ICED13
INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED13
19-22 AUGUST 2013, SUNGKYUNKWAN UNIVERSITY, SEOUL, KOREA

A DECISION SUPPORT SYSTEM FOR MARKET SEGMENT DRIVEN PRODUCT DESIGN

Ningrong LEI, Seung Ki MOON
Nanyang Technology University, Singapore

ABSTRACT

This paper presents a decision support system (DSS) for market segment driven product design. The input for the proposed system is historic market data and design parameters for a new product. Through market segmentation, with Principal Component Analysis (PCA) and k-means, as well as AdaBoost classification, the DSS determines to which market segment a new product belongs. To demonstrate the feasibility of the proposed system, we have conducted a case study, based on US automotive market data. In this case study, the proposed DSS achieved a classification accuracy of 92.40%. The high accuracy levels make us confident that the proposed system can benefit enterprise decision makers by providing an objective second opinion on the question: To which market segment does a new product design belong? Having the information about the market segment implies that the competition is known and marketing can position the product accurately. Furthermore, the design parameters can be adjusted such that (a) the new product fits this market segment better or (b) the new product is positioned in a different market segment. Therefore, the proposed system enables market segment driven product design.

Keywords: decision making, new product development, market segmentation, data mining

Contact:
Ningrong Lei
Nanyang Technology University
Mechanical and Aerospace Engineering
Singapore
639798
Singapore
nlei1@e.ntu.edu.sg

1 INTRODUCTION

We are living in the age of the "buyer's market" where the producer of goods must satisfy individual customer requirements (Nayak et al., 2002). This creates competitive market places which require manufacturers to introduce specific products that meet consumer demand accurately (Ben-Arieh et al., 2009). Therefore, producers have to establish their market and subsequently subdivide this market so that they can address the needs, posed by such a market segment, with a specific product. As a direct consequence, the producers are continuously monitoring their target markets and gathering data from both consumers and competitors (MacDuffie et al., 1996). The data forms the basis for market segmentation and market segment driven product design.

Since the early 1960s, market segmentation is widely considered to be a key marketing concept and a significant amount of marketing research literature focused on this topic (Green and Wind, 2004). Meyer and Lehnerd (1997) introduced a market segmentation grid to leverage product families across multiple market segments. Kumar et al. (2009) used nested logit techniques to model market segmentation, and they integrated the market segmentation with rigorous demand models to design a product family. Despite the general acceptance of market segmentation and its undisputed value for marketing, a large number of companies fail to segment their markets effectively (Green and Wind, 2004). One of the causes for this failure comes from the fact that managers do not base their strategies on the evolving need of target segments. Yankelovich and Meer (2006) postulated that effective segmentations should be dynamic. There is a constant need to reshape the market segment according to market conditions, such as fluctuating economics, emerging consumer niches, and new technologies. Today, these fundamental considerations are more pressing than ever, because through globalization markets evolve at an increasing rate (Kaynak and Hassan, 1994). Yankelovich and Meer (2006) pointed out that management decisions were affected by the availability and use of dynamic market data. However, many companies lack both the dynamic market data and the expertise to extract useful information from the market data to make informed decisions and act on them (Lei and Moon, 2012). Therefore, it is essential to make market data directly accessible to business leaders.

This paper presents a Decision Support System (DSS) for market segment driven product design. We adopt the position that data mining and decision support tools make market data directly available to business leaders and this can improve management decisions. To support this statement, we propose a system that automates the processes of market segmentation and market segment membership determination of new product designs. A combination of Principal Component Analysis (PCA) and k-means clustering identifies market segments. An AdaBoost classifier determines the market segment to which a new product design belongs. The input of the proposed DSS is market data, i.e. information about the products in a market, and a subset of the design specifications for a new product. The output is the market segment to which the new product belongs. This information can benefit enterprise decision making in a number of ways. The market segment indicates the competition for the new product by revealing the market structure of the competitors, this helps to identify a market segment for the new product correctly. Furthermore, the proposed DSS can be used in 'what if' scenarios, where the design specifications of the new product are altered and the outcome of the market segment decision is observed. Another benefit is that the proposed DSS works even with incomplete parameters, that means all the benefits listed before can be realized early in the design phase. To demonstrate the practicality of the proposed DSS for market segment driven design, we present a case study based on the US automotive market in 2010. The proposed system obtained 92.4% accuracy, in classifying a new product specification into one of six market segments, with a full set of parameters. The result was obtained with 10 fold stratified cross validation. Therefore, the proposed DSS can be applied to other markets such as consumer electronics market.

