Consumer Complaints and Traffic Fatalities: Insights from the NHTSA Vehicle Owner's Complaint Database

Mahtab Ghazizadeh, John D. Lee
Department of Industrial and Systems Engineering
University of Wisconsin-Madison, WI, USA

Driving simulators, crash databases, and more recently, naturalistic studies all help understand how changes to vehicle design affect driving safety. The rapid computerization of cars makes it increasingly important to capitalize on these sources and exploit others. The present study explores a rarely analyzed data source on traffic fatalities: National Highway Traffic Safety Administration's vehicle owner's complaint database. The textual data within the event description field of each complaint is extracted and analyzed using a text mining approach that involves the use of latent semantic analysis (LSA) for reducing the dimensionality of the problem. Hierarchical clustering is then employed to identify clusters of complaints that share content. Clusters are described in terms of the most frequent terms and the time trends of the complaints within them. The analysis highlights how text mining analysis can help unlock the wealth of information contained in consumer complaint databases.

INTRODUCTION

Recent advances in vehicle technology, ranging from navigation systems and internet connectivity to collision warning and adaptive cruise control, promise to change the driving experience considerably in the coming years. Many of these changes will extend the impressive contributions that technology has made to automotive safety, but the full effect of these changes is uncertain. Driving simulators, naturalistic data, and crash databases all contribute to a better understanding of how drivers respond to changing vehicle technology, but other sources are needed for a more complete understanding. Systematic analysis of open-ended descriptions, such as those contained in customer surveys and complaint databases, could provide a valuable resource to complement other sources (Lehto, Park, Park, & Lehto, 2007).

The National Highway Traffic Safety Administration's (NHTSA) vehicle owner's complaint database encompasses nearly 900,000 records of component failures (as of October 7, 2011), based on Vehicle Owner's Questionnaire (VOQ) complaint entries since January 1, 1995. Among the failures, around 3,000 involve deaths and more than 47,000 involve injuries. The complaint database has rarely been formally analyzed to uncover general trends. However, a few investigations have used subsets of the data, most recently in the case of Toyota's unintended acceleration problem (Kane, Liberman, DiViesti, & Click, 2010).

The goal of this research is to extract information about vehicle component failures from the vehicle owner's complaint database. No a priori hypothesis was made regarding the outcomes and the analysis was conducted in a fully explorative fashion. However, because particular technological changes have taken place over the past decades, changes in the frequency of complaints related to these changes were expected.

The focus of this study is on a subset of complaints: those involving deaths. Fatal cases signify the most dire consequences of component failures and as such, deserve special attention. They also point to the most important areas of safety enhancement and can be highly informative for the safety administration authorities. Therefore, this study can serve as a first step toward the goal of extracting meaningful patterns of component failures from the rich volume of data contained in the vehicle owner's complaint database.

METHODS

Vehicle Owner's Complaint Database

The vehicle owner's complaint database contains more than 883,000 component failure cases, reported since 1995 (NHTSA, 2011). Vehicle owners or their representatives (the latter mainly the case with those killed in the incident) filed these complaints using one of several channels (e.g., NHTSA's website, hotline VOQ, direct letters, and letters forwarded from a congressional office). There is a field named 'Description of the complaint' that describes the component failure and its consequences. The field named 'Specific component's description' describes the failed component (e.g., Suspension: Front). This field is not a part of the complaint entry form and is not filled in by the person filing the complaint, but is completed by analysts at the Office of Defects Investigation. The database is designed such that each record corresponds to a component failure and because more than one component can fail in a single incident, more than one data record can exist for one complaint. In this study, the fatal failures were analyzed, encompassing 2,740 components from 1,910 complaints. Table 1 summarizes the most frequently failed components leading to deaths.

The five components in Table 1 collectively account for more than one third (33.87%) of the fatal cases. It should be noted that the coding of the 'Specific component's description' field is such that component descriptions are not at the same level of detail, e.g., 'Tires' and 'Tires: Tread/belt' are both possible values for this field. As such, the tire-related failures, for example, exceed the 9.45% shown for 'Tires' in the table and encompass a larger proportion of cases. The same is true for air bags and others.

