SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

Sentiment Analysis of Opinions about Self Driving Cars

Nachiket Kawitkar and Swapneel Deshpande

USERS PROGRAM

Sentiment Analysis of Opinions about Self-driving cars

Nachiket Kawitkar and Swapneel Deshpande

MS in Business Analytics, Oklahoma State University

ABSTRACT

- Self-driving cars are no longer a futuristic dream. This technology is said to enable aging or disable people to drive around without being dependent on anyone while also might affecting the accident rate due to human error. We are analyzing people's opinions about self driving cars to generate a summary describing the sentiments of the buyers which will help these skeptical buyer in deciding if they would like to invest in a self driving car.
- The Text Parsing and the Text Filter nodes in SAS Enterprise Miner allows us to analyze the most common as well as rarely used terms with additional features like spell check, synonym grouping and concepts links.
- The Text Cluster enables to group closely related terms into similar areas like technology advancement, robotics, safety aspects of driverless vehicles and so on.
- The Text Rule based node in SAS Enterprise Miner and Statistical models in SAS Sentiment Analysis Studio is used to classify more than 7,000 opinions voiced by people on the Internet about self-driving cars. Smoothed Relative Frequency and No Feature Ranking model was selected as the best model with overall precision of 70.44%.

PROCESS FLOW

Detailed process structure followed for analysis using the nodes in SAS Enterprise are shown in Fig 1.

- The excel data is converted into SAS dataset named SDC_DATA_MODEL where each observation is an opinion. This dataset is then used in the SAS Enterprise Miner project to perform Sentiment Analysis.
- The data is partitioned for training (70%) and validation (30%) purpose using the Data Partition node.
- Text Parsing and Text Filter nodes allows to clean the unstructured text data and use spell check option along with synonym grouping.
- Text Rule Builder node allows to analyze the opinions and test the accuracy of predictions made by the model to assign them as positive or negative opinion.

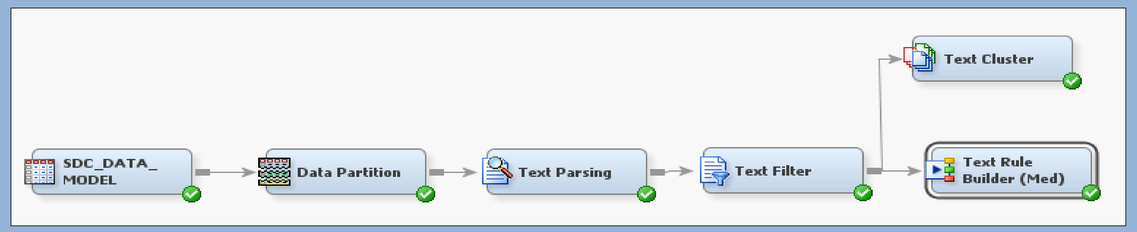


Fig 1. Process Flow

TEXT PARSING AND TEXT FILTER

Terms

Term	Role	Freq	# Docs	Keep
google		534	499	Y
ãâ		573	487	Y
driverless		441	427	Y
ì	Miscellaneous Proper Noun	518	379	Y
+ driverless car	Noun Group	382	375	Y
ãâã		362	353	Y
+ self-driving car	Noun Group	248	246	Y
google car	Miscellaneous Proper Noun	239	237	Y
+ google car	Noun Group	231	228	Y
ã		257	212	Y
+ drive		201	198	Y
future		149	146	Y
+ want		136	135	Y
ù	Miscellaneous Proper Noun	153	130	Y

Fig 2. Text Parsing

Terms

Term	Role	Status	Freq	Weight
+ be		Drop	966	0.000
s		Drop	720	0.000
+ drive car	Noun Group	Keep	625	0.092
ãâ		Drop	573	0.000
not		Drop	412	0.000
ì	Miscellaneous Proper Noun	Drop	518	0.000
+ see		Drop	382	0.000
ãâã		Drop	362	0.000
+ just		Drop	348	0.000
+ have		Drop	327	0.000
+ do		Drop	271	0.000
ã		Drop	257	0.000
+ drive		Keep	204	0.177

Fig 3. Text Filtering

Text Parsing

- Text Parsing node is used to clean the unstructured data by changing the node properties like *standard* for 'Find Entities', 'detect different parts of speech' is set to *No* and ignoring *abbr, prop* and *num* the parts of speech.
- The frequency document matrix generated allows to understand the terms kept and dropped by the node and also tells corresponding frequency and number of documents in which it appears.

Text Filter

- Text Filter node provides the functionality to eliminate the less frequent and irrelevant terms using the interactive filter option in the properties panel.
- By enabling the Spell Check option in Text Filter properties panel, the misspelt word are corrected as shown below in Fig 4 'provlem' to 'problem', 'bwest' to 'best', 'automous' to 'autonomous' and so on.
- We can also group the terms with same meanings manually by using the interactive filter. As shown in Fig 5 the terms car, automobile, van, vehicle, etc. are grouped together and represented as term 'car'.