The rest of the paper is organized as follows: Section 2 gives a review of data mining methods in the area of product design. Section 3 introduces the proposed DSS. Section 4 discusses a case study that was used to test the proposed DSS for market segment driven product design. Conclusions and further work are presented in Section 5.

2 LITERATURE REVIEW

Data mining, which deals with the discovery of hidden knowledge, unexpected patterns and new rules from large databases, is regarded as the key element of knowledge discovery in databases (Adriaans and Zantinge, 1996). Therefore, this method can be used for analyzing the ever-growing quantity of high dimensional data. It can bring significant gains to organizations through better-target marketing and enhanced internal performance (Adriaans and Zantinge, 1996). Zhang et al. (2007) used a fuzzy clustering method to analyze customer purchase behaviors, based on this information they established market segments. Moon et al. (2010) introduced a knowledge discovery method which employs fuzzy clustering and association rule mining to improve module-based product platform design. Shao at al. (2006) integrated fuzzy clustering and variable precision rough sets to discover the mapping between customer groups and product configuration alternatives.

K-means (Hartigan and Wong, 1979), is one of the most popular and efficient clustering methods, it uses prototypes (centroids) to represent clusters by optimizing the squared error function. High dimensional data are often transformed into lower dimensional data via PCA where coherent patterns can be detected more clearly (Jolliffe, 2002). Self-organized feature map is also widely used for dimension reduction and clustering (Aguado et al, 2008). Balakrishnan et al. (1996) employed the frequency-sensitive competitive learning algorithm and the k-means method for clustering the simulated data and real-world problem data. They also presented the combination of these two methods. But which method is better to solve the real-world problem was not determined.

A number of methods have been proposed to solve the market segmentation and new products positioning problems, however each carries its own advantages and shortcomings (Punj and Steward, 1983). The literature review shows that data mining is only the first step to provide meaningful management decision support. Data mining extracts specific features, from potentially large data, and these features represent important information. However, the next task is to interpret these features. In this paper, we combine data mining with automated decision making and thus the burden of feature interpretation is shifted to the decision making algorithm. Therefore, the result from the decision making is easy to interpret and there is little or no room for errors through misinterpretation. For example, the result from the proposed DSS is the market segment of a new product design. This is much easier to understand for enterprise decision makers than clustering results.

3 PROPOSED DECISION SUPPORT SYSTEM

This section describes the methods used in the proposed DSS for market segment driven product design. Figure 1 shows an overview block diagram of the proposed DSS. The proposed system takes market and design data as input and it produces decision support information as output. The market data contains all relevant properties of the products in a specific market. This data is fed into the objective market segmentation step which labels the products according to the market segment they belong to. Using the labeled data, a classification algorithm is trained. The training process yields cluster information which is used in the classification step. To be specific, the classification step decides to which market segment a new design belongs based on both the information extracted from the classifier training step and the design data. As corollary, the proposed system provides number, size, characteristics, and leading brands of each market segment. Both classification result and corollary information constitute the output, i.e. the decision support.

