# Attribute identification and predictive customisation using fuzzy clustering and genetic search for Industry 4.0 environments

Alfredo Alan Flores Saldivar [1], Cindy Goh [1], Yun Li [1,*]
1School of Engineering, University of Glasgow, Glasgow G12 8LT, U.K.
a.flores-saldivar.1@research.gla.ac.uk, Cindy.Goh@glasgow.ac.uk,
*Corresponding author: Yun.Li@glasgow.ac.uk

Hongnian Yu [2]
[2] Faculty of Sciences & Technology
Bournemouth University
Talbot Campus, Poole BH12 5BB, U.K
yuh@bournemouth.ac.uk

Yi Chen [3]
[3] School of Computer Science and Network Security,
Dongguan University of Technology,
Songshanhu, Guangzhou 523808, China
chenyi@dgut.edu.cn

***Abstract*— Today´s factory involves more services and customisation. A paradigm shift is towards "Industry 4.0" (i4) aiming at realising mass customisation at a mass production cost. However, there is a lack of tools for customer informatics. This paper addresses this issue and develops a predictive analytics framework integrating big data analysis and business informatics, using Computational Intelligence (CI). In particular, a fuzzy c-means is used for pattern recognition, as well as managing relevant big data for feeding potential customer needs and wants for improved productivity at the design stage for customised mass production. The selection of patterns from big data is performed using a genetic algorithm with fuzzy c-means, which helps with clustering and selection of optimal attributes. The case study shows that fuzzy c-means are able to assign new clusters with growing knowledge of customer needs and wants. The dataset has three types of entities: specification of various characteristics, assigned insurance risk rating, and normalised losses in use compared with other cars. The fuzzy c-means tool offers a number of features suitable for smart designs for an i4 environment.**

**Keywords—Smart manufacturing, Industry 4.0, smart design, big data analytics, fuzzy clustering, genetic search.**

## I. Introduction

Historically industrial revolutions had led to a paradigm shift, starting with the steam-motor improvement in the 18th century, then mass production systems in the early 19th century because of electricity commercialization, and to the advancement of ICT and introduction of automation systems in the late 20th century. Innovation in manufacturing industry has been building innovative advances that revolutionised the way products were manufactured, services were given and business were made. Advances in ICT technologies have currently and repeatedly progressed in numerous fields, those include software and hardware; that might bring a revolution or evolution to manufacturing industry. For this revolution, smart manufacturing could have the driving force. Integration of various technologies can promote a strategic innovation of the existing industry through the convergence of technology, humans, and information. On the other hand, lean manufacturing targeted cost saving by focusing on waste elimination, this during 1980's and 1990's. In contrast, smart manufacturing represents a future growth engine that aims for sustainable growth through management and improvement of the major existing factors, like: quality, flexibility, productivity, and delivery based on technology convergence as well as numerous elements over societies, environment and humans [1].

Recently i4 has been not much more than a concept [2]. The main idea of i4 is the combination of several technologies and concepts such as Smart Factory, CPS, industrial Internet of Things (IoT), and Internet of Services (IoS) interacting with one another to form a closed-loop production value chain [3]. Differing from other ambitious strategies like the Advanced Manufacturing Partnership in the US [3] and the "Manufacturing 2025" plan in China, is the benefit inside production line: variety vs productivity. Not many industries can produce individual goods in a completely automated fashion. For this to become a reality, not only the machines but occasionally even the parts themselves need to become smart [4].

The focus of this paper is to address the integration of several technologies in a closed-loop cycle such that information from existing inputs, can be retrieved to obtain better prediction for decision-making and customized the intelligent design of products. This framework is proposed under the i4 principles due to the capacity of integration with cloud computing, big data analytics, ICT, CPS, and business informatics inside manufacturing production systems. The aim of this research is to utilize fuzzy c-means and Genetic Algorithm (GA) selection for customized designs for smart manufacture, where prediction and selection of best attributes and customers' needs and wants can be achieved.

In Section II of this paper, challenges and trends of i4 are discussed, together with the issues surrounding mass customisation. In Section III, we tackle the issue of smart design for mass customisation and present a self-organizing tool for predicting customer needs and wants. We demonstrate the effectiveness of the proposed methodology through a case study in Section IV. Lastly, Section V draws conclusions with discussions on future work.

A. A. Flores Saldivar is grateful to CONACYT for a Mexican Government research scholarship.

## II. CUSTOMISATION FOR INDUSTRY 4.0

Coined in the late 80's, the term mass-customized production has become a subject of research along with the proliferation of information throughout the IoT in the 21st century affecting business strategies and acquiring goods & services [5]. This implicates that mass customisation in manufacturing's supply chain, material flow and information concerns, and connection between product types had a direct effect on customer satisfaction [6].

Customized manufacturing describes a process for which all involved elements of the manufacturing system are designed in a certain way that enable high levels of product variety at mass production costs [5] - the reason why companies today are facing challenges as a result of customers' increasing demand for individualized goods and services. With the development and introduction of CPS into the manufacturing process, manual adjustments and variations on product quality can be minimized by connecting the virtual part of the process through computer-aided design (CAD) and comparing the desired information to target optimal features. Finally, all the streamed data that intervene with the process helps to monitor the manufacturing process and apply changes if necessary. From here, the idea of having a closed loop to constantly retrieve information in the customized design and customer satisfaction results in more informed processes and leads to reliable decisions [7].

The next section describes how data and CPS can be integrated into a framework for manufacturing application.

