Exploring User Capability Data with Topological Data Analysis

U. Persad, J. Goodman-Deane, P. M. Langdon and P. J. Clarkson

Abstract This paper presents an analysis of user capability data using Topological Data Analysis (TDA) (unsupervised machine learning) to extract insight. The aim was to explore the global shape and sub-groupings (clusters of profiles) of people using data collected from the Cambridge Better Design Pilot Study of 362 people from across England and Wales. The resulting topological network demonstrated the global shape of the sample and distribution of sensory, cognitive and motor capability across the sample. The TDA network was automatically grouped into 14 distinct clusters, and distinguishing features of each cluster was extracted. The results demonstrate the value of applying TDA to analyse and visualise user capability data, and it is proposed that the cluster descriptions could be used for developing empirically based design tools such as personas for Inclusive Design.

1 Introduction

Inclusive Design is becoming more important with the ageing of the world's population and improvements in medical care. Designers are required to respond to this population shift by executing an Inclusive Design process to produce practical inclusive consumer products across a range of sectors. To enact this approach, design and manufacturing business practices are required to become more people

U. Persad (✉)
The University of Trinidad and Tobago, Arima, Trinidad and Tobago
e-mail: umesh.persad@utt.edu.tt

J. Goodman-Deane · P. M. Langdon · P. J. Clarkson
Cambridge Engineering Design Centre, the University of Cambridge, Cambridge, UK
e-mail: jag76@cam.ac.uk

P. M. Langdon
e-mail: pml24@eng.cam.ac.uk

P. J. Clarkson
e-mail: pjc10@eng.cam.ac.uk

P. Langdon et al. (eds.), *Breaking Down Barriers*,
https://doi.org/10.1007/978-3-319-75028-6_4

and population aware to accommodate the mainstream approach that Inclusive Design advocates.

However, there remains a need for a better understanding of how data on human capability variation across populations can support the inclusive approach (Johnson et al. 2010). Current user data is fragmented and lacking (Johnson et al. 2010), and though designers and researchers have made the best use of such data for product evaluation (Waller et al. 2010), the lack of integrated data with proven predictive value continues to plague the field (Persad et al. 2011; Tenneti et al. 2013). In addition, the databases on user capability require transformation into visual representations that are easy to understand and use. To this end, this paper presents an exploratory study using a relatively new unsupervised machine learning technique termed Topological Data Analysis (TDA) on a recent pilot study of user capabilities across the UK population (Tenneti et al. 2013).

2 Background

Previous work in this area explored the underlying structure of disability data via hierarchical cluster analysis (Langdon et al. 2006). By using numerical methods of classification, it is possible to extract the underlying structure in data without any prior assumptions. However, the resulting clusters and interpretations were found to be difficult for practitioners to understand (Langdon et al. 2006). Given recent progress in machine learning and big data analytics, a new method has emerged to understand the global structure in datasets. This method combines the mathematical field of Topology with Machine Learning and Visualisation resulting in TDA (Carlsson 2009; Lum et al. 2013).

In essence, the TDA method is built on the principle that data has a multidimensional shape, and this shape conveys meaning. Fundamentally, TDA is a geometric method to detect patterns and shapes within the data. By recognising these shapes and patterns in the data, important features and groupings could be identified. Lum et al. (2013) describe three key ideas of topology that make extracting of patterns via shape possible. First, TDA defines a metric space between all multidimensional points in a dataset, i.e. the ‘distance’ between any pair of points. Since this is a coordinate-free way of defining the data, the TDA depends only on the distance function that specifies the shape. Second, TDA shapes and representation are invariant under small deformations. Third, TDA generates a compressed representation of the shape of the data using a simplicial complex or network. Shapes, such as circular segments (loops) and linear segments (flares), appear in the data visualisation leading to new insight.

The advantage of TDA is that it can detect patterns missed by traditional multidimensional methods, such as PCA, MDS and cluster analysis. Specifically, clustering methods produce several distinct and unrelated groups without clearly showing how these groups relate to each other. Therefore, TDA provides a new tool in the data science toolbox for understanding multidimensional datasets as found in the field of ergonomics/human factors. The study presented in this paper uses TDA to explore the Better Design pilot survey data of 362 people from across England and Wales (Tenneti et al. 2013).

