Title:
Data-driven approach toward identifying CAD user archetypes

Authors:
Robert Celjak, rc216951@stud.fsb.hr, University of Zagreb, FMENA
Nikola Horvat, nikola.horvat@fsb.hr, University of Zagreb, FMENA
Stanko Škec, stanko.skec@fsb.hr, University of Zagreb, FMENA
Tomislav Martinec, tomislav.martinec@fsb.hr, University of Zagreb, FMENA

Keywords: Computer-aided design (CAD), Collaborative CAD, Clustering, Data mining, CAD log data

DOI: 10.14733/cadconfP.2023.302-306

Introduction and related work:
Computer-Aided Design (CAD) models are essential in the design process [1], as they allow designers to create, modify, and test designs virtually. In addition, they also simplify the process of communicating the current work with other designers [2]. With the emergence of cloud computing and a growing industry trend of collaborative design tools (e.g., *Onshape*) usage, CAD models can now be synchronously updated by different designers [2]. Although this transition provides new opportunities in the domain of CAD modeling (e.g., working on the same model in the same CAD environment in real-time), it is not yet clear how to fully utilize the capabilities offered by the new technologies. Hence, it is essential to examine in more detail how exactly designers utilize a synchronous collaborative CAD environment. Insights from such an analysis would then allow for a better understanding of the CAD environment's effect on the underlying design processes of individuals and teams, as well as on some of the design metrics (e.g., design quality, user productivity). Therefore, the overarching aim would be to provide a better understanding and best practice guidelines for the use of synchronous collaborative CAD environments.

CAD modeling is not a single task, but rather a unique problem-solving situation in which designers iteratively engage. The usage of collaborative tools provides a basis for a better understanding of how designers approach this task in the context of CAD actions performed and which design approaches they utilize. Therefore, identifying designers' behavior patterns, i.e., designer archetypes, is essential as it provides insight into the characteristics, tendencies, and capabilities of a CAD tool user. As different designers adopt different design approaches or possess diverse skillsets [3], it is crucial to compose design teams in a way that team members (TMs) complement each other. In the case of redundant skillsets or similar design approaches of different TMs, the quality of the overall design may not be sufficient and a decrease in the productivity of a designer may occur. To avoid the latter from happening, Stone et al. [4] set out to establish a method to determine the optimal number of designers on a part CAD model based on the complexity of it and its features. Teams varied in size from one to four TMs, however, no statistically significant patterns were found. Furthermore, Deng et al. [2] conducted research where they tracked CAD actions performed by design teams composed of experienced or/and novice designers during a design challenge and compared their CAD models. They found that teams of experienced designers tend to spend a greater proportion of their actions in Assemblies and tend to perform more iterative design processes. The authors also propose a metric for determining how iterative a design process is in the form of a ratio between creation and revision actions performed on the models. Johnson et al. [5] conducted similar research and compared CAD models generated by teams of engineers and students working on an identical design task. They found that a team of engineers performed a greater number of CAD actions and that they generated CAD models of better quality in the context of reusability for further adjustments. Rahman et al. [3] expand on this as they developed a framework for clustering designers with similar sequential design patterns. They clustered CAD users based on their behavioral patterns and found that these patterns can lead to overall designs of similar quality, where quality is determined by product usability and profitability.

Therefore, the analytical CAD actions data gathered from design teams during modeling sessions offer great potential to enhance the current understanding of designer behavior within the CAD environment. Although the aforementioned studies did provide some insight into characterizing and clustering of user types, a gap in the research was identified as CAD user archetypes in the context of their role in the design team have still not been determined. This paper thus aims to explore the potential of tracking CAD actions during a design course to identify different CAD user archetypes within design teams.

Research methods:

This study was set up to identify different archetypes of CAD users during a project-based design course. This was enabled by tracking CAD actions performed by students in a collaborative CAD environment. The main goal of the design course was to familiarize students with CAD modeling software (*Onshape*) in the context of generating a functional CAD model. After introducing the students to the capabilities, workspace, and functions of *Onshape* and teaching them CAD methodology, 42 undergraduate engineering students were divided into 14 teams of three. They were assigned design tasks in the form of patent sketches which served as a basis for modeling a functional CAD model of the product. The final CAD assemblies, shown in Fig. 1, were reviewed by teaching assistants and senior engineers from industrial companies. The review involved examining if the CAD model fulfills the task requirements and whether the CAD model is suitable for manufacturing. Teams were graded based on the quality of their CAD models in the context of adhering to the task's requirements. The points they were evaluated with a range from 28.67 to 44, out of possible 50.