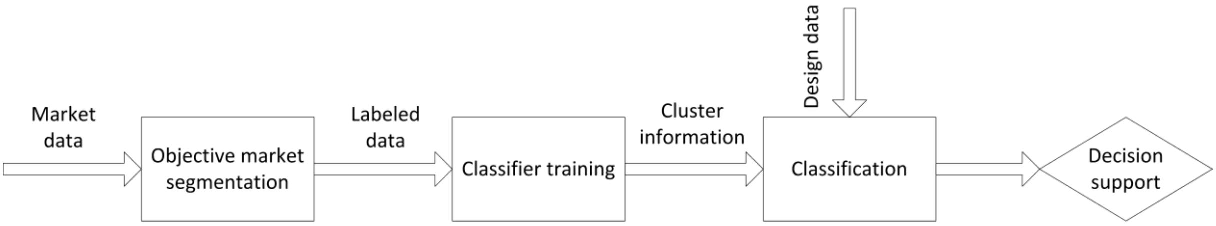

Figure 1. Overview diagram of the proposed DSS system

The next sections introduce the algorithms which are used in the DSS. Section 2.1 introduces PCA, as a dimension reduction technique, which is part of the objective market segmentation step. Section 2.2 discusses k-means clustering as a method for discriminating individual market segments. The AdaBoost training algorithm is introduced in Section 3.3. Section 3.4 introduces stratified cross validation as a method for classification testing.

3.1 Principal Component Analysis (PCA)

PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (Jolliffe, 2002). The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e. uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables.

3.2 K-means

MacQueen (1967) has introduced k-means clustering as a data mining method which is based on cluster analysis. It partitions n observations into k clusters such that each observation belongs to the cluster which has the nearest mean. This results in a partitioning of the data space into Voronoi cells (Hartigan and Wong, 1979). Algorithmically k-means uses a two-phase iterative algorithm to minimize the sum of point-to-centroid distances, summed over all k-clusters.

3.3 AdaBoost

Kearns and Valiant (1994) investigated whether or not it is possible to boost the prediction quality of a weak learner, even if the prediction accuracy, of this learner, is just slightly better than a random guess. This sparked a number of improvements on boosting algorithms. For example, Freund and Schapire (1993) introduced the AdaBoost algorithm, which solved many of the practical shortcomings of earlier algorithms.

The AdaBoost is a machine learning algorithm which feeds the input training set to a weak learner algorithm repeatedly. During these repeated calls, the algorithm maintains and updates a set of weights for the training set. Initially, all weights are equal. However, after each call, the weights are updated such that the weights of incorrectly classified examples are increased. This forces the weak learner to focus on the hard examples in the training set. In this study, we have used the AdaBoost implementation from Paris (2012).

3.4 Stratified cross validation

Stratified cross validation is a technique that assesses how the results from a statistical analysis will generalize to an independent data set (Schaffer, 1993). We use stratified cross validation to test the AdaBoost classification accuracy, because Kohavi (1995) reported that stratified cross validation performs better (has smaller bias and variance) than regular cross validation. The algorithm starts by partitioning the labeled product data, from the market segmentation step, into 10 equally sized disjoint subsets called *folds*. During the partitioning, the algorithm ensures that each class (market segment) is uniformly distributed over all folds (Breiman, 1984). Each of the 10 folds is then in turn used as the test set, while the remaining 9 folds are used as the training set. The AdaBoost classifier is then constructed from the training set, and its accuracy is evaluated on the test set. This process repeats 10 times, with a different fold used as the test set each time. The estimated true accuracy by this method is the average over the 10 folds.

4 CASE STUDY

This section discuses implementation and testing of the proposed DSS. Figure 2 shows both implementation and test setups. We used US automotive market data from 2010 as input, and the test setup output was classification result data. These results were obtained in a two-step process. The first processing step established the market segments by subjecting the automotive market data to PCA and k-means. The result from the first processing step was a label vector that assigns each product, from the market data, to a specific market segment. In the second processing step, the labeled market data was used to train and test the AdaBoost classifier.

