A Study of Applying Data Mining to Discover the Influencing Factors of Laminated Glass Strength

Hsing-Hung Lin, K. T. Fan, and D. T. Chang

***Abstract*—Being one of the mainstream materials, laminated safety glass has been widely applied in construction industry. However, the bonding part between the laminated material and glass is often ineffective and appears unexpected delamination situation that affect safety greatly. Based on the database of shipment and customer service system in the case company, the object of this paper is analysis the relevant data to find out the factors that affect the bonding strength of the laminated glass. After improved the factor detecting by this study, the edge delaminated ratio is increased from 10.99% to 0.9%, which greatly reduces the delamination circumstance of the laminated glass after installation on the building due to poor adhesion.**

***Index Terms*—Laminated glass, delamination, adhesive strength.**

I. Introduction

With the rise of safety awareness, safety in all industries, not only personal safety but also product safety, have received considerable attention. Glass in building materials has also evolved from an early general glass into a tempered glass because of safety requirements and then evolved into a laminated safety glass. Laminated safety glass uses polyvinyl butyral film as the interface between glass and glass to achieve a safe level. The early safety glass is so-called tempered glass, which is safer than ordinary glass but produces a blemish and glass around Vulnerability, so that the glass blew or external damage caused by glass splash has become a major safety concerns. At the lamination stage, many complex factors will directly affect the strength of adhesion, and then cause some abnormal situations such as delamination, deterioration, etc., resulting in reduced safety strength, the apparent appearance of abnormal phenomena. The main purpose of this study is to find out the possible effects of delamination. In the research, the factors such as manufacturing process, glass source, warpage and soaking water have been excluded.

In the industry of laminated glass often encountered the occurrence of commonly known as edge delaminated situation that is the glass and film separated off after installed in several years or even months. The industry often cannot clarify the possible causes, which caused many complaints, compensation, and even endangering safety. The cause of edge delaminated is nothing more than the glass itself warping, installation caused by external force warp, glass thickness and film thickness... and other processing factors. However, some key factors directly impact the adhesive of glass and films were ignored.

Data mining is the process of extracting the useful data, patterns and trends from a large amount of data by using techniques including classification, clustering, association and regression [1]. Soni and Ganatra presented that according to the applications, various algorithms and tools can be utilized [2]. Classification is a data mining tool based on machine learning that is used to classify each item in a dataset into a set of predefined classes [3]. Lots of algorithms have been developed for classification including decision tree, k-nearest neighbors, Bayesian and Neural-Net. The decision tree has become a significant approach of data mining that is commonly used in marketing, surveillance, fraud detection, science discovery. The basic learning approach of decision tree is greedy algorithm, which use the recursive top-down method to construct decision tree [4]. Modern classification trees can partition the data with linear splits on subsets of variables and fit nearest neighbor, kernel density, and other models in the partitions [5]. The advantages of decision tree algorithm include producing accurate outcomes, taking less memory, high classifying speed, short searching time and strong learning ability.

The purpose of this paper is finding out the influencing factors of the delaminating except the manufacturing processes. The data mining methods were applied to the information of the delaminated situation obtained by the glass manufacturer.

II. Related Works

A. Laminated Glass

The laminated glass can be also referred to as the laminated safety cemented glass which mainly uses the polyvinyl butyral film (PVB film for short) as the material to make the two pieces of glass stick together. By the high molecular OH bond in the PVB film and the OH bond in the glass component achieve bonding effect (Fig. 1).

Today, all metropolitan areas and even residential areas are flooded with tall buildings made of glass and steel. When the earthquake struck the glass frame in the building was crushed and the glass shattered as rain fell.

The unique bonding properties of laminated glass provide excellent safety protection. At the same time laminated glass used in the construction of the middle of the film because of the relationship with a good sound insulation, UV absorption,

Manuscript received January 30, 2018; revised May 1, 2018.

Hsing-Hung Lin is with the Big Data Lab. Institute of Chunghwa Telecom. Taoyuan, Taiwan, R.O.C. (e-mail: hsinhung@gamil.com).

K. T. Fan is with Special Function Products Division, Chang Chun Petrochemical Co., Ltd. Miaoli, Taiwan, R.O.C. (e-mail: princo689@yahoo.com.tw).

D. T. Chang is with the Industry Management Department, Chunghua University, Hsinchu, Taiwan (e-mail: dtchang@chu.edu.tw).

DOI: 10.7763/IJET.2018.V10.1102

anti-theft ... and other special features.

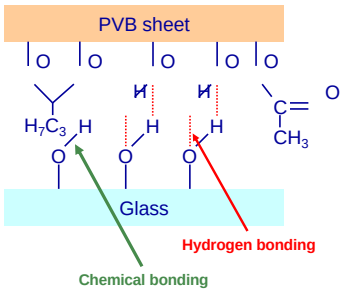

Fig. 1. Diagram of glass lamination.