3 Methodology

The dataset of 362 people contained capability variables describing the age, gender, vision, hearing, cognition and motor function of participants. Thirty nine (39) variables were selected for inclusion fulfilling the assumption that the measures were ratio or interval level data in order to be compatible with the TDA analysis. Apart from age and gender, the other variables selected were as follows:

Sensory Variables: Near-vision comfort (high contrast): majority of the day setup; near-vision comfort (low contrast): majority of the day setup; distance vision, distance vision comfort: majority of the day setup; distance vision comfort: general setup (distance aid if participants did the distance aid test, majority setup otherwise), hearing at different volumes (no background noise) and hearing at medium volume at different levels of background noise.

Motor Variables: Moberg test results: right hand, Moberg test results: left hand, grip strength: comfort non-dominant hand, grip strength: comfort dominant hand, grip strength: threshold non-dominant hand, grip strength: threshold dominant hand, getting out of a chair (with arms), getting out of a chair (without arms), reaching floor level, out in front (Left arm), out in front (Right arm), above head (Left arm) and above head (Right arm).

Cognitive Variables: Immediate recall memory, delayed recall memory, number of letters scanned and search efficiency (executive function measures), literacy: number of correct answers, numeracy: number of correct answers, perseverance when things go wrong, ability to find a solution when confronted with a problem, confidence in learning to use technology products, anxiety about new technology products, experience with the following—make calls on a mobile phone, send text message on a mobile phone, take pictures with a digital camera or phone, use a remote control for digital TV, use the Internet, listen to MP3 tracks on a portable device, use a gaming console, such as XBOX, playstation or Wii and use satellite navigation, like a tom-tom.

The data was imported and analysed in Ayasdi platform, a TDA software tool for analysis and visualisation (AYASDI 2017). This TDA method requires no prior assumptions allowing the data to speak for itself.

The first step in the analysis was to select an appropriate metric for the data that could account for missing values and deal with continuous variables that measure different phenomena. Only variables with interval or ratio level data were used in the analysis. A norm angle metric was selected where the procedure first normalises all capability variables in the dataset to have a mean of 0 and a variance of 1 (making the variables comparable). The norm angle distance is then calculated as the angle distance between the mean-centred, variance-normalised points.

This metric handles nulls by projecting the pair of rows to the intersection of their non-null columns.

The second step in the analysis was to select an appropriate lens for the data. A lens is a filter that converts the dataset into a vector, where each row in the original dataset contributes to a real number in the vector turning every row into a single number. Neighbourhood Lenses 1 and 2 were selected. These lenses generate an embedding of high-dimensional data into a two-dimensional plane by embedding a k-nearest neighbours graph of the data using Ayasdi's proprietary graph layout algorithm. These lenses work to emphasise the metric structure of the data. The software then used a mapping and clustering algorithm to group people into connected clusters (nodes) producing TDA network visualisations. For exploratory data analysis, the resolution and the gain of the lenses were varied to produce networks of varying levels of detail. These networks could be coloured by any variable in the dataset.

For clustering the network, the Community auto-grouping algorithm was used via the provided Python SDK. This network algorithm, based on Louvain modularity optimisation, operates on the topological model's graph structure. It tries to find the best grouping of nodes that have high intragroup connectivity and low intergroup connectivity, resulting in highly connected clusters. The clusters were compared to the rest of the dataset using comparison tests of P value and KS scores (Kolmogorov–Smirnov tests) to determine which variables differentiated each cluster (with $p < 0.05$ representing significant variables on the KS test). Cases in each of the 14 clusters were exported out of the AYASDI platform software and further analysed in MS Excel and JASP for descriptive statistics. In the next section, these results are presented.

4 Results

Figure 1 shows the TDA network produced from the Better Design data. The overall shape of the network indicates a structure similar to a neuron with a core at the left end with three small protruding 'flares' and a fourth large protruding 'flare' out to the right end with smaller flares protruding from it. Some nodes were also not included in the main structure seen as singletons above the main structure. The network is coloured by 'Rows per Node', which translates to the number of people grouped in a node. The main structure on the left contains the most people in each node shown with the red and yellow colouring. Flares (i.e. sub-groupings of interest) therefore contain fewer individuals than the inner core on the left.

Figure 2 shows 14 clusters resulting from the application of the Community Algorithm on the TDA network structure. For ease of interpretation, these clusters have been grouped into three categories with average age less than 40 years, average age 40–60 years and average age greater than 60 years. These are shown in Figs. 3, 4 and 5, respectively, with qualitative cluster descriptions.

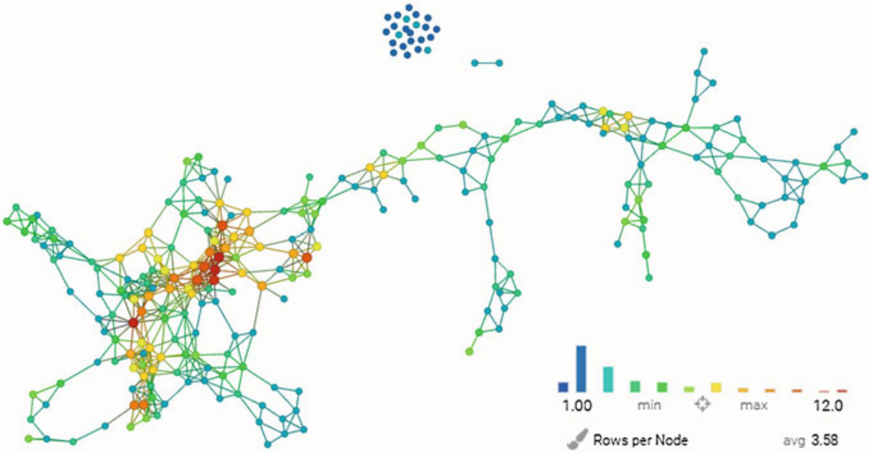


Fig. 1 TDA network coloured by rows per node. Metric: norm angle, Lens 1: neighbourhood Lens 1 (res: 30, gain: 2.5), neighbourhood Lens 2 (res: 30, gain: 2.5)

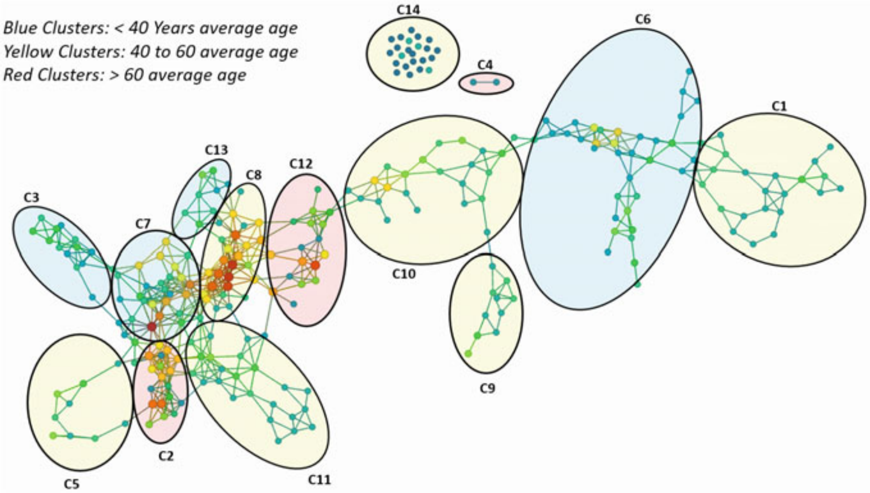
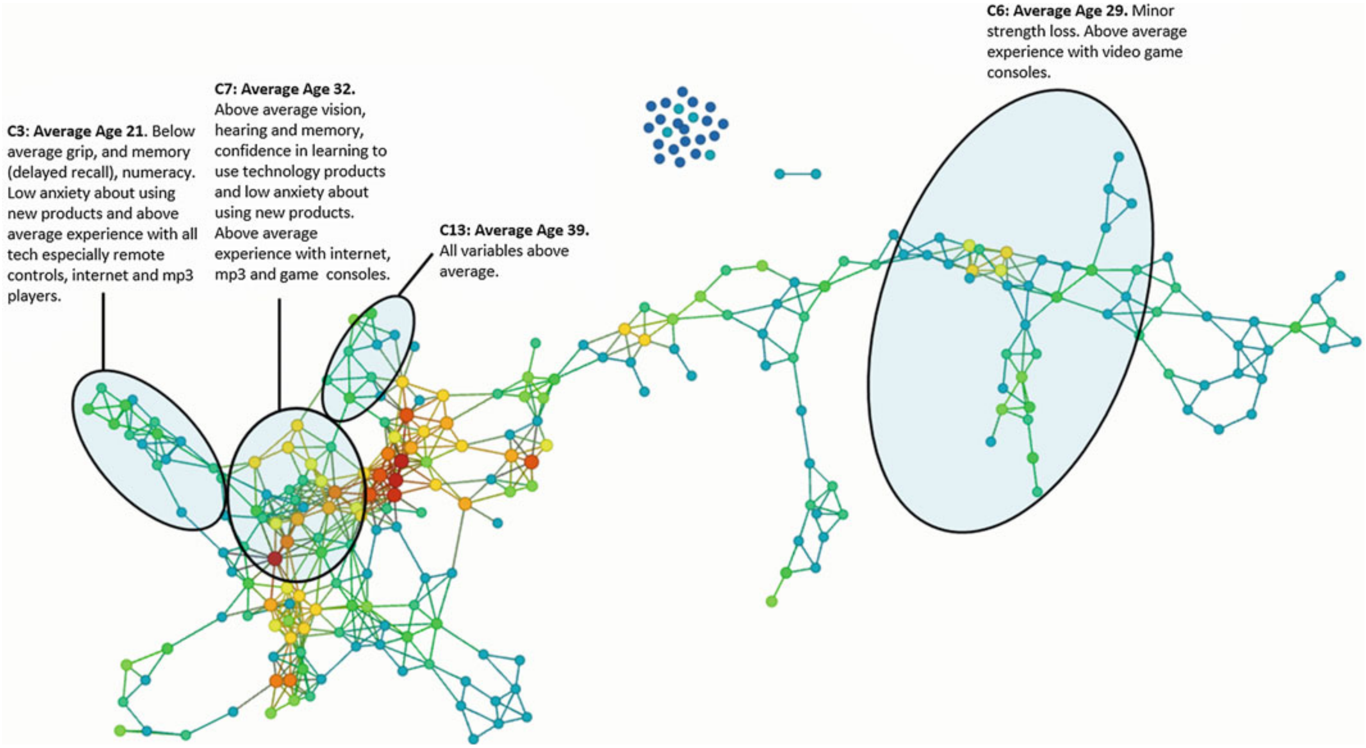

Blue Clusters: < 40 Years average age
Yellow Clusters: 40 to 60 average age
Red Clusters: > 60 average age
C14
C4
C6
C1
C13
C8
C12
C3
C7
C10
C9
C5
C2
C11

Fig. 2 TDA clustering of topological network of the Better Design data using auto-grouping (community algorithm) showing the resulting 14 distinguishing clusters in three broad age groups: <40 years, 40–60 years and >60 years

Four clusters are shown in Fig. 3 for the less than 40 years age group. In this age group, people showed sample average and above scores or experience with digital technology. However, clusters 3 and 7 show some minor capability loss in cognitive and motor capabilities. Seven clusters are shown in Fig. 4 for the 40–60 years age group. In this age group, the clusters ranged from above sample



C3: Average Age 21. Below average grip, and memory (delayed recall), numeracy. Low anxiety about using new products and above average experience with all tech especially remote controls, internet and mp3 players.

C7: Average Age 32. Above average vision, hearing and memory, confidence in learning to use technology products and low anxiety about using new products. Above average experience with internet, mp3 and game consoles.

C13: Average Age 39. All variables above average.

C6: Average Age 29. Minor strength loss. Above average experience with video game consoles.