Copyright 2012 by Human Factors and Ergonomics Society, Inc. All rights reserved. DOI 10.1177/1071181312561475

Table 1. The most frequently failed components in the fatal cases*

Component	Number or cases	% of cases
Tires	259	9.45
Air bags: Frontal	230	8.39
Tires: Tread/belt	215	7.85
Vehicle speed control	139	5.07
Air bags	85	3.10

* Based on the 'Specific component's description' field

The complaint database also includes information regarding the vehicle (e.g., manufacturer's name, vehicle make and model, and model year), incident (e.g., involvement of fire or crash), number of injuries and/or deaths, fuel type, and other descriptors of the failure. Our analysis focuses exclusively on the open-ended description provided by drivers in the field 'Description of the complaint'. The analysis is performed in an exploratory fashion with no a priori hypotheses; however, the extent to which the results replicate known component failure trends, the analysis would be considered successful in tracking emerging vehicle problems.

Data Preprocessing

The raw data from the 'Description of the complaint' field was prepared for the text mining analysis with a number of preprocessing steps using the tm (text mining; Feinerer, Hornik, & Meyer, 2008) and Snowball packages (Hornik, 2009) in R 2.14.1 (R Development Core Team, 2011). The corpus structure is the data structure for managing documents in the tm package (Feinerer, 2011). As such, a corpus was first built based on the text of each complaint. Then every letter was made lowercase and punctuation was removed. In the next step, stop words, i.e., frequently occurring unimportant words, were removed from the corpus. Two types of stop words were considered: generic (e.g., 'the') and custom (e.g., 'vehicle'). Then, a stemming procedure was run on the data to retrieve word radicals, i.e., the suffix-erased and more general form of words. For example, 'accelerator', 'accelerate', and 'acceleration' were all reduced to their radical, 'acceler'.

In the next step, a term-document matrix was built using the preprocessed text corpus. This matrix provides a complete mapping of the terms found in the corpus to the individual complaints, in terms of the frequency of each term in each complaint (Feinerer, 2011). Sparse terms were removed from at the 95% maximal allowed sparsity cutoff. This means that those terms that were observed in fewer than 5% of documents were eliminated. The choice of sparsity cutoff is somewhat arbitrary and one can choose the sparsity cutoff that gives a manageable number of terms.

Analysis

A text mining technique incorporating latent semantic analysis (LSA; Dumais, 2004) was used to find the dominant trends in the 'Description of the complaint' field. LSA is a technique used for reducing the dimensionality of a term-document matrix by using a transformed terms vector. The main feature of LSA is dimensionality reduction, where a reduced-rank singular value decomposition (SVD) is performed on the transformed term-document matrix such that only the *k* largest singular values are retained. The reduced-dimension SVD representation is the best *k*-dimensional approximation to the original matrix, in terms of least-squares. The distances (dissimilarities) between documents (i.e., complaints) were then calculated in this reduced-dimensional space. A hierarchical clustering technique using cosine distance was used to group complaints into clusters based on their dissimilarities (Tan, 2011). The cluster dendrogram (i.e., the tree representation of clusters resulting from hierarchical clustering) had all the cases as one group on the highest level. These cases were further split into smaller clusters, as described below. The analysis was conducted using the tm and lsa packages in R (Feinerer, et al., 2008; Wild, 2011).

RESULTS

The term-document matrix (after removing sparse terms) included 50 terms. The most frequent terms were: 'seat', 'tire', 'passeng' (stemmed form of the word 'passenger'), 'control', and 'crash'. It should be noted that because the corpus of terms was stemmed, terms such as 'passeng', 'acceler', and 'injur' were observed in the term-document matrix. Hierarchical clustering identifies clusters at several levels. The choice of stopping criterion (i.e., the number of clusters) depends on the goal and scope of the modeling effort. Here, an 8-cluster solution provides a concise yet informative picture of the complaints data: at the highest level (node 1), there is only one node encompassing all the 1,910 complaints. The subsequent 7 cuts (at nodes 1-7) create hierarchical clusters. On the lowest level of the tree, there are 8 nodes (nodes 8-15). Figure 1 illustrates the dendrogram of clusters and Table 2 describes each of the nodes in Figure 1 in terms of its three most frequent terms and their frequencies.

Table 2. Hierarchical clusters description

Node ID	# of cases in node	Most frequent terms[1]	Frequency of the most frequent terms
1	1910	seat, tire, passeng	661, 572, 509
2	1684	seat, passing, contact	636, 455, 375
3	1381	brake, contact, deploy	366, 322, 315
4	1202	brake, control, contact/crash[2]	345, 285, 264
5	1128	control, crash, contact	281, 259, 257
6	791	fire, truck, passeng	238, 218, 198
7	337	acceler, contact, crash	194, 185, 181
8	226	tire, rear, tread	492, 130, 118
9	303	seat, belt, child	584, 303, 161
10	179	air, bag, deploy	232, 223, 184
11	74	brake, stop, pedal	159, 25, 23
12	100	control, roll, speed	86, 45, 24
13	237	contact, crash, acceler	183, 175, 173
14	137	recal, engin, ford	108, 60, 59
15	654	fire, truck, passeng	229, 216, 195

Notes: 1. The three most frequent terms in descending order of frequency.
2. Contact and crash are each repeated 264 times.