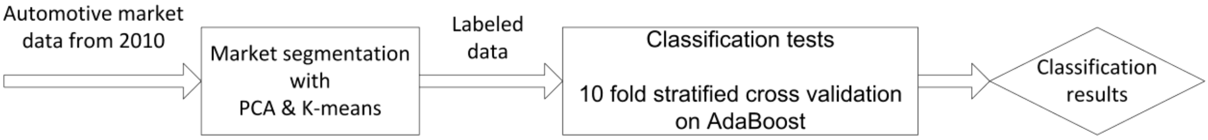

Figure 2. Overview diagram of the test setup

4.1 Automotive market data

In this case study, we used Ward's Automotive Group (2010) data. Table 1 provides example data from the Acura brand. The table lists 16 properties of five car models. These properties are:
Wheel Base (WB), Length, Width, Height, Weight, Cylinder Type (CT), Cylinder Arrangement (CA), Number of Cylinders (NC), Cubic Inches Displacement (CID), Liter, Valves per Cylinder (V/C), Horse Power (HP), RPM, miles per gallon in the City (C), miles per gallon on the Highway (H), Price (P).

Table 1. Properties of the five car models from the Acura brand. The car models are: (1) RL, (2) TL, (3) TL SH-AWD, (4) TSX, and (5) TSX V-6. The data is based on Ward's Automotive group (2010)

Model	WB	Length	Width	Height	Weight	CT	CA	NC	CID	Liter	V/C	HP	RPM	C	H	P
1	110.2	195.7	72.7	57.2	4083	3	3	6	224	3.7	4	300	6300	16	22	47640
2	109.3	195.3	74	57.2	3721	3	3	6	212	3.5	4	280	6200	18	26	35915
3	109.3	195.3	74	57.2	3889	3	3	6	223	3.7	4	305	6300	18	26	39465
4	106.4	185.6	72.4	56.7	3400	1	1	4	143.6	2.4	4	201	7000	20	28	30120
5	106.4	185.6	72.4	56.7	3680	3	3	6	212	3.5	4	280	6200	18	27	35660

The complete market data lists all the products (car models) from the following brands:
Acura, Audi, Bentley, BMW, Buick, Cadillac, Chevrolet, Chrysler, Dodge, Ford, Honda, Hyundai, Infiniti, Jaguar, Kia, Lexus, Lincoln, Maybach, Mazda, Mercedes-Benz (MB), Mercury, Mini, Mitsubishi (MS), Nissan, Pontiac, Porsche, Rolls-Royce (RR), Scion, SMART, Subaru, Suzuki, Toyota, Volkswagen (VW), Volvo.
The overall number of products in the market was: 608.

4.2 Results

This section descries both market segmentation and classifier testing results. The result of the market segmentation step was a labeled data set, where the labels expressed the market segment.

4.2.1 Labeled data

The market segmentation step generated six distinct market segments for the automotive market data from 2010. These six clusters were based on the results of PCA processing. Figure 3 shows the first three principle components as axis of a three dimensional space and the points in this space were the result of applying PCA to the automotive market data. The additional lines in Figure 3 indicate the properties with which the initial automotive data was structured.
The first three principle components were used as input to the k-means clustering algorithm. We found the amount of clusters (market segments) by calculating the silhouette mean for different target clusters. Figure 4 a) shows the silhouette mean over the number of clusters. In this case we tested from 2 to 25 target clusters. The higher the silhouette mean value is the better does the number of clusters reflect the data structure. The marker × in Figure 4 a) indicates that the highest silhouette mean of 0.79 was obtained with six clusters. Hence, the market data should be partitioned into six market segments. Figure 4 b) shows the silhouette plot for these six clusters.