### A. *CPS and data analytics framework for smart manufacturing*

In recent years, the use of sensors and networked machines has increased tremendously, resulting in high volumes of data known as big data being generated [8]. In that way, CPS, which exploits the interconnectivity of machines, can be developed to manage big data to reach the goal of resilient, intelligent, and self-adaptable machines. Boost efficiency in production lines for meeting customers' needs and wants is key in i4 principles, and since CPS are still in experimental stage, a proposed methodology and architecture described in [9] which consists of 2 main components: (1) the advanced connectivity that guarantees real-time data procurement from the physical world and information feedback from the digital space; and (2) intelligent data analytics, management, and computational capability that constructs the cyber space. Fig. 1presents the value creation when combining CPS from an earlier data acquisition, and analytics.

From the above framework, the smart connection plays an important role, hence aqcuiring reliable and accurate data from machines including components and customers' feedback telling the insides of the design that best approaches to their needs and wants. Here is where enterprise manufacturing systems interviene such as enterprise resource planning (ERP), manufacturing exectution system (MES), and supply chain management (SCM). Data is obtained from those types of systems that update information in real time and provide a reliable inside of the product, from there all that collected data can be transformed into action [9].

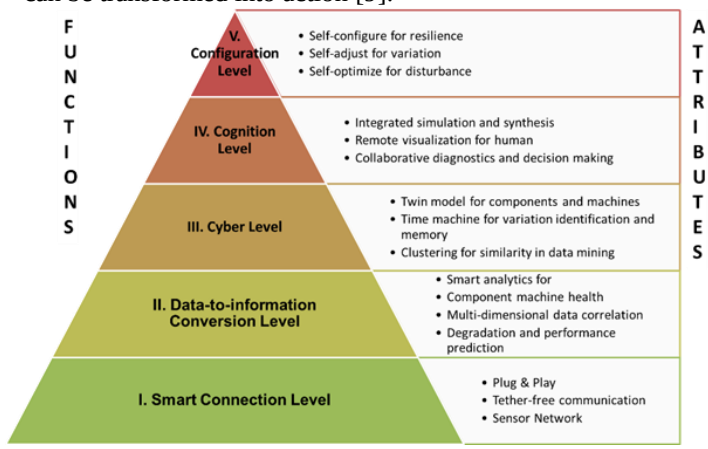



Fig. 1. Architecture for implementing CPS [9]

i4 also describes the overlap of multiple technological developments that comprise products and processes. The purpose of this paper is to provide a robust methodology to find possible solutions to fill the missing gaps that big data offers to individualistic manufacture (customized production). The next section discusses the relation between smart products and machine learning for i4 environments.

### B. *Smart products and product lifecycle for Industry 4.0*

Defined by [10], a smart product is an entity (software, tangible object, or service) made and designed for self-organized embedding (incorporation) into different (smart) environments in the direction of its lifecycle. The smart product provides boosted simplicity and openness through improved Product-to-user & Product-to-product interaction by means of proactive behaviour, context-awareness, semantic self-description, Artificial Intelligence (AI) planning, multimodal natural interfaces, and machine learning.

The interaction with their environment is what makes a product smart. Under the i4 principles, each product is tag with an identity for example, using Radio Frequency Identifiers (RFID). This result in the increase in volume, variety and velocity of data creation, which poses a challenge for identifying best, attributes in smart product designs to detect exactly what customers really want as an individual product. Today with the IoT, data is collected constantly creating a continuous stream of data, leading to an evolve data that comprises videos, sounds and images that can trigger best design for products, better quality, meet customer needs and wants, and process operations [11].

The digitalization of the value chain, how to optimize a process, and bring flexibility lead to a whole value chain fully integrated. Customers and suppliers are included in the innovation of the product, through social software [12]. Then cloud services connect to the networked product in the use phase. During its entire lifecycle the product stays connected and maintain data collection, here big data can be used to create a feedback loop into the production phase, using algorithms and models that are able to process data in an unprecedented velocity, volume and variety [13].

Creating smart products for i4 technologies also lead to determine the necessary base technologies, those can be named

as follows: mobile computing, big data and Cloud Computing [11]. More than providing scalable compute capacity, i4 aims to provide services that can be accessed globally via the Internet, here lies the importance of cloud computing and mobile computing [14]. For this in [11] is proposed the framework depicted in Fig. 2.

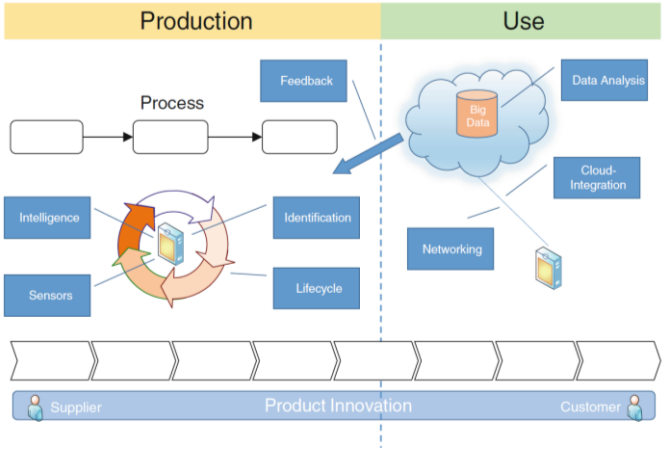



Fig. 2. Framework for smart product's innovation [11]

The management and analysis of data is key to this work. CPS will only implement mass production, but mass customisation needs to be designed beforehand, and it is often found that customer is not clear what their needs and wants are [15]. Eventually, how data is managed will lead to evolution for the innovation floor by this constant communication and linkage that IoT enables.