Fig. 3 Cluster descriptions for the <40 years age group

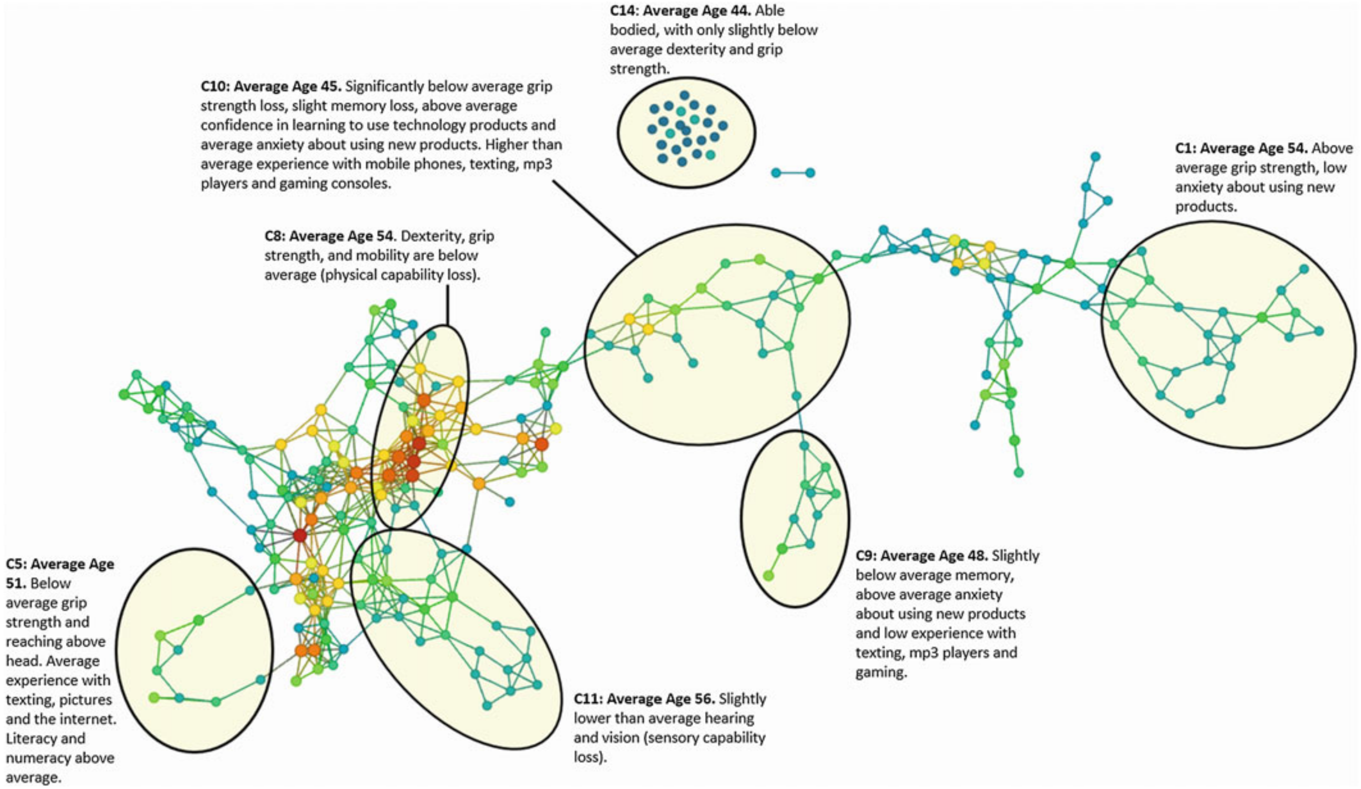
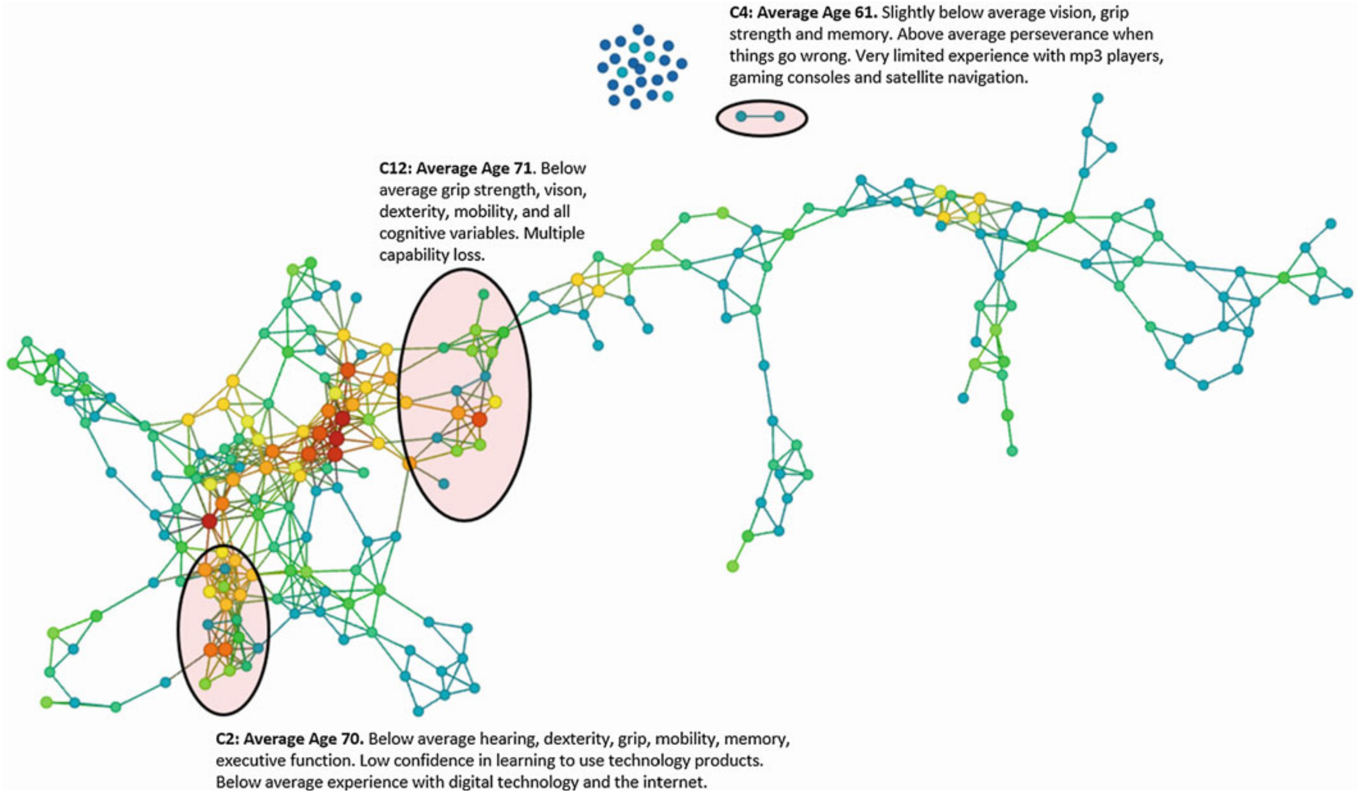

C14: Average Age 44. Able bodied, with only slightly below average dexterity and grip strength.
C10: Average Age 45. Significantly below average grip strength loss, slight memory loss, above average confidence in learning to use technology products and average anxiety about using new products. Higher than average experience with mobile phones, texting, mp3 players and gaming consoles.
C1: Average Age 54. Above average grip strength, low anxiety about using new products.
C8: Average Age 54. Dexterity, grip strength, and mobility are below average (physical capability loss).
C5: Average Age 51. Below average grip strength and reaching above head. Average experience with texting, pictures and the internet. Literacy and numeracy above average.
C9: Average Age 48. Slightly below average memory, above average anxiety about using new products and low experience with texting, mp3 players and gaming.
C11: Average Age 56. Slightly lower than average hearing and vision (sensory capability loss).

Fig. 4 Cluster descriptions for the 40–60 year age group



C4: Average Age 61. Slightly below average vision, grip strength and memory. Above average perseverance when things go wrong. Very limited experience with mp3 players, gaming consoles and satellite navigation.

C12: Average Age 71. Below average grip strength, vison, dexterity, mobility, and all cognitive variables. Multiple capability loss.

C2: Average Age 70. Below average hearing, dexterity, grip, mobility, memory, executive function. Low confidence in learning to use technology products. Below average experience with digital technology and the internet.

Fig. 5 Cluster descriptions for the >60 years age group

average sensory, cognitive and motor capability (cluster 1), through the minor capability losses (clusters 9 and 14), to more severe combinations of capability losses (Clusters 10, 8, 11 and 5). Three clusters are shown in Fig. 5 for the greater than 60 years age group. Clusters 2 and 12 highlight moderate to severe multiple capability losses compared to the sample averages, coupled with very limited experience with digital technology and confidence in using new products. Cluster 4, however, shows a tech-savvy subgroup that perseveres with new technology even though they have limited experience and multiple minor capability losses.

5 Discussion and Conclusion

The results presented summarised the Better Design study data in terms of TDA networks and 14 clusters. These clusters provide evidence of the structure of capability distribution in populations. The advantage of the Better Design dataset is that it contained multiple measures across capability domains resulting in rich descriptions of each cluster.

In each age group, there was a spread of capability from single minor capability loss to multiple capability loss. In addition, the attitudes and experience of people in each age group can provide designers with the data that they need to create designs that are usable, accessible and easy to learn. The variation exhibited by the data underscores the importance of Inclusive Design approaches when designing for the wider population. Supporting and capturing the richness of user diversity in design approaches, methods and tools will become more important as populations age and healthcare improvements enable longer life.

The results demonstrate the usefulness of the TDA approach using machine learning and network visualisation to explore and extract insight from user capability data. The data science and machine learning approach show promise for application in future ergonomics/human factors studies that capture large multivariate datasets. The Better Design pilot study points the way to future large-scale data collection efforts with multiple sensory, cognitive and motor variables. Given that analysis and visualisation tools such as TDA will make it easier to see the global structures inherent in data, it will encourage a move to methodologies that allow the data to ‘speak for itself’ and build new theoretical and practical insights.

So and Joo (2017) demonstrate that creativity in the design process could be improved through the use of personas. Personas capture qualitative details of key users that allow designers to focus on designing for ‘real’ people rather than a nebulous group (Goodman-Deane et al. 2010, 2014). The cluster information provided in this paper could add a quantitative dimension to the creation of personas by integrating sensory, cognitive and motor capability values in persona descriptions. This data-driven approach could ensure that designers account for the full range of user capabilities while engaging in Inclusive Design. It could also support market segmentation (Goodman-Deane et al. 2010, 2014).

Further work on the Better Design data will focus on relating user capabilities to rated product difficulties with an eye to developing predictive models for analytical product evaluation. In this endeavour, TDA will also play a major role.

References

AYASDI (2017) Ayasdi platform. www.ayasdi.com. Accessed on 15 Sept 2017

Carlsson G (2009) Topology and data. Bull Am Math Soc 46(2):255–308

Goodman-Deane J, Langdon PM, Clarkson PJ (2010) Key influences on the user-centred design process. J Eng Des 21:345–373

Goodman-Deane J, Ward J, Hosking I, Clarkson PJ (2014) A comparison of methods currently used in inclusive design. Appl Ergon 45:886–894

Johnson D, Clarkson PJ, Huppert F (2010) Capability measurement for inclusive design. J Eng Des 21:275–288

Langdon PM, Persad U, Clarkson PJ (2006) Developing a model of capability for inclusive design: the hidden structure of the ONS disability data. In: Disability Studies Association conference 2006: disability studies: research and learning, Lancaster University, Lancaster, UK

Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, Carlsson G (2013) Extracting insights from the shape of complex data using topology. Sci Rep 3:1236

Persad U, Langdon PM, Clarkson PJ (2011) Investigating the relationships between user capabilities and product demands for older and disabled users. HCI International 2011, Springer, Orlando, FL, US, 9–14 July 2011

So C, Joo J (2017) Does a persona improve creativity? Des J 20(4):459–475

Tenneti R, Goodman-Deane J, Langdon P, Waller S, Ruggeri K, Clarkson PJ, Huppert HA (2013) Design and delivery of a national pilot survey of capabilities. Int J Hum Factors Ergon 2:281–305

Waller SD, Langdon PM, Clarkson PJ (2010) Using disability data to estimate design exclusion. Univ Access Inf Soc 9:195–207