation papers to some extent) or c-index is better, I have shown that a change is necessary, since top aggregates do not show either. Research aggregates need to include a new type of index, whether it is the c-index or a different one, that follows this model in order to better predict a researcher's success.

References

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (pp. 207-16). New York, NY: ACM.

Baughman, A.K., Chuang, W., Dixon, K.R., Benz, Z., & Basilico, J. (2014). DeepQA Jeopardy! gamification: A machine-learning perspective. *IEEE Transactions on Computational Intelligence and AI in Games,* 6(1), 55-66. Retrieved from http://cis.ieee.org

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131-152. Retrieved from http://link.springer.com/journal/11192

Greene, S. (2001). Argument as conversation: The role of inquiry in writing a researched argument. In W. Bishop & P. Zemliansky (Eds.), *The subject is research: Processes and practices* (pp. 145-64). Portsmouth, NH: Boynton/Cook.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569-16572. Retrieved from http://pnas.org

Rokach, L., Kalech, M., Blank, I., & Stern, R. (2011). Who is going to win the next Association for the Advancement of Artificial Intelligence Fellowship Award? Evaluating researchers by mining bibliographic data. *Journal of The American Society for Information Science & Technology*, *62*(12), 2456-2470. doi:10.1002/asi.21638

Steck, H. (2010). Training and testing of recommender systems on data missing not at random. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 713-22). New York, NY: ACM.

Weingart, S. (2012, September 20). How many citations does a paper have to get before it's significantly above baseline impact for the field? [Web log post]. Retrieved from http://www.scottbot.net/HIAL/index.html@p=22108.html

Zachary, G. (2015, June 17). Let's shape AI before AI shapes us. *IEEE Spectrum*. Retrieved from http://spectrum.ieee.org/robotics/artificial-intelligence/lets-shape-ai-before-ai-shapes-us

Alex Coleman

Alex Coleman is currently a freshman at the University of Central Florida, studying Computer Science. In his free time outside of classes, he competes with UCF's programming team in ACM's International Collegiate Programming Contest, a contest that integrates programming and problem solving. After graduating with his bachelor's degree from UCF, Alex hopes to get a job working at Google or Microsoft doing software development, where he will apply knowledge gained from UCF's programming team in real world situations.