The wordcloud package (Fellows, 2012) was used to visualize the contents of each node. The combination of hierarchical clustering and word clouds in Figure 1 provides a

visual representation of cluster content. At the highest level of the tree, all 1,910 complaints are represented by node 1. The most frequent terms in the whole corpus are ‘seat’, ‘tire’, and ‘passeng’, occurring 661, 572, and 509 times, respectively. At the second level, the complaint cases are divided into two clusters: node 2 consists of 1684 complaints with the most frequent terms being ‘seat’, ‘passeng’, and ‘contact’, repeated 636, 455, and 375 times, respectively and node 8 consists of 226 complaints with the terms ‘tire’, ‘rear’, and ‘tread’ being most frequent, repeated 492, 130, and 118 times, respectively. The subsequent cuts divide the cases further, so that at lowest level, nodes 8-15 represent 8 distinct clusters in the data, each having one of the terms ‘tire’, ‘seat’, ‘air’, ‘brake’, ‘control’, ‘contact’, ‘recal’, and ‘fire’ as their most frequent term.

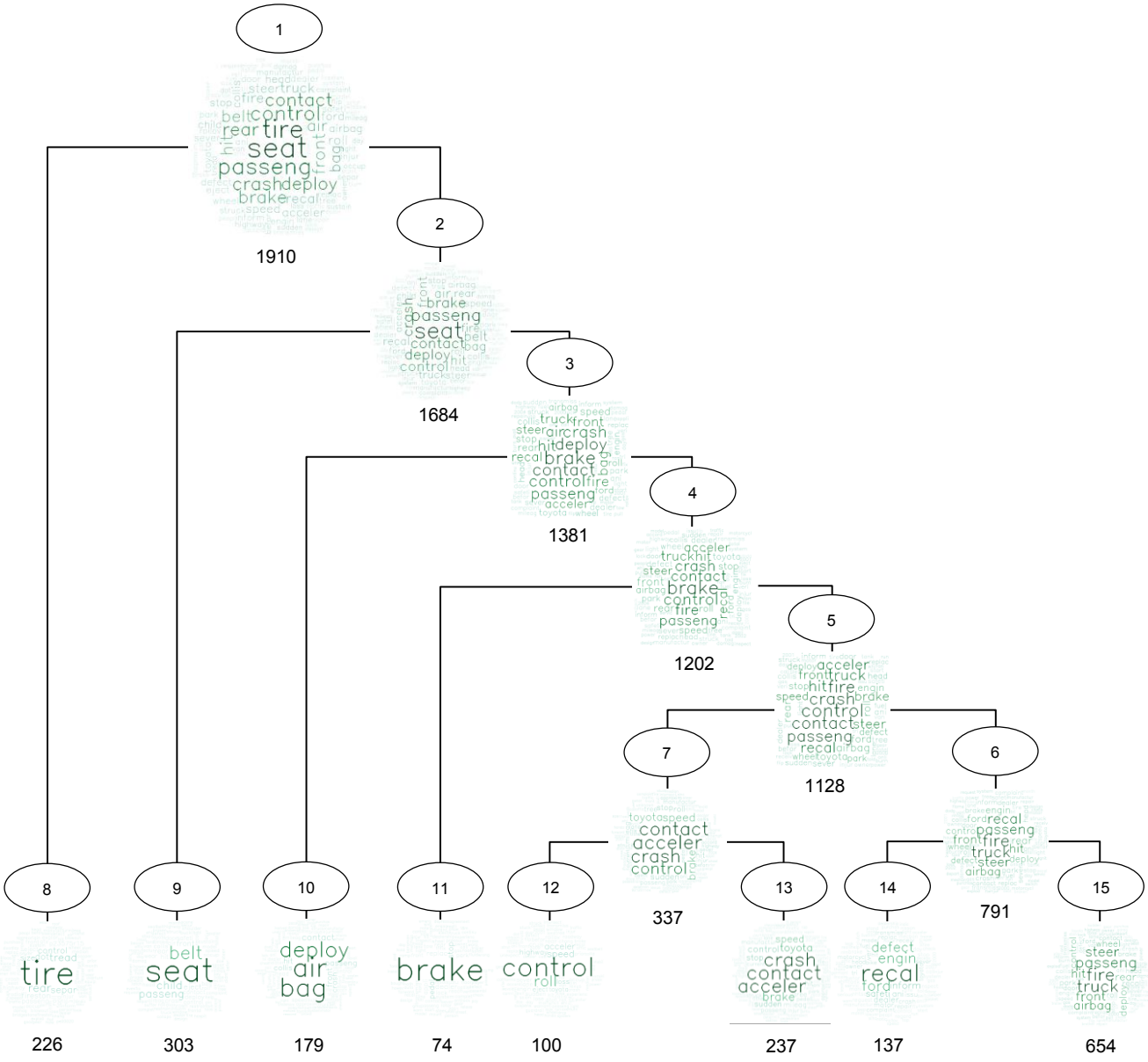

Figure 1. Dendrogram of the fatal complaint clusters
Note: Node IDs are shown in ovals above nodes. The number of cases in each node is shown below the node.

Figure 2 shows the time trend of the number of complaints within each of the 8 final clusters (nodes 8-15) for years 1995 to 2010. The temporal patterns of complaints differ across clusters—the Tire cluster shows a large spike around the year 2000, the Contact cluster has its peak in year 2009, and the Fire cluster is consistently high through 1998-2005. The Seat, Air bag, and Brake clusters show a decreasing trend over the years, although with some fluctuations. The Control cluster, on the other hand, shows intermittent ups and downs, with no apparent overall decrease or increase. The Recall cluster shows a steady increase from 1995 to 2004, but then drops in 2005. This diversity in time trends suggests that the complaints within different clusters, not only differ in terms of content, but also have different temporal patterns, adding another dimension to the comparison of the clusters.

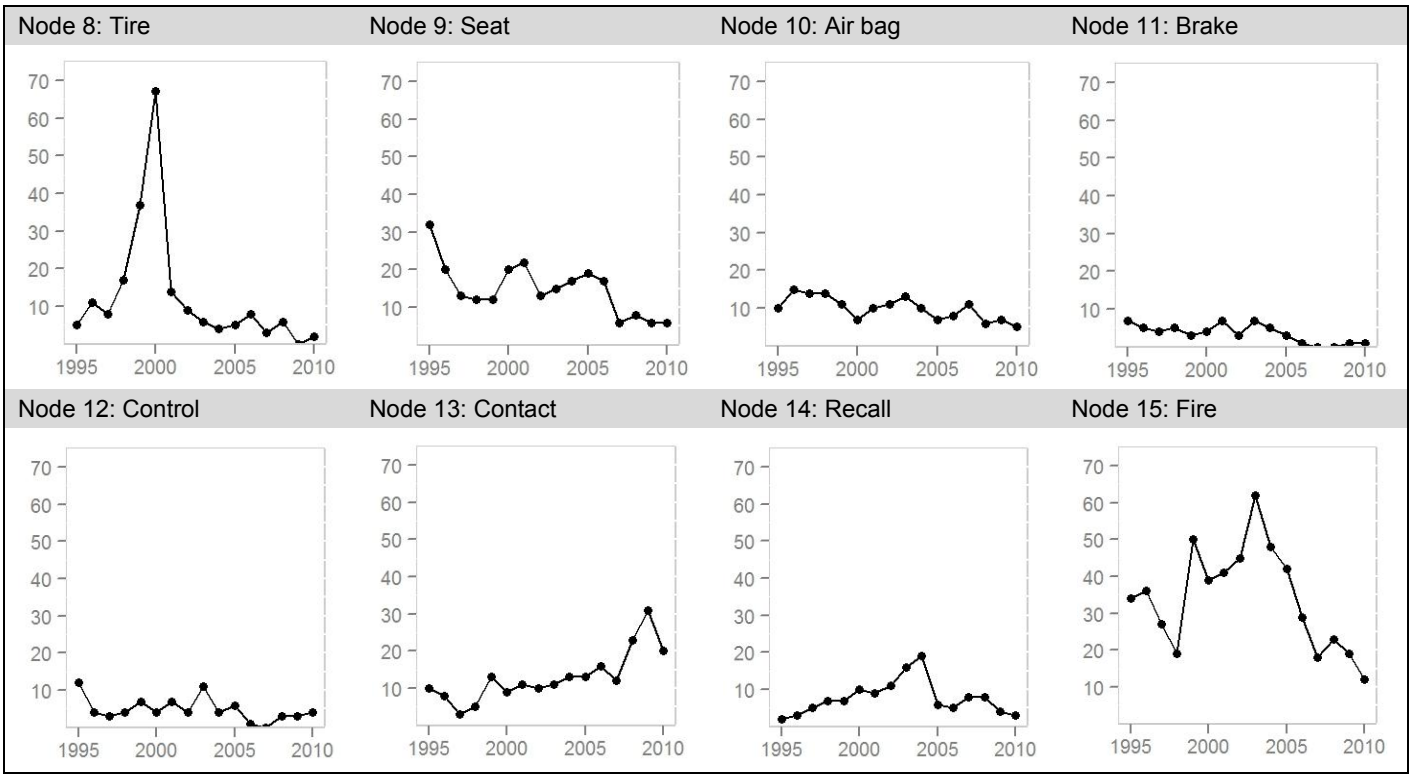

Figure 2. The number of complaints within each cluster as a function of the component failure year

Table 1 summarizes of the most frequent component failures based on the 'Specific component's description' field. These components all belonged to three groups: tires, air bags, and the vehicle speed control system. When referring to Figure 1, node 8 ('tire', 'rear', 'tread') branches at the very top of the dendrogram. The 226 complaints within this cluster constitute the cohort of tire-related complaints and deserve a careful study to gain insight into the deadly tire failure scenarios. Air bag (nodes 10) and Control (node 12) both show up in the clusters at the lowest level, showcasing their significance in fatal component failures. However, the same pattern of consistency is not evident between the most frequent components in the 'Specific component's description' field and all those highlighted in this study. For example, the term 'seat' is the most frequent term in the preprocessed corpus (repeated 661 times) and plays a prominent role in clustering, but interestingly 'seat' (or seat belt) is not among the most frequently cited components under 'Specific component's description'. This observation points to the potential divergence of 'Specific component's description' from the underlying trends in the relatively lengthy accounts of the incidents in the 'Description of the complaint' field that it attempts to summarize.

Investigating the complaints within each cluster over time can offer more insight into the nature of fatal component failures. For example, a review of the Tire-cluster complaints in years 2000 (67 complaints), 2005 (5 complaints), and 2010 (2 complaints) shows that the complaints have been mainly about tread separation and tire blowout throughout the years, and that this problem has been considerably less frequent in recent years. The peak frequency occurs in 2000, which corresponds to the year in which NHTSA contacted Ford and Firestone about the high incidence of Firestone tire failures on Ford products, resulting in a recall (NHTSA, 2000). Another example is the Contact cluster which includes many cases of unintended acceleration and shows an upward trend in recent years which peaked in 2009. These similarities are important in that they demonstrate the success of the analytical method in describing the known trends of component failures in the past decade.