The use of windshield is usually to protect the driver from the wind, extreme temperatures, oncoming debris such as dust, stones, mosquitoes caused by the destruction of the windshield after foreign bodies and broken glass splash injury, the general glass can not To meet the safety requirements, the harder glass will cause the driver to see the sight in an instant because of blew or rupture caused by external force, increasing the possibility of danger and secondary injury.

The strengthening process to reinforcement general glass changes the surface stress of the glass results in the release of stress on the four sides of the glass and makes edge warping situation. Glass and PVB film while laminated because of the large angle caused by the result that glass cannot be completely laminated. (Fig. 2.)

Fig. 2. Tree shape delamination.

Because of this phenomenon, industry often uses strong clip to assist, but it will only make the so-called pseudo-glue in the glued glass, which takes a long time to produce the glue shape like fan shape. (Fig. 3.)

Fig. 3. Fan shape delamination with moisture.

Fig. 4. Tree shape delamination with moisture.

If the glass is installed outdoors and exposed to moisture more serious cases can lead to emulsification and subsequent failure to reduce the safety of the situation (Fig. 4).

B. Data Mining

Data mining is the process of discovering potentially useful, interesting, and previously unknown patterns from a large collection of data involving tools from statistical models, mathematical algorithm and machine learning models. Data mining tools predict future trends and behaviors, allowing business to make proactive, knowledge-driven decisions [6]. Data mining is the main key step to reach the knowledge discovery.

Data mining is an interdisciplinary field merging idea from statistics, machine learning, information science, visualization and other disciplines [7]. Data mining technology provides a user-oriented approach to the novel and discovers hidden patterns in the data set [8].

C. Classification and Decision Tree

Classification is one important branch of data mining techniques that assign instances to a set of groups that makes analysis and prediction more precisely. Classification is a kind of supervised machine learning that the labeled instances should be provided in advance. The set involved labeled samples called training set. The training set is used to train the classification model called classifier. While the classifier is generated, the remaining samples said testing set is employed to test the classifier. For each instance, the prediction value get by the classifier is compared with the actual value. The accurate rate is applied to evaluate the performance of the classifier.

Based on the ID3 algorithm, an improved algorithm is proposed by Quinlan in 1993, namely C4.5 algorithm [9]. C4.5 algorithm uses Information gain ratio as the test attribute selection criteria, and each time selects an attribute with the highest information gain ratio as the test attribute for a given set. The algorithm also can handle continuous and default attribute values and adds to the pruning techniques.

Extended from C4.5 algorithm, C5.0 algorithm is the classification algorithm which applies in big data set. C5.0 performs better than C4.5 on the speed, memory consumption and efficiency. C5.0 model works by splitting the instances based on the features that provide the maximum information gain. The case subset that got from the former split will be split afterward. The process will continue until the instance subset cannot be split and is usually according to another feature. Finally, examine the lowest level split; those instance subsets that don't have remarkable contribution to the model will be rejected. C5.0 is easily handled the multi value attribute and missing attribute from data set [10]. The main pros and cons were revealed as Table I. [11], [12].

TABLE I: PROS AND CONS OF DECISION TREE ALGORITHM

Pros	1. There are no requirements of domain knowledge in the construction of decision tree. 2. It minimizes the ambiguity of complicated decisions and assigns exact values to outcomes of various actions. 3. It can easily process the data with high dimension. 4. It is easy to interpret. 5. Decision tree handles both numerical and categorical data.

Cons	1. It is restricted to one output attribute. 2. It generates categorical output. 3. It is an unstable classifier i.e. performance of classifier is depending upon the type of dataset. 4. If the type of dataset is numeric than it generates a complex decision tree.

III. Methods

A. Data Preparation and Discovery

The original data set obtained from laminated glass company involved 64968 instances. However, lots of instances of the data set consist with empty data. After discussed with the production engineers, we eliminated those instances that involved empty data. At last, we got 34176 instances. The features of the data set are those twelve factors could influence the strength of laminated glass that include *OH, Viscosity, Volatiles, Pummel, Haze, YI, transmission, Creep.23, Creep.80, ionA, ionB, ionC*.

The popular big data analytic software R has been applied to analysis the laminated glass data set. Fig. 5 reveals R code for data discovery.

```
lsg <- read.csv("data_20170710_2.csv")
str(lsg)
summary(lsg)
lsg$Viscosity <- as.numeric(lsg$Viscosity)
```

Fig. 5. R code for data discovery.

The field "*ionA"*, "*ionB"* and "*ionC"* are category features while the "*Abnorm*" is the target that we want to classify from other features. The all features of the data set descriptions are as TABLE II.