Next section reviews the machine learning techniques together with Computational Intelligence (CI) for addressing prediction in customized production.

### *C. Computational intelligence for customized production*

Discussed previously, the main components of the i4 or factory of the future vision are: CPS with the ability to connect everything through the IoT and IoS, in digitalized environment, comprising decentralized architectures and real-time capability to analyse huge quantities of data (big data analytics) in a modular way.

In this context classical and novel Machine Learning and CI techniques, among which Artificial Neural Networks (ANN), which have been developed exactly to extract (hidden) information from data for pattern recognition, prediction issues, and classification find a natural field of application. Such techniques have a huge potential to provide a clear improvement of many transformation processes, as well as to services by providing reliable insides of what customers' really need & want.

Addressing prediction in larger datasets can be but one application of Machine Learning techniques, but first it's necessary to understand the characteristics of the data in order to find the most suitable method according to data inputs [16]. A good understanding of the dataset is crucial to the choice and the eventual outcome of the analysis. Within the context of i4.0, there are two main sources of data: human-generated data and machine-generated data, both present huge challenges for data processing. Many of the algorithms developed so far are iterative, designed to learn continually and seek optimized outcomes. These algorithms iterate in milliseconds, enabling manufacturers to seek optimized outcomes in minutes versus months.

Facing the era of the IoT in [17] is discussed the integration of machine learning databases, applications, and algorithms into cloud platforms and most of all automate process because of the feasibility of controlling high-complex process. An architecture is proposed by [17] and presented in Fig. 3.

This presented framework englobes four key components: customer relationships, design & engineering, Manufacturing & supply Chain, and Service & Maintenance. The Enterprise business process are connected inside the cloud that retrieves information already processed from the industrial equipment. Here is used intelligence in the form of systems service agent. Then local technicians report events, status or alarms if necessary for remote experts to evaluate each event; in this process business intelligence takes part when accessing all the data that the platform Hadoop processed to generate prediction models. Finally a cloud-based machine learning platform facilitates the analysis and new knowledge is obtain, which experts as well need to verify the reliability of prediction obtained.

Machine learning can also be implemented inside Business Intelligence where prediciton must be achieved, and also by using descriptive statistics that tell insights of customer relations. In [18] is suggested the following approaches for identifying customer relations:

- Use linear models for data analysis, which regularly performed in simple ways, and since from linear statistics are implicit numerous assumptions about mutually independence between variables and normally distributed values, those can be helpful for initial stage of exploration.
- Dealing with stochastic distributions, the hidden Markov models (HMM) [19] focus on the analysis of temporal sequences of separate (discrete) states. As well, those are used for creating predictions on time-stamped events.
- When analysing customer satisfaction, the use of Bayesian networks are suggested in [20], which are based on a graphical model representing inputs as nodes with directed associations among them. Nevertheless, because those are developed for academic level and do not provide needed levels of intuition, automation, and integration into corporate environments; accessible Bayesian network software is not suitable, enabling this can create them accessible to business users.

Discussed in [18], customers play a significant role in Smart Manufacturing environments, because of the improvement of customer-business relations and as well the responsiveness of business to take actions in real-time when needed based on customer lifecycle. Since this is not a trivial task that can be implemented overnight using existing business informatics models. Two main factors can be attributed to this[3]: (i) the lack of an automated closed-loop feedback system that can intelligently inform business processes to respond to changes in real-time based on the inputs (for example, data trends, user experience, etc.) received, and (ii) existing analytical tools cannot accurately capture and predict consumer patterns.

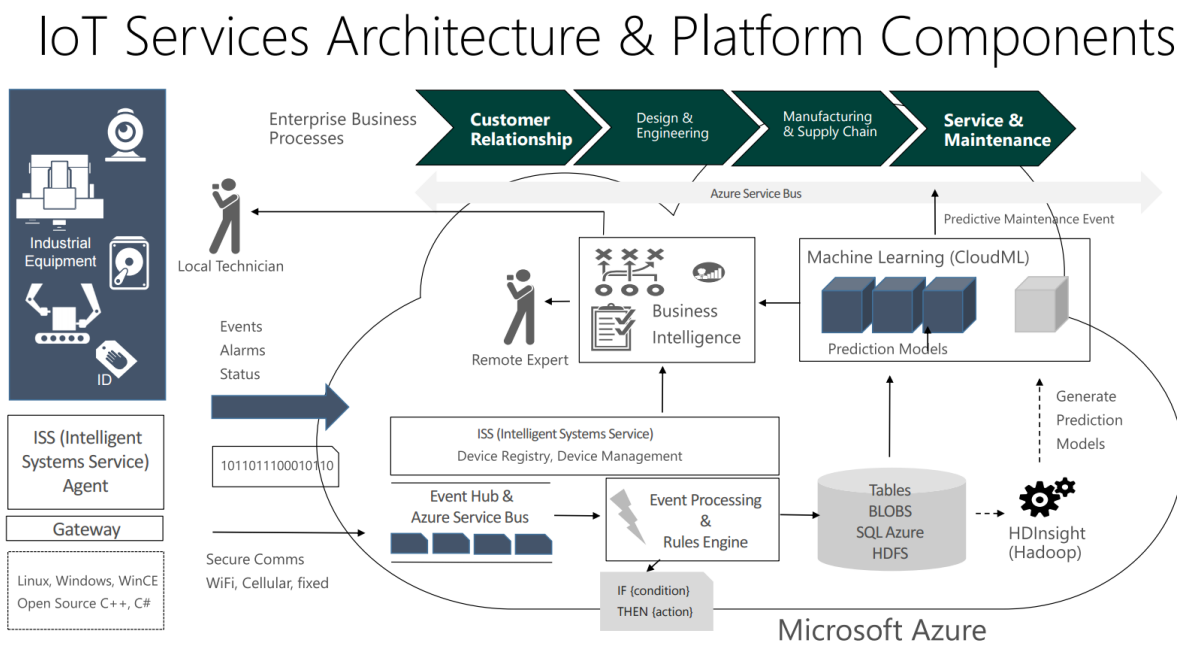



Fig. 3. Architecture for IoT services proposed by Microsoft [17]

The use of digital models is a possible way forward for (i), a digital model able to achieve automation in a closed-loop.