An important consideration when analyzing traffic crashes is the context in which the incidents occur. For example, consider the following entry as 'Description of the complaint' field for a complaint involving one death:

"… was driving 55-70 mph on state [highway] 72 when steering made a full circle then a clicking noise; the passenger said that they then realized that steering wheel had locked up following loss of control. The truck then flipped over back end over front then rolled over on side, [losing] both left wheels, brakes/drums/pads and tires. The driver was thrown from truck after the 2nd flip forward; he was revived on scene [but] then later died in flight to hospital. Passenger survived with minor upper body injuries."

All that can be found in the 'Specific component's description' field for this complaint is 'Steering: column locking: anti-theft device'. However, the description above additionally provides information regarding vehicle's speed, type of road, sequence of events, type and extent of damage to the component, etc. This information can be used in investigating the circumstances under which certain components are prone to fail and how best to guide safety initiatives relevant to product design and manufacturing. As

such, the rich textual description of failure is a valuable source of information that should not be ignored.

DISCUSSION AND CONCLUSION

An exploratory analysis of the complaint database, with the goal of extracting patterns of component failures, identified eight clusters using a text mining approach. The analysis focused on a small, but important, subset of complaint cases: those involving at least one death. Distinct clusters were described in terms of the most frequent term, as well as temporal patterns. The analysis identified a cluster of complaints associated with tread separation problems that peaked in 2000. It also identified a cluster of events that has been increasing over the last several years, associated with unintended acceleration and loss of control.

Further analyses that use the same procedures on larger portions of the dataset can facilitate comparisons between the contents of complaints that have no mention of severe consequences and those leading to deaths (or injuries). This larger dataset might have greater sensitivity to emerging problems noted by drivers. More generally, such text analysis could be applied to other consumer complaint databases or even twitter feeds (Golder & Macy, 2011). Such data sources might indicate emerging problems months or years before they are revealed in crash or naturalistic driving data.

The proportion of failed components in 'Specific component's description' is different in the subset of cases associated with fatal crashes and the full complaint database. In the full database (not analyzed here), 'Power train: Automatic transmission' is the most frequently reported component (5.33% of cases) and 'Service brakes, hydraulic: Antilock' and 'Engine and engine cooling: Engine' follow, accounting for 3.77 and 3.56%, respectively. 'Tires' appears as the fourth frequent component (3.38% of cases). This inconsistency shows that at least based on 'Specific component's description', there is little to no overlap between major failures associated with deaths and those that are not. Although it is unsurprising that certain types of failures can lead to severe consequences whereas others might just disturb a smooth travel, it would be worthwhile to conduct separate analyses based on different subsets of the data and compare the findings.

Another factor that warrants consideration is the cause-effect relationships between incidents and component failures. More specifically, in some cases the failed component is responsible for the incident, whereas in others, the crash occurs and the failed component only contributes to the outcome severity. An example of the former relationship is a brake failure that causes the vehicle to collide with the vehicle ahead, whereas an example of the latter is when a vehicle is in a crash but the air bag fails to deploy. Therefore, although the failed component is potentially responsible for the causalities in both cases, the dependencies between the components and events are not all alike.

Many of the decisions throughout the analysis were based on heuristics. For example, the choice of stop words was partially driven by finding extremely frequent terms that are not informative (e.g., 'vehicle'). The choice of stop words can influence the outcomes, both in the preprocessing phase and in clustering. For example, the term 'passeng' which is the stemmed form of the word 'passenger' occurs frequently in cluster 15. Eliminating this term might have changed the composition of clusters. Decisions such as the stopping criterion in clustering and the sparsity cutoff introduce additional degrees of freedom. As such, even though the modeling framework used here is a quantitative one, major qualitative considerations enter into the analysis. An explicit consideration of the philosophy and techniques of qualitative data analysis and the mixed methods approach could extract more insights from the data than either a pure quantitative or qualitative approach (Johnson & Onwuegbuzie, 2004).

Future studies can augment this work along several dimensions. One promising direction would be to consider information contained in other fields of the database (e.g., vehicle type) to find potentially meaningful associations with clusters. For example, specific car makes or models might be more frequently found in certain clusters. Additionally, it would be informative to know the severity of consequences in each cluster, in terms of the number of people injured and killed. Another direction would be to make comparisons between the complaint failure patterns in different subsets of the data, e.g., those resulting in deaths, those resulting in injuries but no death, and those with no severe consequence. Finally, the association between time trends observed in the number of complaints in each cluster could signal emerging problems associated with new automotive technology.

REFERENCES

Dumais, S. T. (2004). Latent semantic analysis. *Annual review of information science and technology*, *38*(1), 188-230.

Feinerer, I. (2011). Introduction to the tm Package Text Mining in R, from http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, *25*(5), 1-54.

Fellows, I. (2012). wordcloud: Word Clouds. R package version 2.0, from http://CRAN.R-project.org/package=wordcloud

Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*(6051), 1878-1881.

Hornik, K. (2009). Snowball: Snowball Stemmers. R package version 0.0-7, from http://CRAN.R-project.org/package=Snowball

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, *33*(7), 14.

Kane, S., Liberman, E., DiViesti, T., & Click, F. (2010). Toyota Sudden Unintended Acceleration. Rehoboth, MA: Safety Research & Strategies, Inc.

Lehto, X., Park, J., Park, O., & Lehto, M. (2007). Text analysis of consumer reviews: the case of virtual travel firms. *Human Interface and the Management of Information. Methods, Techniques and Tools in Information Design*, 490-499.

NHTSA. (2000). Firestone Tire Recall. Retrieved March 12, 2012, from http://www.nhtsa.gov/PR/FirestoneRecall

NHTSA. (2011). Vehicle Owner's Complaint Database. Retrieved October 7, 2011, from http://www-odi.nhtsa.dot.gov/downloads/

R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Tan, A. (2011). *Text mining: The state of the art and the challenges.* Paper presented at the PAKDD 1999 Workshop on (2011).

Wild, F. (2011). lsa: Latent Semantic Analysis. R package version 0.63-3, from http://CRAN.R-project.org/package=lsa