Fig. 1: Final CAD models generated by design teams.

The usage of *Onshape* during the project-based design course enabled non-intrusive data collection due to its capability of capturing CAD actions performed by users within the modeling workspace and storing them in an audit trail. In addition to actions performed by users, actions automatically executed by the software itself (e.g., *Update Metadata*, *Content update*, etc.) were also stored in the audit trail. Data was gathered and stored in a structured way adequate for data analysis. In this case, the gathered dataset consists of the action name, the timestamp when the action was performed, the document and the tab in which the action was performed, and the designer who performed the action. As the paper focuses only on designers' actions, the automatically generated actions were excluded from the analysis. Moreover, redundant actions were identified, when *OnShape* registers a single action (e.g. *Insert Sketch*) as four unique actions (*Add part studio feature*, *Commit add* or *edit of part studio feature*, *Insert feature: Sketch*, *Add or modify a sketch*). Redundant actions were therefore excluded from the analyzed dataset, to avoid their effect on data analysis. Finally, given that the emphasis of the study was put on the modeling of parts and assemblies, the actions related to the generation of technical documentation were also excluded. The remaining actions were then classified using an adopted classification of CAD actions proposed in previous work [6]. To streamline the data analysis process, irrelevant actions from the adopted classification were excluded (e.g., *Organizing* and *Viewing* CAD action classes). The final classification, which is depicted in Fig. 2, comprises only those actions that fall within the scope of the metrics defined in the succeeding subsection. The classification enabled for the data analysis to be conducted on different levels of granularity, i.e., class, subclass, or action level. The modified classification includes four general classes of CAD actions: *Creating*, *Editing*, *Deleting,* and *Reversing*. *Creating*, *Editing*, and *Deleting* were further segmented on a *Part* and *Assembly* level, while *Editing*, *Deleting,* and *Reversing* were consolidated into a *Revising* class for the purposes of the *Creation* to *Revision* metric adopted by Deng et al. [2].

Creating		Revising				
		Editing		Deleting		Reversing
Part	Assembly	Part	Assembly	Part	Assembly	
Add part studio feature Copy paste sketch	Add assembly instance Add assembly feature Linked document insert Paste: instance	Start edit of part studio feature Move part	Start edit of assembly feature Set mate values Configure suppression state Start assembly Move to origin Load named position Replace part Fix part Unfix part Suppress part Unsuppress part	Delete part studio feature	Delete assembly feature Delete assembly instance	Cancel Operation Reset mates to initial positions Restore Previous

Fig. 2: Classification of CAD actions [6].

Analysis of design teams and their members was conducted by using the audit trail data about 74055 CAD actions performed by 14 student teams during the design course and clustering similar CAD users based on several metrics:

- *User action percentage* - the percentage of performed CAD actions by a user in relation to the team total.
- *Creation/revision ratio* - the ratio of *creating* to *revising* classes of CAD actions performed which shows how much new geometry was generated by the user versus how much of existing geometry the user modified.
- *Part contribution* - the ratio of using the *Part* subclass of CAD actions to the total usage of *Part* and *Assembly* subclasses of an individual user shows how much they contribute in the context of generating part geometry. This metric implicitly provides an insight into the usage of the *Assembly* class of CAD actions (*Assembly contribution* = 1 – *Part contribution*).

Data analysis was conducted using the Python programming language and its libraries. This is followed by using the *elbow method* to determine the optimal number of clusters in a *k-means* clustering algorithm by plotting the sum of squared distances (WCSS) between the data points and their cluster centroid. From the plot, it can be observed that the WCSS starts to decrease at a slower rate and forms an "elbow" shape on the plot, indicating that adding more clusters will not significantly decrease the WCSS. This point is considered the optimal number of clusters because adding more clusters will not result in a significant decrease in WCSS, but rather result in an increase in the complexity of the model. Finally, the results were laid out in two 2D scatter plots in which the TMs were clustered based on *User action percentage* in both plots, *Creation/revision ratio* in one graph and *Part contribution* in the other graph. Moreover, TMs are labeled according to their ranking based on the number of points they were evaluated with for their designs by senior engineers from industrial companies. This is used as an additional lens that might provide insights into differences and similarities of team compositions of high- and low-evaluated teams. The labels are shown in the following format: *Team x* represents the team rank among all teams, whereas, in the case of an identical number of points, teams were labeled as *Team xa* and *Team xb*.