A solution for (ii) when analyzing business contained in data using intelligence should be considered as the use of gathered information into data finally into action. Intelligence in this sense comes from the expert knowledge that can also be integrated in the analysis process, the knowledge-based methods used for analysis, and the new knowledge created and communicated by the analysis process.

The next section presents the used methodology for addressing prediction in customer relations, determining what customers' needs and wants are, and selecting best attributes.

## III. Methodology and Approaches

With all the revised methods and tools from different research, it was determined to use machine learning as unsupervised learning. In specific, it was used fuzzy c-means for clustering and genetic algorithms for selection of best attributes once the fuzzy clustering finished classifying. Following the next sections, the fuzzy c-means is described, together with the Genetic Algorithm (GA) selection. After the tools used, a proposed framework is shown, which integrates the i4 principles for design and manufacture, data analytics, machine learning, Computer Automated Design (CAutoD), among others. With this, the closed-loop for automation can finally close the missing gap for determining customers' needs and wants in order to achieve customized design and processes.

### A. *Fuzzy c-means approach*

Cluster approaches can be applied to datasets that are qualitative (categorical), quantitative (numerical), or a mixture of both. Usually the data (inputs) are observations of some physical process. Each observation consists of $n$ measured variables (features), grouped into an $n - dimensional$ column vector $z_k = \lfloor z_{1k}, \dots, z_{nk} \rfloor^T, z_k \in R^n$ [21].

$N$ Observations set is denoted by $Z = \{z_k | k = 1,2, \dots N\}$, and is represented as a $n \times N$ matrix:

$$Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1N} \\ z_{21} & z_{22} & \cdots & z_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ z_{n1} & z_{n2} & \cdots & z_{nN} \end{pmatrix} \quad (1)$$

Many clustering algorithms have been introduced and clustering techniques can be categorized depending on whether the subsets of the resulting classification are fuzzy or crisp (hard). Hard clustering methods are based on classical set theory and require that an object either does or does not belong to a cluster. Hard clustering means that the data is partitioned into a specified number of mutually exclusive subsets. Fuzzy clustering methods, however, allow the objects to belong to several clusters simultaneously with different degrees of membership [21]. Fuzzy clustering assigns membership degrees between 0 and 1 that indicates their partial membership. Vital for cluster analysis is cluster partition, as well for identification techniques that are based on fuzzy clustering.

Most analytical fuzzy clustering algorithms are based on the optimization of the basic c-means objective function, or some modification of the objective function. The optimization of the c-means functional represents a nonlinear minimization problem, which can be solved by using a variety of methods including iterative minimization [22]. The most popular method is to use the simple Picard iteration through the first-order conditions for stationary points, known as the FCM algorithm. Bezdek [23] has proven the convergence of the FCM algorithm. An optimal c partition is produced iteratively by minimizing the weighted within group sum of squared error objective function:

$$J = \sum_{i=1}^{n} \sum_{j=1}^{c} (u_{ij})^m d^2(y_i, c_j) \quad (2)$$

Where $Y = [y_1, y_2, \dots, y_n]$ is the dataset in a d-dimensional vector space, $n$ is the number of data items, $c$ is the number of

clusters, which is defined by the user. Where $2 \leq c \leq n, u_{ij}$ is the degree of membership of $y_i$ in the $jth$ cluster, $m$ is a weighted exponent on each fuzzy membership, $c_j$ is the center of the cluster $j, d^2(y_i, c_j)$ is a square distance measure between object $y_i$ and cluster $c_j$.

The following steps were used inside Matlab for the fuzzy c-means algorithm:

1) Input ➔ $c$= centroid matrix, $m$= weighted exponent of fuzzy membership, $\in$ = threshold value used as stopping criterion, $Y = [y_1, y_2, \ldots, y_n]$: data
   Output ➔ $c$ = update centroid matrix
2) Randomly start the fuzzy partition matrix $U = [u_{ij}^k]$
3) Repeat
4) Calculate the cluster centres with $U^k$:

$$c_j = \sum_{i=1}^{n} (u_{ij}^k)^m y_i / \sum_{i=1}^{n} (u_{ij}^k)^m \quad (3)$$

Update the membership matrix $U^{k+1}$ using:

$$u_{ij}^{k+1} = 1 / \sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{(m-1)}} \quad (4)$$

Where

$$d_{ij} = \|y_i - c_j\|^2 \quad (5)$$

Until $max_{ij} \|u_{ij}^k - u_{ij}^{k+1}\| < \in$

5) Return $c$

After that, the best attributes are selected using GA toolbox in Matlab. The process is described in Fig. 4.

### B. *Framework for predicting potential customer needs and wants*

In Fig. 5 is depicted the framework proposed to solve several of the afore-mentioned challenges in i4. Based on i4 and Smart Manufacturing key objective, i.e. achieve self-prediction, and self-configurable in order to manufacture products and provide services tailor-made at mass production rates.