EMWS1.TextFilter_spellDS

Parent # Docs	Term	# Docs	Parent
6.0	changeist	1.0	changes
24.0	provlem	1.0	problem
4.0	compite	1.0	compete
24.0	transportnation	1.0	transportation
15.0	bwest	1.0	best
79.0	automous	1.0	autonomous
10.0	crzy	1.0	crazy

Fig 4. Spell Check in Text Filter

TERM	FREQ	# DOCS	KEEP	WEIGHT
car	3748	3371	☑	0.015
automobile	5	4		
van	4	4		
automobiles	4	4		
vehicles	65	63		
cars	1588	1489		
vans	2	2		

Fig 5. Synonym Grouping in Text Filter

Sentiment Analysis of Opinions about Self-driving cars

Nachiket Kawitkar and Swapneel Deshpande

MS in Business Analytics, Oklahoma State University

TEXT CLUSTER

- Text Cluster node shows the grouping of terms belonging to a certain topic.
- Using the Expectation-Maximization Cluster Algorithm, 7 clusters are obtained having well distributed frequencies except for cluster 4.
- Cluster table describes different clusters as shown in Fig 6. Cluster 7 shows the excitement in people while Cluster 6 talks about technology changes and Cluster 4 describes the happiness of work done on this innovation.

Cluster ID	Descriptive Terms		Frequency	Percentage
1	+drive +want +wait +wheel +cool +'steer wheel' +steer	...	549	14%
2	+driving +'drive car' +spot +google +car +future +uber	...	459	12%
3	+driver +day +great +hope +look +self-driving +road	...	791	20%
4	+love +big +innovation +fun +live +design +work	...	41	1%
5	+accident awesome +life +idea +mile +human traffic	...	377	10%
6	+car +'self-driving car' +self-driving +technology +verge +google +test	...	1056	27%
7	+people +good +robot +amaze +excite +driving +'drive car'	...	610	16%

Fig 6. Text Clusters

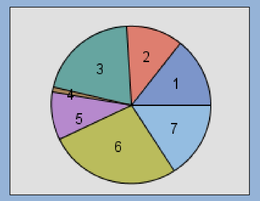


Fig 7. Cluster Frequencies

CONCEPT LINKS

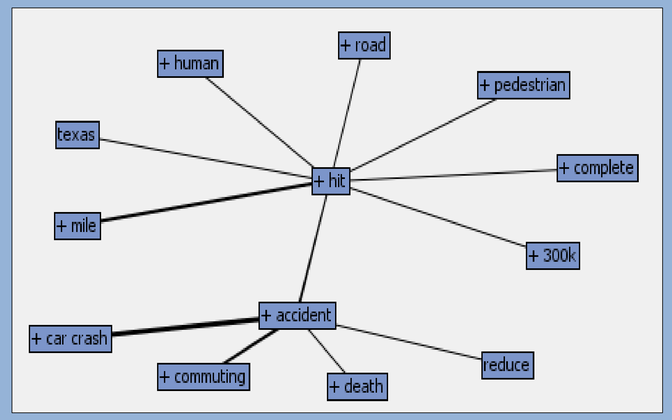
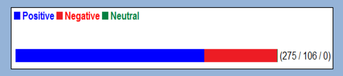


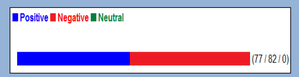
Fig 8. Concept Link for term 'hit'

- Concept Links provides a overview of association of the term at the center with the other terms in the document.
- The strength of association between the linked terms is shown by the width of the link.
- In the Fig 8, the term 'hit' is associated with human, road, pedestrian, accident, etc.
- On further exploring the term 'accident', it is discovered that the terms like car crash and death are also associated with the parent term.
- All these terms are closely associated with driving hazards caused during car driving.

SENTIMENT ANALYSIS

- SAS Sentiment Analysis Studio gives a quick overview of classification of the opinions into positive and negative.
- Keeping 20% of the data aside, Statistical model is built which uses 80% of the remaining data for training and 20% for validation purpose.
- With 70.44% overall precision, Smoothed Relative Frequency and No Feature Ranking model is chosen as the best model.
- The 20% of the data for testing purpose which produces the below results for positive and negative opinions respectively.





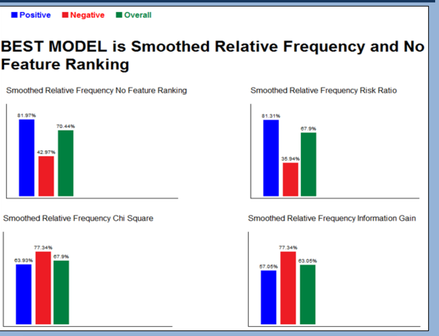


Fig 9. Graphical results of Statistical Model

CONCLUSIONS

- Using SAS® Sentiment Analysis Studio, the reviews of any text online can be quickly classified into a positive or negative sentiment. A quick summary can be generated which reflects the sentiments of the person writing this opinion. Such analysis can be extremely helpful to the audience that depends on others opinions before they make any purchase, especially incase of newer technologies like self-driving cars as they are a considerable investment.
- Online Opinions give insights into the peoples expectations from this newer car technology. This information can be leveraged by the auto makers to include different functionalities in their products and shape their marketing campaigns to cater to the needs and expectations of their customers
- Depending on how often they are utilized together, a relationship can be defined in-between terms using concept links. For example the term like "accident", "car crash" "Commuting" are strongly associated with "car". This indicates the fear about safety of these newer and technologically advanced self-driving cars

REFERENCES

- Guizzo, Erico. "How google's self-driving car works." *IEEE Spectrum Online, October* 18 (2011).
- Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS® by Goutam Chakraborty, Murali Pagolu, Satish Garla
- SAS Institute Inc 2014. Getting Started with SAS® Text Miner 13.2. Cary, NC: SAS Institute Inc.

Acknowledgement : We wish to express our sincere thanks to Dr. Goutam Chakraborty for his valuable guidance

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL