

SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

A Text Analytics Approach to Unraveling Trends in
Government Funding to Small Businesses

Sairam Tanguturi
Sagar Rudrake
Nithish Yeduguri

#SASGF



Text Analysis

Sairam Tanguturi, Sagar Rudrake, Nithish Yeduguri

Oklahoma State University

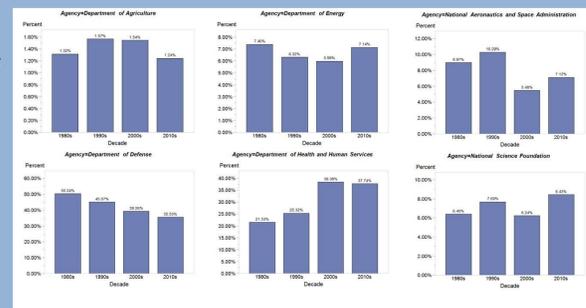
ABSTRACT

Government funding is a highly sought after medium by entrepreneurs and researchers aspiring to make a breakthrough in the market and compete with larger organizations. This Government funding is channelized via federal agencies who seek out promising research and products that better the present framework. The aim of this study is to analyze and study the abstracts which earned Govt. funding in the past three and a half decades in the fields of Defense and Health & Human Services and to understand the factors that are responsible for any awards made to any organization. The top two contenders for government funding are the fields of defense and health, which is why they are the focus of this study.

This study will benefit researchers and entrepreneurs to understand topics that are currently getting more Govt. funding. We will analyze the trends of awards over the past years and through the text mining approach we will find out the technologies that are being researched, the possible new additions to the defense arsenal from the small businesses in the defense sector, and cause, diagnosis and treatment of ailments in the health sector.

METHODOLOGY

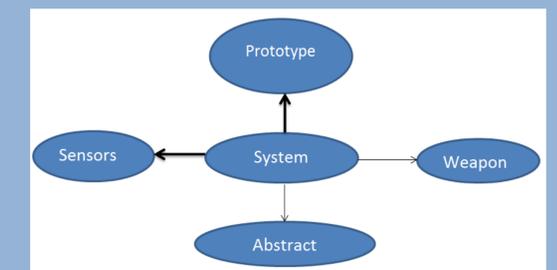
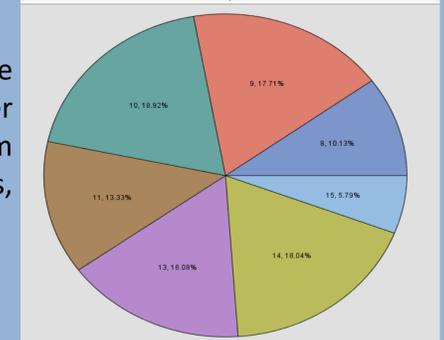
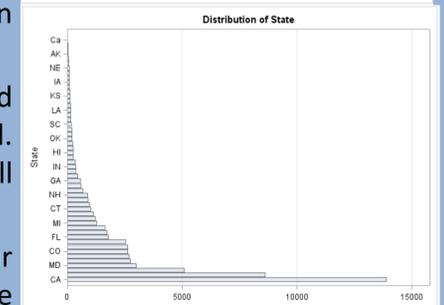
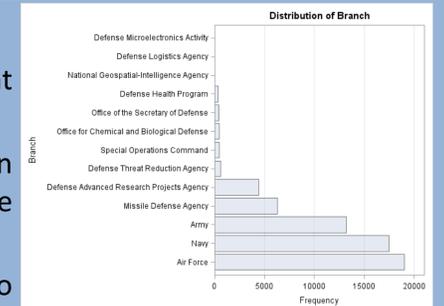
- The Data obtained from SBIR has various data about funding offered for various departments. By using basic descriptive statistics we found that the Department of Defense and the Department of Health and Human Sciences received highest margins of the total funding offered by SBIR.
- We decided to analyze these two fields separately since there was little to no overlapping with the titles of these two fields.
- We also used PROC SQL and Descriptive statistics to understand other trends within the data.
- For text mining we used Text Topic and Text Cluster nodes on both the datasets in SAS Enterprise Miner¹.
- With the Text Parsing node we instructed SAS Miner to ignore terms all the terms other than Nouns and Verbs since these two are the ones that would give us a fair idea about the keywords in the research abstracts that won grants/funding/awards.



RESULTS

DEFENSE DEPARTMENT

- To begin with, we see the distribution of funds within the various branches that come under the Defense Department.
- As expected, major share of the total funds awarded were earned for research in the fields of Air force, Navy and Army respectively. Missile Defense and Defense Advanced Research Projects followed the top three fields.
- Further, we checked which states received the most number of awards in order to see whether location of the business seem to have any impact on the distribution of the funds.
- As shown in the graph on the right, small businesses based in California received the highest number of awards followed by Massachusetts, Virginia and Maryland. The heavy inclination towards California could be because of high number of small businesses present in the state.
- Figure on the right shows a pie chart distribution of the text clusters within our defense dataset. The top clusters contain terms that occur together in the documents within the dataset.
- Cluster number 10, which contains terms like sensor, performance, and test has the highest frequency percentage of 19%. The frequent occurrence of this cluster indicates that the abstracts that won most awards seem to be contain terms from this cluster. This goes to show that the Govt. is encouraging research on sensors, better performance of equipment and performing more testing.
- Concept link diagrams are really helpful in understanding the relationship between various terms that we analyze.
- Our focus is mostly on weapons and technology that interests the defense department, hence we take a look at the concept link diagram on the right.
- The term System has the highest frequency and is strongly related to terms like Prototype and Sensors. Defense agencies seem to be highly interested in these terms as development of new Sensor systems and Weapons Systems would give them the technological edge over their opponents.