In the first block of the proposed framework, customer needs and wants are first captured and processed to extract key design characteristics. These information are then fed into a Computer Automated Design (CAutoD) engine [24] where the design requirements, features and performance objectives are mapped into 'genotypes' for further analyses. This process, which is commonly known as rapid virtual prototyping uses intelligent search algorithms such as the GA or Particle Swarm Optimization (PSO) to explore the design search space for optimal solutions. In the proposed framework, this process takes place over the Cloud and produces a set of an optimized virtual prototype at the end of the search.

The second block of the closed loop in Fig. 5 shows the virtual prototype, which is obtained from the selection and design process in CAutoD Through the integration of CPS or Cyber-Physical Integration (CPI), the virtual prototype in the second block is transformed into a physical product, i.e. the Smart Product as shown in Fig. 5.

The next part of the framework refers to Business Informatics and how the smart products are connected to the IoT. Here is where big data comes in, through the performance of the product and the feedback from the customer, more features can be considered. This covers the necessary attributes for the product to be manufactured in optimal ways.

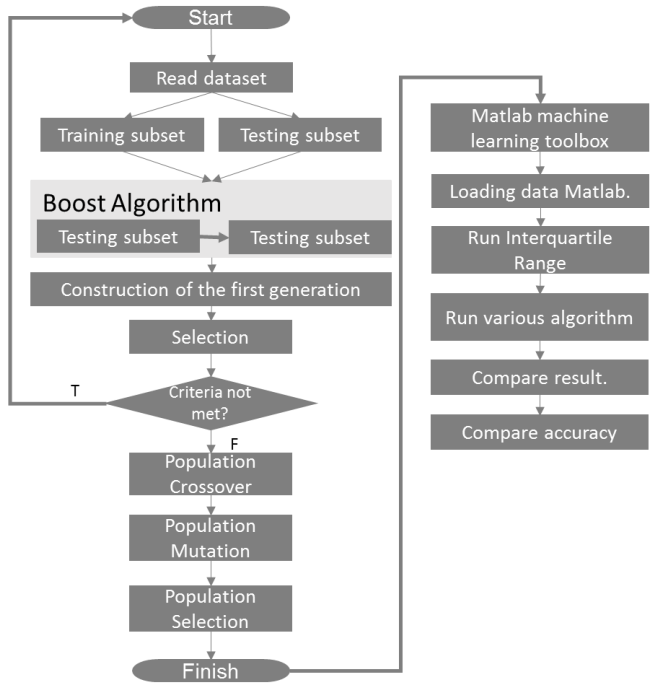



Fig. 4. Genetic search framework using Matlab

Following this, the response obtained from the customer is automatically fed back to the system for further analysis and to fine-tune the virtual prototype. It is necessary to perform the analysis. This analysis is related to prediction, by using node or dynamic analysis that can perform clustering, selection and detection of patterns and visualize it. After that, the fuzzy c-means clustering completes the update of selected attributes by comparing the latest input to the existing cluster and tries to identify one cluster that is most similar to the input sample. Then several features are fed back into the cloud again.

The analysis can result in two outcomes [3]: (i) Similar clusters found. If it is the case, this will be reflected as an existent attribute and the algorithm will update the existing cluster using information from the latest sample. (ii) Non-similar clusters found. The algorithm will hold its operation with the current sample until it sees enough out-of-cluster samples.

When the number of out-of-cluster samples exceeds a certain threshold, it means that there exists a new behaviour in the data that has not been modelled. The algorithm will then create a new cluster to represent the new behavior.

The data that is presented in the following section is used to solve the clustering problem with a fuzzy c-means network designed using the machine learning toolbox in Matlab. Fuzzy c-means are widely used to produce a concise representation of a system's behaviour, by grouping n clusters with every data-point in the dataset belonging to every cluster to a certain degree [22].

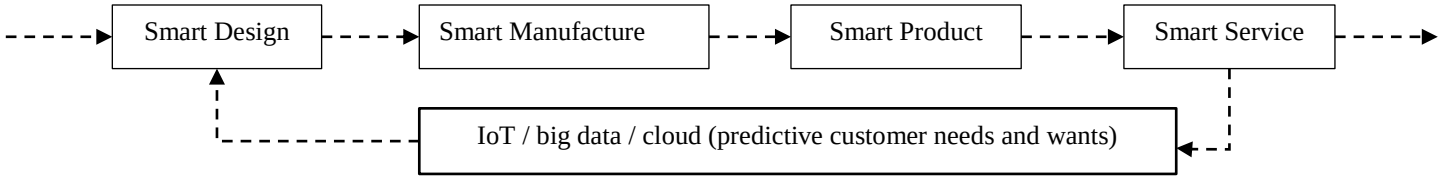



Fig. 5. Industry 4.0 value chain with predictive customer needs and wants fed back for automated customisation [3].

## IV. Case Study

Cluster analysis with fuzzy c-means was performed to the data set found in [25]. This data set consists of three types of entities: (a) the specification of an auto in terms of various characteristics, (b) its assigned insurance risk rating, (c) its normalized losses in use as compared to other cars. The second rating corresponds to the degree to which the auto is riskier than its price indicates. Cars are initially assigned a risk factor symbol associated with its price. Then, if it is more risky (or less), this symbol is adjusted by moving it up (or down) the scale. Actuaries call this process "symbolling". A value of +3 indicates that the auto is risky, -3 that it is probably safer.

The third factor is the relative average loss payment per insured vehicle year. This value is normalized for all autos within a particular size classification (two-door small, station wagons, sports/speciality, etc...), and represents the average loss per car per year.

Database contents are shown in TABLE I.

TABLE I. Automobile Data

| Attribute | Attribute Range | Attribute | Attribute Range |
|---|---|---|---|
| symbolling | -3, -2, -1, 0, 1, 2, 3. | curb-weight: | Continuous from 1488 to 4066. |
| normalized -losses: | Continuous from 65 to 256. | engine-type: | dohc, dohcv, l, ohc, ohcf, ohcv, rotor. |
| make | alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo | num-of-cylinders: | Eight, five, four, six, three, twelve, two. |
| fuel-type | Diesel, gas. | engine-size: | Continuous from 61 to 326. |
| Aspiration | Std, turbo. | fuel-system: | 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi. |
| num-of-doors | Four, two. | bore: | Continuous from 2.54 to 3.94. |
| body-style | Hardtop, wagon, sedan, hatchback, convertible. | stroke: | Continuous from 2.07 to 4.17. |
| drive-wheels | 4wd, fwd, rwd. | compression -ratio: | Continuous from 7 to 23. |
| engine-location | Front, rear. | horsepower: | Continuous from 48 to 288. |
| wheel-base | Continuous from 86.6 120.9. | peak-rpm: | Continuous from 4150 to 6600. |
| Length | Continuous from 141.1 to 208.1. | city-mpg: | Continuous from 13 to 49. |
| Width | Continuous from 60.3 to 72.3. | highway-mpg: | Continuous from 16 to 54. |
| height | Continuous from 47.8 to 59.8. | price: | Continuous from 5118 to 45400. |

This dataset comprises 205 instances, 26 attributes as shown in TABLE I.

The results of the fuzzy c-means are shown in Fig. 6. Here, the partition of the 3 clusters can be noticed. The scatter plot shows the connections between all the instances. From here, Matlab function for fuzzy c-means update the cluster centres and membership grades of each data point, clusters are iteratively moved from the centre to the right location inside the dataset. The selected parameters for the fuzzy c-means were 3 clusters, exponent =3, the maximum of iterations = 100, and minimum improvement= 1e-05. Since iterations are based on minimizing an objective function that represents the distance from any given data point to a cluster centre weighted by that data point's membership grade. Membership function plots obtained are presented in Fig. 7, here for each cluster shows when it reached the maximum of iterations, or when the objective function improvement between two consecutive iterations is less than the minimum amount of improvement specified. Once the clustering was done, it was processed the training data to obtain the attribute classification inside Matlab toolbox for machine learning, were it was as well embedded

parallel routine for speeding up the whole process. Testing with several classifier algorithms, the results are presented in Fig. 8.

All those values colored in green show the corrected classified instances, based on the attribute that best reflected the desired selection: manufacturer or make. The red slots represent the incorrect instances. Here the manufacturer (make) was selected as the predictive variable in order to provide which of the observed brands are more attractive to customers based on all the considered variables.

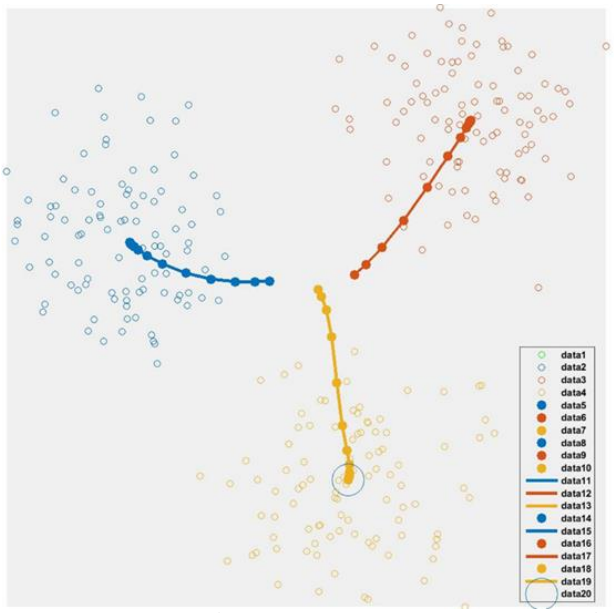
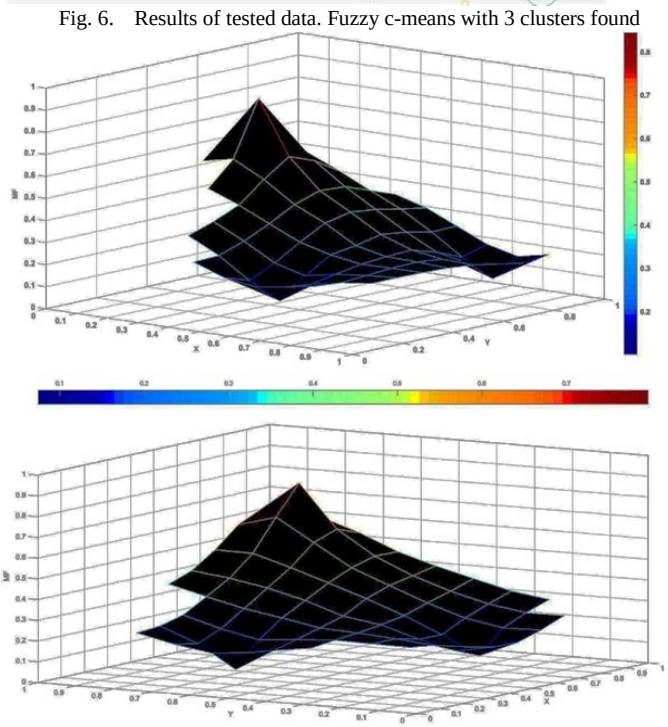



Fig. 6. Results of tested data. Fuzzy c-means with 3 clusters found



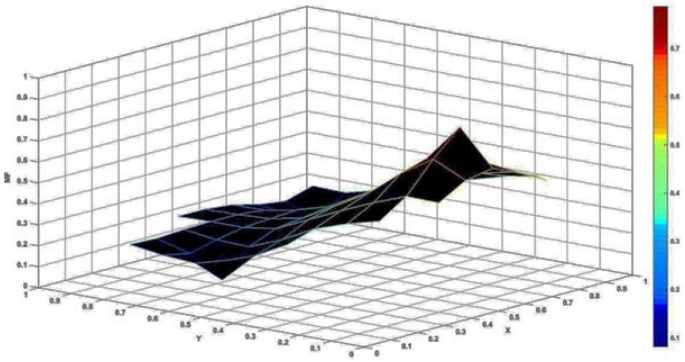


Fig. 7. Membership function. From top to bottom: cluster 1, 2 and 3 results.

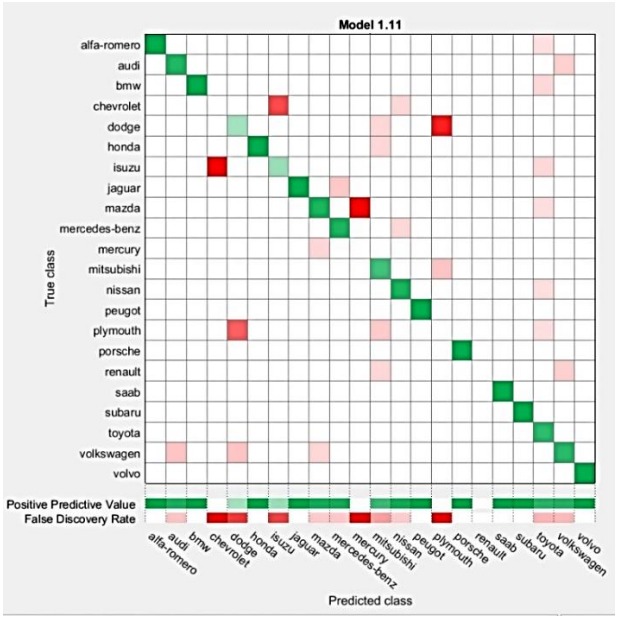



Fig. 8. Confusion matrix obtained for positive predictive values

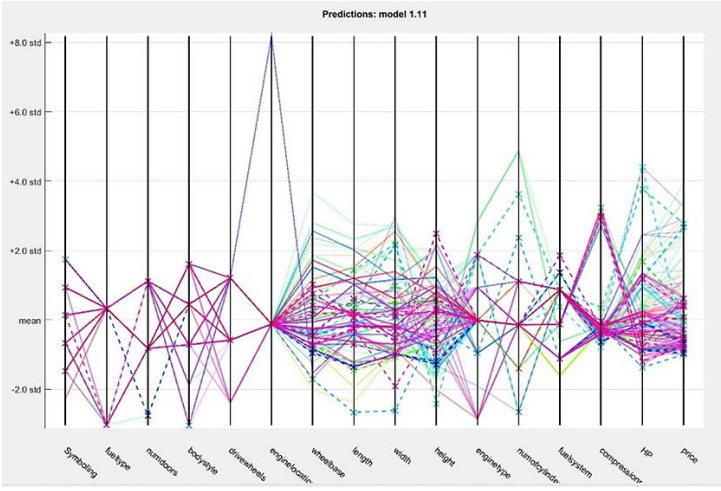



Fig. 9. Paralell coordinates plot for membership functions.

For this plot in Fig. 9 is inferred what type of attributes represent the most corrected classified instances to the predictive model. The selected response variable was the Manufacturer, and each colour represents the brand related to

the predictors (fuel-type, number of doors, body style, engine locations, HP, etc.). For which the strongest relation is found with the engine location, number of cylinders and the HP variables. Moreover, once the attribute selection was performed using the GA selection, it was selected the following instances: num-of-doors, drive-wheels, height, engine-type, num-of-cylinders. Those were performed with a crossover probability of 0.6, a max of generations of 20, mutation probability of 0.033, initial population size of 20, and an initial seed.

## V. Discussion and Conclusion

The use of fuzzy c-means to identify clustering, classify attributes and then select instances using GA search has delivered promising performance. It is found that visualization of results facilitates the analysis in real time. Identification of values for customers' acquisition of a car based on categorical and numerical inputs can be achieved with fuzzy clustering.

Through the development of a predictive tool for mining customers' subconscious needs and wants, selection of best designs can thus be achieved in a smart way. The following features are summarised through the development of this work:

1. In the case study, the results reveal that customer behaviour is based on 5 attributes (number-of-doors, drive-wheels, height, engine-type, number-of-cylinders).
2. Fuzzy c-means has performed a good partition on the dataset and has identified 3 clusters for classification.
3. A feedback design process is suitable for automation with CAutoD.
4. Intelligent search within the design process allows needs and wants to be predictively covered, with virtual prototypes further tuneable by the customer.
5. A CPS interconnected to the designed virtual prototypes would implement customisation efficiently.
6. A smart product may be gauged with business informatics and reliable data constantly, which can be fed back to smart design with IoT in the loop of the i4 value chain.
7. Since the "Internet of Everything (IoE)" facilitates connection through the cloud, it could make it faster to satisfy customer needs and wants.
8. Customer-oriented decision by the manufacturer becomes easier to make, with customer-driven informatics, design and automation.
9. Big data analytics help visualize the influence of product characteristics, clustering and interpretation of subconscious customer needs and wants.