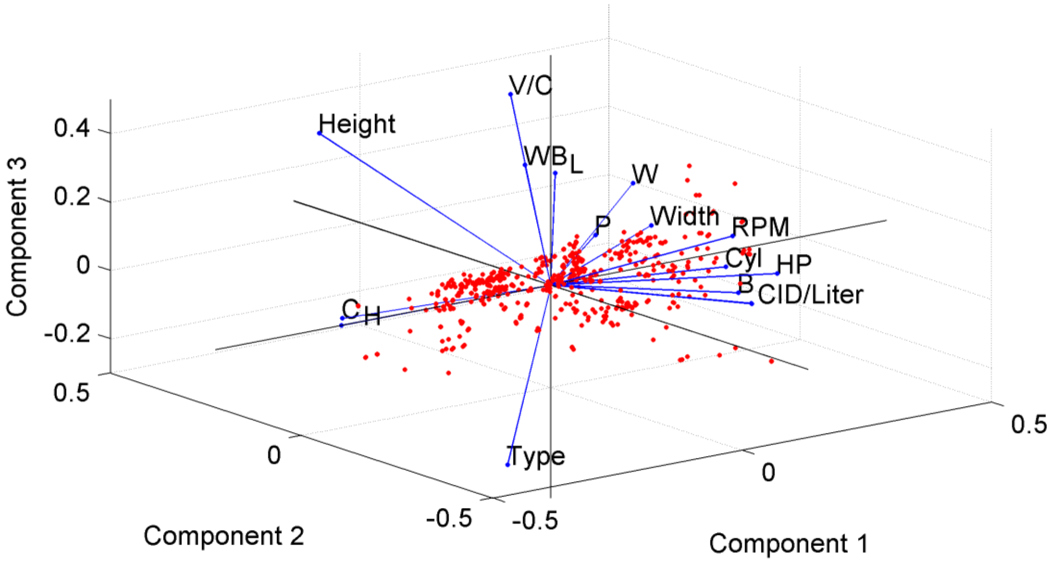

Figure 3. PCA result for the first three principal components. Each variable is labeled with the appropriate automotive data property

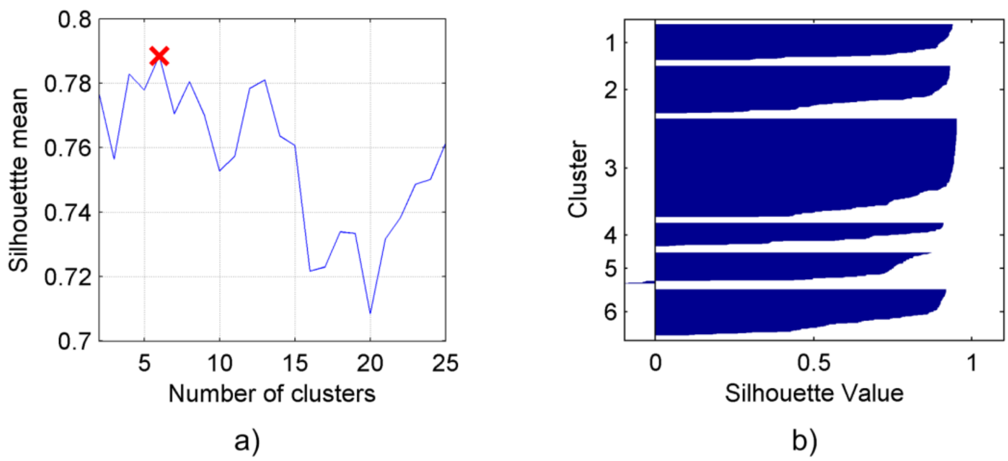

Figure 4. Clustering results
Graph a) Dependency of the silhouette mean on the number of clusters.
Graph b) silhouettes of the six clusters

Table 2 presents the market segmentation result. It shows the mapping of the individual car model to specific market segments (clusters). The market segments from 1 to 6 contain 77, 102, 211, 52, 67, and 99 products respectively. For example, four car models from the Acura brand were mapped to cluster 1 and one car model was mapped to cluster 4. According to American automobile classification, the market segments from C1 to C6 represent ultra luxury market, full-size luxury market, SUV market, entry-level luxury, subcompact, and roadster respectively. Further analysis shows that each market segment is represented by specific brands, the respective market segment identification brands are: Maybach and RR, Jaguar and Bentley, Kia, BMW, Mini, Porsche and Infiniti respectively. These leaders were established by determining which brand has the highest percentage of car models in a specific cluster. For example, Jaguar has 100% of its car models in cluster 2; therefore this brand is a market segment leader. Each segment has its primary distinguishing features. For instance, for C1, the price of the models is over 100,000 US dollar, and the length of the wheelbase is at least 195 inches. For C5, the length of the wheelbase is less than 165 inches.