Empirical study results and discussion:

The results are structured as follows. Users were clustered in a 2D scatter plot based on *User action percentage* and *Part contribution*. Furthermore, the latter metric was replaced by the *Creation/revision ratio* and was also used to show user clusters in a form of a 2D scatter plot. The elbow method suggested three clusters for both graphs. Furthermore, both the *Within-Cluster Sum of Squares* and the *Silhouette Score* method indicate favorable values in the context of the results of the *k-means* clustering algorithm. The interpretation of obtained clusters does not imply their complete discrete distribution but rather describes behaviors typical for different user archetypes. Hence, the interpretation is based on the data points distanced further from the center of the plot as those closer to the center of the plot are similar regarding certain metric values. The first graph, depicted in Fig. 3 (a), shows clusters of similar users in terms of *User action percentage* and *Part contribution* metrics. Clusters are characterized by the following values in the context of *User action percentage*. *Cluster 1* consists of users who registered above-average values of 40 to 75 %, while clusters 2 and 3 consist of users who performed below average, with values of 12 to 42 % and 6 to 35%. As for the *Part contribution*, Cluster 1 consists of average users who generated 47 to 77 % of geometry related to parts, Cluster 2 consists of below-average to average users who generated 30 to 60 %, while Cluster 3 consists of above-average users who generated 64 to 98 %. By analyzing the clusters, it may be implied that the role of TMs in *Cluster 3* was to generate parts, while the role of TMs in *Cluster 2* was to assemble these parts. TMs in *Cluster 1* show versatility as they register a rather high percentage of *User action percentage* and similar percentages of *Part contribution* and *Assembly contribution* metrics. Out of 14 teams, only one team had all their TMs in the same cluster (*Team 4a – Cluster 2*). This suggests that TMs of Team 4a have not agreed on specific roles in the context of part and assembly modeling, but rather each TM equally performed both part and assembly CAD actions. Furthermore, 8 teams have two TMs clustered together, out of which 5 teams have two TMs in *Cluster 3* and one TM in *Cluster 1*. Out of these 5 teams (1a, 2, 5a, 9 and 11), two of them are within the top three evaluated teams and two of them are within the bottom three evaluated teams. Finally, the remaining 5 teams have all their TMs clustered in different clusters. In the case of these 5 teams, the lowest evaluated team ranks 9th.

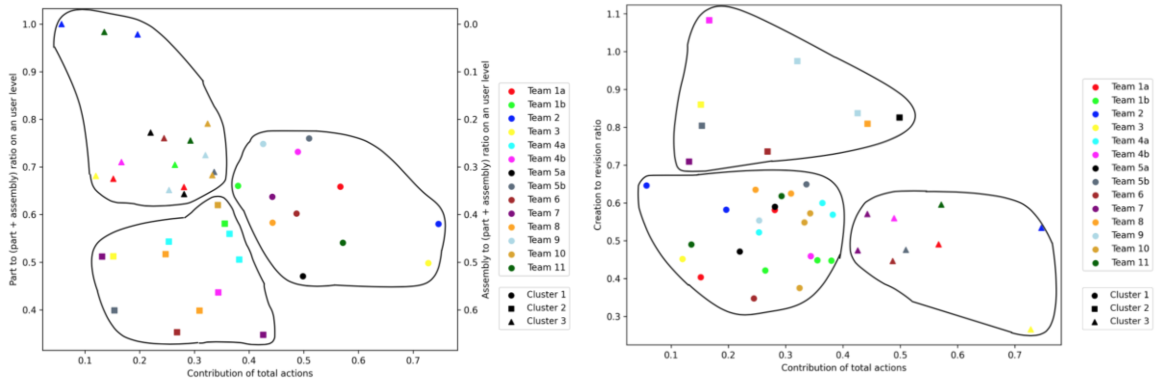

Fig. 3: (a) 2D scatter plot of clusters of designers in the context of part and assembly actions performed (b) 2D scatter plot of clusters of designers in the context of *creation* to *revision* ratio.