Text Analysis

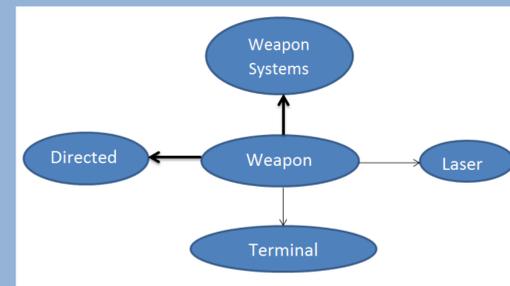
Sairam Tanguturi, Sagar Rudrake, Nithish Yeduguri

Oklahoma State University

RESULTS CONTINUED

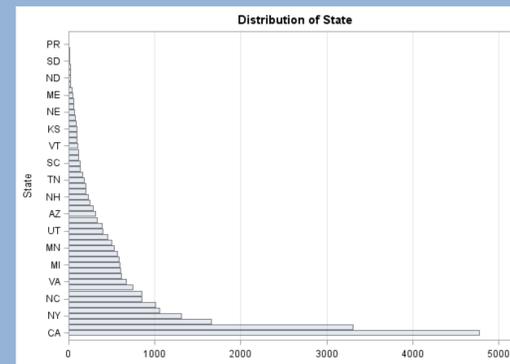
DEFENSE DEPARTMENT (contd...)

- As we are interested in the weapon technology that receives more funding lets take a look at the vector based representation of the term 'Weapons'.
- The terms like Directed, Weapon Systems are strongly correlated to the term 'weapon', while terms like Terminal and Laser have a relatively weak correlation.
- So businesses researching on Directed Weapon Systems have a higher chance of getting a funding than laser weapon technology.

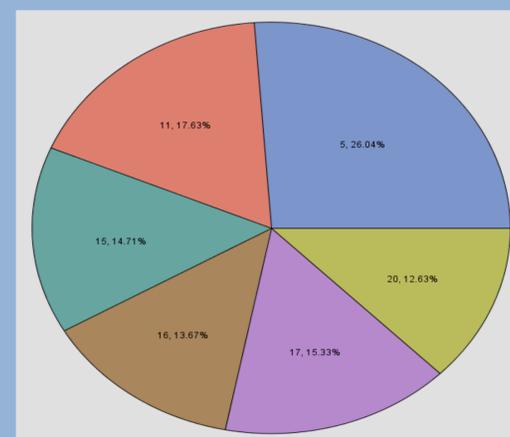


HEALTH AND HUMAN SERVICES DEPARTMENT

- Again we start off with observing the distribution of funds within different states. The distribution is noted in the graph on the right.
- California received the maximum number of awards from the Health and Human Services Department. Again the reasons for this seem to be the same as stated in the defense section.



- The Pie Diagram on the right shows the cluster frequencies of different clusters occurring within the abstract variable of the database.
- The 5th Cluster has the highest frequency of about 26% and it contains terms like 'research data based proposed', 'treatment health', 'therapy'.
- This seems to indicate that most abstract winning awards are focusing on these terms. These do feel like a new trend since, usually it is expected that the field of Health research would focus on new diseases and their research. This data seems to be deviating from the old trend.



CONCLUSIONS

Conclusion:

- The research fields in both the departments of Defense and Health and Human Services seem to be investing in newer research areas.
- Potential Entrepreneurs in the Defense sector will have highest chance of funding if their research is focused on developing new sensor systems which can withstand severe tests and prove high performance.
- Most of the funding seems to be going to abstracts which have a working prototype of their system. So, it is advisable to include a working prototype model in the abstract rather than just providing the idea of it.
- After sensors, most funding goes to development of new laser guided weapon systems. Hence, entrepreneurs should focus their research on developing better laser guidance systems for the Defense department's weapons.
- In the case of health and human services, funding is attracted by therapeutic systems. So, a higher focus on advanced and better performing therapeutics is advisable.

SHORTCOMINGS & FUTURE SCOPE

The data provided by the Government lists only abstracts which were funded. Any information about the rejected abstracts is not available for analysis. Hence, demographic and geographic conclusions might be skewed. Also, providing information about what attracts the Government's attention the *least* is not plausible.

Deeper text analysis of all the sub-divisions of the Defense and Health sectors to pin point the research areas exactly is what we hope to achieve in the near future.

References

1. Sabo, T. (2014). Uncovering Trends in Research Using Text Analytics with Examples from Nanotechnology and Aerospace Engineering, Paper SAS061(-2014), 1-16. Retrieved March 15, 2016, from <http://support.sas.com/resources/papers/proceedings14/SAS061-2014.pdf>
2. Data Source: Small Business Innovation Research(www.sbir.gov)



SAS[®] GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

LAS VEGAS | APRIL 18-21

#SASGF