Table 2. The table details which brand has how many cars in what cluster. C1 to C6 stands for Cluster 1 to Cluster 6, N stands for the total number of products for a specific brand

Brand	C1	C2	C3	C4	C5	C6	N	Brand	C1	C2	C3	C4	C5	C6	N
Acura	4	0	0	1	0	0	5	Maybach	7	0	0	0	0	0	7
Audi	6	8	10	4	14	7	49	Mazda	0	0	13	8	0	2	23
Bentley	2	11	0	0	0	0	13	MB	5	16	0	4	0	4	29
BMW	0	13	2	16	1	7	39	Mercury	1	0	3	0	0	1	5
Buick	3	1	0	2	0	2	8	Mini	0	0	0	0	9	0	9
Cadillac	0	6	0	0	0	6	12	MS	2	0	6	3	5	0	16
Chevrolet	8	2	15	0	0	2	27	Nissan	0	2	19	1	1	9	32
Chrysler	7	0	4	0	1	1	13	Pontiac	0	0	5	0	0	0	5
Dodge	9	4	7	0	1	0	21	Porsche	0	5	0	0	0	16	21
Ford	12	2	10	0	0	7	31	RR	4	1	0	0	0	0	5
Honda	4	0	7	2	16	0	29	SCION	0	0	3	0	0	0	3
Hyundai	0	4	14	1	0	4	23	SMART	0	0	5	0	0	0	5
Infiniti	0	2	0	0	0	12	14	Subaru	0	0	5	7	11	3	26
Jaguar	0	14	0	0	0	0	14	Suzuki	0	0	16	0	0	0	16
Kia	0	0	19	0	0	0	19	Toyota	0	0	17	0	4	6	27
Lexus	0	8	2	3	0	6	19	VW	0	0	17	0	4	2	23
Lincoln	3	2	0	0	0	2	7	Volvo	0	1	12	0	0	0	13

4.2.2 Classification results

We constructed eight different cases to model scenarios which are likely to happen during market segment driven product design for the automotive market. All cases used 10 fold stratified cross validation to establish the results. In the first scenario, we supposed that all product properties are known, so all available product properties were used to train and test the AdaBoost. In this configuration, the classification algorithm achieved a classification accuracy of 99.09% and 92.40% for training and testing respectively. Table 3 details both properties used and classification performance for all the scenarios. In scenario 2, where we left out the price parameter, the AdaBoost achieved a classification accuracy 93.80% for testing, this is actually 1.40% higher than the testing accuracy for the complete set of parameters. That means, without the price the prediction is more accurate. Scenarios 3 and 4 were based on the design information from both chassis and engine. Scenarios 5 and 6 were based on engine data alone. Scenarios 7 and 8 were constructed only with chassis information.

Table 3. Accuracy results for the AdaBoost training and test phase for eight different cases from 1 to 8. Each case has a different subset of product properties, an element of this subset is indicated by '•'

Scenario No.	WB	Length	Width	Height	Weight	CT	CA	NC	CID	Liter	C/V	HP	RPM	C	H	Price	Training accuracy	Testing accuracy
1	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	99.09%	92.40%
2	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		98.49%	93.80%
3	•	•	•	•	•	•	•	•	•	•	•	•	•				97.55%	93.00%
4	•	•	•	•		•	•	•	•	•	•	•	•				97.78%	93.60%
5						•	•	•	•	•	•	•	•			•	98.45%	93.00%
6						•	•	•	•	•	•	•	•				95.99%	92.20%
7	•	•	•	•												•	81.28%	69.60%
8	•	•	•	•													76.53%	69.20%

5 CONCLUSION

This paper proposes a DSS for market segment driven product design. The proposed system consists of a market segmentation and a decision making step. The market segmentation step is based PCA and k-means clustering. The decision making step is based on the AdaBoost meta classification algorithm. We used data from the North American automotive market in 2010 to construct a case study with which the proposed DSS was tested. During these tests, we found that, with a full set of product parameters, the proposed system is able to achieve a classification accuracy of 92.40%. Another noteworthy result was that the classification accuracy increased to 93.83% when the price was removed from the input parameter list. That means, in this DSS price is not a good market segment indicator. Despite this exception, we found that the classification accuracy, i.e. the ability of the system to find the correct market segment, diminishes in scenarios with less design parameters.