## References


1. Kang, H.S., et al., *Smart manufacturing: Past research, present findings, and future directions.* International Journal of Precision Engineering and Manufacturing-Green Technology, 2016. **3**(1): p. 111-128.
2. Kull, H., *Intelligent Manufacturing Technologies*, in *Mass Customization: Opportunities, Methods, and Challenges for Manufacturers*. 2015, Apress: Berkeley, CA. p. 9-20.
3. Flores Saldivar, A.A., et al. *Self-organizing tool for smart design with predictive customer needs and wants to realize Industry 4.0*. in *World Congress on Computational Intelligence*. 2016. Vancouver, Canada: IEEE.
4. Kull, H., *Introduction*, in *Mass Customization: Opportunities, Methods, and Challenges for Manufacturers*. 2015, Apress: Berkeley, CA. p. 1-6.
5. Möller, D.P.F., *Digital Manufacturing/Industry 4.0*, in *Guide to Computing Fundamentals in Cyber-Physical Systems: Concepts, Design Methods, and Applications*. 2016, Springer International Publishing: Cham. p. 307-375.
6. Yang, B. and N. Burns, *Implications of postponement for the supply chain.* International Journal of Production Research, 2003. **41**(9): p. 2075-2090.
7. Flores Saldivar, A.A., et al. *Identifying Smart Design Attributes for Industry 4.0 Customization Using a Clustering Genetic Algorithm*. in *International Conference on Automation & Computing*. 2016. University of Essex, Colchester city, UK: IEEE.
8. Lee, J., et al., *Recent advances and trends in predictive manufacturing systems in big data environment.* Manufacturing Letters, 2013. **1**(45): p. 38-41.
9. Lee, J., B. Bagheri, and H.-A. Kao, *A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems.* Manufacturing Letters, 2015. **3**: p. 18-23.
10. Mühlhäuser, M., *Smart Products: An Introduction*, in *Constructing Ambient Intelligence: AmI 2007 Workshops Darmstadt, Germany, November 7-10, 2007 Revised Papers*, M. Mühlhäuser, A. Ferscha, and E. Aitenbichler, Editors. 2008, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 158-164.
11. Schmidt, R., et al., *Industry 4.0 - Potentials for Creating Smart Products: Empirical Research Results*, in *Business Information Systems: 18th International Conference, BIS 2015, Poznań, Poland, June 24-26, 2015, Proceedings*, W. Abramowicz, Editor. 2015, Springer International Publishing: Cham. p. 16-27.
12. Nurcan, S. and R. Schmidt, *Introduction to the First International Workshop on Business Process Management and Social Software (BPMS2 2008)*, in *Business Process Management Workshops: BPM 2008 International Workshops, Milano, Italy, September 1-4, 2008. Revised Papers*, D. Ardagna, M. Mecella, and J. Yang, Editors. 2009, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 647-648.
13. LaValle, S., et al., *Big data, analytics and the path from insights to value.*, in *MIT Sloan Management*. 2011, MIT Sloan Management Review: North Hollywood, CA. p. 15.
14. Schmidt, R., et al. *Strategic Alignment of Cloud-Based Architectures for Big Data*. in *2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops*. 2013.
15. Isaacson, W. *The real leadership lessons of Steve Jobs*. Harvard Business Review, 2012. **4**, 92-102.
16. Ji-Hyeong, H. and C. Su-Young. *Consideration of manufacturing data to apply machine learning methods for predictive manufacturing*. in *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*. 2016.
17. Shewchuk, J., *Enabling Manufacturing Transformation in a Connected World*, in *Microsoft Internet of Things*. 2014, Microsoft Corporation: United States. p. 25.
18. Nauck, D., et al., *Predictive Customer Analytics and Real-Time Business Intelligence*, in *Service Chain Management*, C. Voudouris, D. Lesaint, and G. Owusu, Editors. 2008, Springer Berlin Heidelberg. p. 205-214.
19. Rabiner, L. and B.H. Juang, *An introduction to hidden Markov models.* ASSP Magazine, IEEE, 1986. **3**(12): p. 4-16.
20. Heckerman, D. and M.P. Wellman, *Bayesian networks.* Commun. ACM, 1995. **38**(13): p. 27-30.
21. Ludwig, S.A., *MapReduce-based fuzzy c-means clustering algorithm: implementation and scalability.* International Journal of Machine Learning and Cybernetics, 2015. **6**(6): p. 923-934.
22. Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1981: Kluwer Academic Publishers. 256.
23. Bezdek, J.C., *Objective Function Clustering*, in *Pattern Recognition with Fuzzy Objective Function Algorithms*. 1981, Springer US: Boston, MA. p. 43-93.
24. Yun Li, K.H.A., Gregory C.Y. Chong, Wenyuan Feng, Kay Chen Tan, Hiroshi Kashiwagi, *CAutoCSD-Evolutionary Search and Optimisation Enabled Computer Automated Control System Design.* International Journal of Automation and Computing, 2004. **1**(17): p. 76-88.
25. Schlimmer, J.C., *Automonile Data Set*, W.s.A. Yearbook, Editor. 1985, UCI Machine Learning Repository: United States of America.