TABLE II: Descriptions of All Features

Features	Type	Examples
Abnorm	Factor with 2 levels	"N","V"
OH	numerical	19.7, 19.6, 19.6, 19.6, 19.3
Viscosity	numerical	368, 346, 348, 388, 366, 316
Volatiles	numerical	0.91, 0.98, 0.97, 0.78, 0.79
Pummel	numerical	9 7 8 5.4 6.4 7.9 5 8 3.9 6 ...
Haze	numerical	0.83, 0.56, 0.48, 0.32, 0.38
YI	numerical	1.19, 1.05, 1.07, 0.79, 0.65
transmission	numerical	87.8, 87.8, 87.9, 87.8, 87.8
Creep.23	numerical	100, 102, 110, 102, 108
Creep.80	numerical	45, 45, 47.5, 37.5, 42.5, 40
Ion A	Factor with 2 levels	"<200ppm",">200ppm"
Ion B	Factor with 2 levels	"<10ppm",">50ppm"
Ion C	Factor with 2 levels	"<5ppm",">60ppm"

B. Data Discovery

The data type of field "Viscosity" should be changed to numerical. The summary of features is presented as Table III.

TABLE III: The Arrangement of Channels

OH		Viscosity		Volatiles	
Min	14.920	Min	1.000	Min	0.010
1st Qu.	19.340	1st Qu.	295.000	1st Qu.	0.540
Median	19.480	Median	329.000	Median	0.690
Mean	19.550	Mean	312.000	Mean	0.719
3rd Qu.	19.620	3rd Qu.	358.000	3rd Qu.	0.850
Max.	28.940	Max.	405.000	Max.	80.000
Pummel		Haze		YI	
Min	0.000	Min	0.120	Min	0.010
1st Qu.	2.000	1st Qu.	0.390	1st Qu.	0.260
Median	5.000	Median	0.490	Median	0.860
Mean	4.515	Mean	1.297	Mean	1.959
3rd Qu.	7.000	3rd Qu.	0.700	3rd Qu.	1.540
Max.	9.000	Max.	67.850	Max.	91.580
transmission		Creep .23		Creep .80	
Min	5.530	Min	20.000	Min	5.000
1st Qu.	88.030	1st Qu.	82.500	1st Qu.	25.000
Median	88.510	Median	87.500	Median	30.000
Mean	89.230	Mean	88.850	Mean	32.260
3rd Qu.	90.700	3rd Qu.	97.500	3rd Qu.	37.500
Max.	98.810	Max.	157.500	Max.	401.000
ionA		ionB		ionC	
<200ppm	20588	<10ppm	13588	<5ppm	13588
>200ppm	13588	>50ppm	20588	>60ppm	20588

IV. Results

This study applied R software to build decision tree model. The R code is as Fig. 6. At first, the total data set contains 34176 cases is used to train the decision tree model through C5.0 algorithm. Fig. 7. presented the summary of training result of C5.0 tree which indicated that the feature "*ionA"* less than 200ppm makes the target *N* with 20588 cases while "*ionA"* more than 200ppm makes the target *V* with 13588 cases. The decision tree plotted as Fig. 8. that contains one node and two leaves. We can find that the key factor influenced the strength of laminated glass is feature "*ionA"*.

```
library(C50)
lsg.model <- C5.0(Abnorm~., data=lsg)
lsg.model
summary(lsg.model)
plot(lsg.model)
```

Fig. 6. The R code of decision tree generation.

```
Decision tree:
ionA = <200ppm: N (20588)
ionA = >200ppm: V (13588)
Evaluation on training data (34176 cases):
        Decision Tree
      ----------------
      Size      Errors
        2    0( 0.0%)   <<

       (a)   (b)    <-classified as
      ----  ----
     20588          (a): class N
        13588    (b): class V
```

Fig. 7. Summary of decision tree.

To verify the result of data mining from the production parameter, the experiment has been conducted in manufacture process by modified the production compound "*ionA*". From the original manufacturing process, there were 511 sets of edge delaminated glass in the production lot 5000

sets. After improving the affecting factor "*ionA*", only 43 sets of edge delaminated glass were found from the production lot 5000 sets. The results reflected in TABLE IV indicate that the influence factor "*ionA*" is the prime cause of edge delaminated.

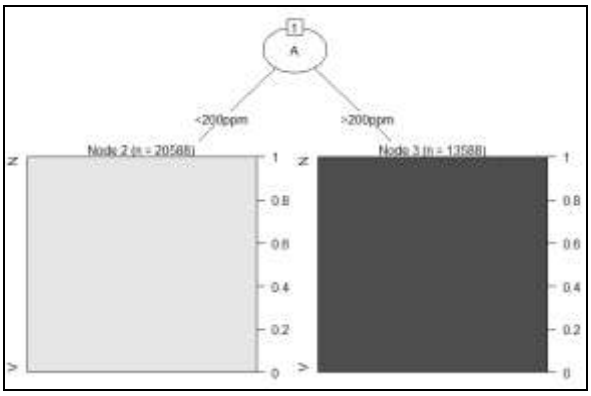

Fig. 8. The plotted decision tree.