With the proposed system, we aim to provide objective support information for enterprise decision makers. The proposed DSS helps management to select target market segments, and suggests the market segment to which a new product design belongs, by providing number, size, characteristics, performance/price tiers, and leading brands of each segment. Through the case study, we established that the proposed system worked accurately even in a case when there was an incomplete set of design data. Therefore, enterprise decision makers can obtain valuable decision support in an early design stage. To be specific, the proposed system suggests a specific market segment for the new product design. Having this market segment implies that the competition, in terms of products and brands, is known as well. Therefore, the proposed tool enables companies to tailor a new product for a specific market segment. Such market segment specific products are one way to be profitable in the age of the buyer's market.

Thought the proposed DSS has aforementioned advantages, it has also a number of distinct limitations. Most of these limitations come from the ideas of objectivity. The fundamental problem is that even digital processing machinery is not entirely objective. These machines have been built by humans to do specific tasks, i.e. a classifier is build according to specific rules and parameters to mimic human decision making. Furthermore, these machines act on specific input data, but this data is selected according to subjective criteria. In the practical test case, we used automotive data from Ward's Automotive Group, a division of Penton Media Inc. (2010), the decision to use this data was entirely subjective. Another limitation of this work is that we did not include a way to rank the individual input parameters. For example, human experts place lots of emphasis on the price when they do market segmentation in contrast our system treats all input parameters with equal importance. The reason why this weighting feature is absent comes from the fact that data weighting is subjective and our aim was to produce a system which is as objective as possible. In other words, weighting the input data would lead to market segmentation and subsequently to decision support which is very much dependent on human experts. But this dependence would limit the usefulness of the proposed system, because these human expert decisions on weighting the input data are not transferable between different markets. In our case, the proposed DSS works for any market where appropriate market data is available with a minimum on subjective human intervention.

Future work will focus on extending the ideas of the proposed DSS to platform based product family design. We plan to use data mining and artificial intelligence methods to automate tradeoff decisions between commonality and distinctiveness within a family of products. Such a system will provide condensed information and thereby it will make large amounts of data directly accessible to the management. This will enable enterprise decision makers to reshape their product families such that they address the moving markets in the globalized economy.

ACKNOWLEDGEMENTS

This work was supported by a SUG Grant from Nanyang Technological University.

ACRONYMS

C	Miles per Gallon in the City
CA	Cylinder Arrangement
CID	Cubic Inches Displacement
CT	Cylinder Type
DSS	Decision Support System
H	miles per gallon on the Highway

HP Horse Power
MB Mercedes-Benz
MS Mitsubishi
NC Number of Cylinders
P Price
PCA Principal Component Analysis
RR Rolls-Royce
V/C Valves per Cylinder
VW Volkswagen
WB Wheel Base

REFERENCES

Adriaans, P. and Zantinge, D. (1996) *Data mining*, Harlow, England, Mass: Addison-Wesley.

Aguado, D., Montoya, T., Borras, L., Seco, A. and Ferrer, J. (2008) 'Using SOM and PCA for analyzing and interpreting data from a P-removal SBR', *Engineering Applications of Artificial Intelligence*, Vol. 21, no. 6, pp. 919-930.

Balakrishnan, P. V. S., Cooper, M. C., Jacob, V. S. and Lewis, P. A. (1996) 'Comparative performance of the FSCL neural net and *K*-means algorithm for market segmentation', *European Journal of Operational Research*, Vol. 93, no. 2, pp. 346-357.

Ben-Arieh, D., Easton, T. and Choubey, A. M. (2009) 'Solving the multiple platforms configuration problem', *International Journal of Production Research*, Vol. 47, no. 7, pp. 1969-1988.