TABLE IV: COMPARISON OF EDGE DELAMINATED GLASS

Before improving the affecting factor "*ionA*"		
Glass(qty)	Delaminated (qty)	Delaminated (ratio)
5000 sets	511 sets	10.22%
After improving the affecting factor "*ionA*"		
Glass(qty)	Delaminated (qty)	Delaminated (ratio)
5000 sets	43 sets	0.9%

V. CONCLUSION

This paper deals with exploring the key factor that influences the strength of laminated safety glass. Data mining technique were utilized to find out the factor from the data set obtained from the glass manufacture. Decision tree algorithm is chosen to classify the original data set due to the target feature consist with binary value. The decision tree classifier was trained by total 34176 cases. The feature "*ionA*" was discovered that directly correlate to the feature "*Abnorm*". By practical production, a significant improvement in the strength of laminated safety glass was obtained. It is recommended that the approach outlined in this study be replicated in other manufacturing plants.

REFERENCES

[1] M. Gera and S. Goel, "Data mining – Techniques, methods and algorithms: A review on tools and their validity," *International Journal of Computer Applications*, vol. 113, no. 18, pp. 22-29, May 2015.
[2] N. Soni and A. Ganatra, "Categorization of several clustering algorithms from different perspective: A review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 8, pp. 63-68, Aug. 2012.
[3] S. Archana and K. Elangovan, "Survey of classification techniques in data mining," *International Journal of Computer Science and Mobile Applications*, vol. 2, no. 2, pp. 65-71, Feb. 2014.
[4] B. R. Patel and K. K. Rana, "A survey on decision tree algorithm for classification," *International Journal of Engineering Development and Research*, vol. 2, no. 1, pp. 1-5, Mar. 2014.
[5] W. Y. Loh, "Fifty years of Classification and regression trees," *International Statistical Review*, vol. 82, no. 3, pp.329-348, Dec. 2014.
[6] L. Zhang and F. C. Tian, "Performance study of multilayer perceptrons in a low-cost electronic nose," *IEEE Transactions on Instrument and Measurement*, vol. 63, no.7, pp.1670-1679, July 2014.
[7] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, San Francisco, CA, USA: Morgan Kaufmann, 2011.
[8] J. Soni, U. Ansari, D. Sharma, and S. Soni. "Predictive data mining for medical diagnosis: An overview of heart disease prediction," *International Journal of Computer Applications*, vol. 17, no. 8, pp. 43-48, Mar. 2011.
[9] J. R. Quinlan, *C4.5: Programs for Machine Learning*, San Francisco, CA, USA: Morgan Kaufmann, 1993.
[10] N. Patil, R. Lathi, and V. Chitre, "Comparison of C5.0 & CART classification algorithms using pruning technique," *International Journal of Engineering Research & Technology*, vol. 1, no. 4, pp. 1-5, June 2012.
[11] D. Tomar and S. Agarwal, "A survey on data mining approaches for healthcare," *International Journal of Bio-Science and Bio-Technology*, vol. 5, no. 5, pp. 241-266, Oct. 2013.
[12] H. Leopord, W. K. Cheruiyot, and S. Kimani, "A survey and analysis on classification and regression data mining techniques for diseases outbreak prediction in datasets," *The International Journal of Engineering and Sciences*, vol. 5, no. 9, pp. 1-11, Sep. 2016.

Hsinh-Hung Lin was born in Taichung, Taiwan in July 1972. He received the B. S. degree from the Department of Computer Science at Tunghai University in 1989, M.Eng. from the Department of Industrial Engineering and Management of Chung Hua University in 1998, and Ph.D. from the Department of Industrial Engineering and Management of National Chiao Tung University in 2010.

Dr. Lin currently works as a research fellow at Big Data Lab in the Institute of Chunghwa Telecom. His current research interests include big data analytics, soft computing, artificial intelligence and systematic innovation.

K. T. Fan was born on Mach 17, 1979, in Taiwan. He received the B.E. degree from the Department of Chemical Engineering at the National United University in 2008, and the master degree from the Department of Industrial Management at Chung Hua University in 2017, respectively. He is currently a technical service staff in Changchun Petrochemical Co., Ltd..

D. T. Chang was born on August 6, 1961, in Taiwan.

Associated Professor Chang pursued his PhD studies at National Tsing Hua University, Taiwan, in 1993, he majoring in industrial engineering. His researches interests include are operations research, database management system, application of world-wide web. His research has published in *IIE Transactions* and *Journal of the Chinese Institute of Engineers*. His current major had changed to focus on the application of world-wide web.