Breiman, L. (1998) *Classification and regression trees*, Boca Raton, Fla., Chapman and Hall/CRC.

Freund, Y. and Schapire, R. (1995) 'A decision-theoretic generalization of on-line learning and an application to boosting', in P. Vitányi, ed., Computational Learning Theory, *Lecture Notes in Computer Science*, Springer Berlin, Heidelberg, Vol. 904, pp. 23-37.

Green, P. E. and Wind, Y. (2004) *Marketing research and modeling: progress and prospects: a tribute to Paul E. Green.*, International series in quantitative marketing: 14, Boston, Kluwer Academic Publishers.

Hartigan, J. A. and Wong, M. A. (1979) 'Algorithm as 136: A k-means clustering algorithm', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, no. 1, pp. 100-108.

Jolliffe, I. T. (2002) *Principal Component Analysis*, Springer Series in Statistics, Springer.

Kaynak, E. and Hassan, S. (1994) *Globalization of Consumer Markets: Structures and Strategies*, New York, International Business Press.

Kearns, M. J. and Vazirani, U. V. (1994) *An introduction to computational learning theory*, Cambridge, MIT Press.

Kohavi, R. (1995) 'A study of cross-validation and bootstrap for accuracy estimation and model selection', *in* 'Proceedings of the 14th international joint conference on artificial intelligence - Volume 2', IJCAI'95, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1137-1143.

Kumar, D., Chen, W. and Simpson, T. W. (2009) 'A market-driven approach to product family design', *International Journal of Production Research*, Vol. 47, no. 1, pp. 71-104.

Lei, N. R. and Moon, S. K. (2012) 'Objective product family design analysis using self-organization map', *in 'Industrial Engineering and Engineering Management, IEEE International Conference*, Hong Kong, China, pp. Ahead of print.

MacDuffie, J. P., Sethuraman, K. and Fisher, M. L. (1996) 'Product variety and manufacturing performance: Evidence from the international automotive assembly plant study', *Management Science* Vol. 42, no. 3, pp. 350-369.

MacQueen, J. (1967) 'Some methods for classification and analysis of multivariate observations', *in* '*Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*', Vol. 1, California, USA, pp. 14.

Meyer, M. H. and Lehnerd, A. P. (1997) *The power of product platforms: building value and cost leadership*, New York, Free Press.

Moon, S. K., Simpson, T. W. and Kumara, S. R. T. (2010) 'A methodology for to support product family design', *Annals of Operations Research*, Vol. 174, no. 1, pp. 201-218.

Nayak, R. U., Chen, W. and Simpson, T. W. (2002) 'A variation-based method for product family design', *Engineering Optimization*, Vol. 34, no. 1, pp. 65-81.

Paris, S. (2012) 'Multiclass gentle Adaboosting', Matlab central, file exchange.

Punj, G. and Steward, D. W. (1983) 'Cluster analysis in marketing research: review and suggestions for applications', *Journal of Marketing Research*, Vol. 20, no. 2, pp. 134-148.
Schaffer, C. (1993) 'Selecting a classification method by cross-validation', *Machine Learning,* Vol. 13, no. 1, pp. 135-143.
Shao, X. Y., Wang, Z. H., Li, P. G. and Feng, C. X. J. (2006) 'Integrating data mining and rough set for customer group-based discovery of product configuration rules', *International Journal of Production Research,* Vol. 44, no. 14, pp. 2789-2811.
Ward's Automotive Group, a division of Penton Media Inc. (2010) '10 model car US specifications and prices – final', Ward_LV_Specs_2010.xls.
Yankelovich, D. and Meer, D. (2006) 'Harvard business review', Vol. 84, no. 2, pp. 122-131.
Zhang, Y., (Roger) Jiao, J. and Ma, Y. (2007) 'Market segmentation for product family positioning based on fuzzy clustering', *Journal of Engineering Design,* Vol. 18, no. 3, pp. 227-241.