The second graph, shown in Fig. 3 (b), depicts user clusters based on *User action percentage* and the *Creation/revision ratio* metrics. *Cluster 1* encompasses TMs whose *User action percentage* ranges from 6 to 40% and whose *Creation/revision ratio* ranges from 0.35 to 0.65. Furthermore, *Cluster 2* is characterized by a slightly greater *User action percentage* with a range of 12 to 50% and a *Creation/revision ratio* from 0.60 to 1.10. Finally, *Cluster 3* encompasses TMs with the most *User action percentage* (42 to 75%) and 0.3 to 0.6 *Creation/revision ratio*. The most common team composition (4 teams) in the context of this graph is the one in which every TM is clustered within a different cluster. Furthermore, it was observed that in half of the teams, one user performs less than 17% of the team's total actions. Out of these 7 teams, in 6 of them that TM tends to also have the highest *Creation/revision ratio*. This suggests that those TMs, clustered in the left part of *Cluster 2*, use a sequential design approach and are either experienced CAD users which know the most straightforward path to completing the task or their work was cursory, i.e., their contribution was too low to have a significant effect on the overall design. The latter scenario is more likely as Deng et al. [2] found that more experienced users tend to register lower values of the *creation* to *revision* ratio. Moreover, this is supported by Johnson et al. [5] who found that experienced users generate more CAD actions and overall CAD models of good quality.

In the context of clustering users based on the percentages of part and assembly classes of CAD actions, the results show that the majority of teams adopt one of two compositions. One composition includes having one TM performing the majority of part CAD actions, one TM performing the majority of assembly CAD actions, and one versatile TM which covers both domains and tends to perform the most CAD actions within the team. This type of composition is characteristic of average and high-evaluated teams in the context of their design quality. The second composition is similar to the first, except for the "assembly specialist TM" being replaced by another "part specialist TM". This type of composition was found in both high- and low-evaluated teams. Furthermore, out of 4 high-evaluated teams, 2 teams adopt one design approach, and 2 teams adopt the other one. Average and low-evaluated teams also show differences in team composition between other teams in the same rank. This finding is aligned with the observation made by Rahman et al. [3] who stated that different design teams adopt different design approaches but still generate overall CAD models of similar quality. Furthermore, a dominant TM in the context of *User action percentage* was identified in 7 out of 14 teams (Δ (dominant TM vs other TMs) > 22%). Dominant TMs tended to have a low value of *creation* to *revision* ratio, which Deng et al. [2] found to be a characteristic of an experienced CAD user. Moreover, Johnson et al. [5] found that experienced users generate more CAD actions and overall CAD models of good quality.

Conclusions:

This research has provided some initial insights into the potential of utilizing non-intrusive data logging for data analysis to identify CAD user archetypes within design teams. The analysis of 14 design teams showed similarities and differences in team composition regarding user archetypes. The results have shown the two most common design teams' compositions in the context of clusters related to the usage of part and assembly classes of CAD actions. The first composition is consisted of a "part specialist", an "assembly specialist" and a versatile TM, while the other replaces the "assembly specialist" with another "part specialist" and shifts assembly tasks onto the versatile TM. Furthermore, a dominant TM has been identified in half of the teams who carried the most workload and is the most experienced member of the team. Scholars and educators can utilize these results to analyze different user archetypes within design teams. Their findings could then be used to pair together designers whose archetypes complement each other's, in order to maximize the design teams' potential. Furthermore, the results presented in this paper could be helpful for engineers and educators to gain insight into different compositions of design teams and their CAD modeling approaches which led to the generation of an overall CAD model of higher quality.

In future work, a more concrete study should be performed. To gain a deeper understanding of the design process, the implementation of user modeling patterns in the context of CAD actions in the analysis is needed. Moreover, the implementation of machine learning algorithms could be used as a tool to process the data and identify different patterns of transitions between CAD actions.

Acknowledgments:

This paper reports on work funded by the Croatian Science Foundation project IP-2018-01-7269: Team Adaptability for Innovation-Oriented Product Development - TAIDE. The authors would like to thank the study subjects for their participation.

References:

[1] S.K. Chandrasegaran, K. Ramani, R.D. Sriram, I. Horváth, A. Bernard, R.F. Harik, W. Gao, The evolution, challenges, and future of knowledge representation in product design systems, Computer-Aided Design. 45 (2013) 204–228. https://doi.org/https://doi.org/10.1016/j.cad.2012.08.006.

[2] Y. Deng, M. Mueller, C. Rogers, A. Olechowski, The multi-user computer-aided design collaborative learning framework, Advanced Engineering Informatics. 51 (2022) 101446. https://doi.org/10.1016/j.aei.2021.101446.

[3] M.H. Rahman, M. Gashler, C. Xie, Z. Sha, Automatic Clustering of Sequential Design Behaviors, 2018. https://doi.org/10.1115/DETC2018-86300.

[4] B. Stone, J. Salmon, A. Hepworth, E. Red, M. Killian, A. La, A. Pedersen, T. Jones, Methods for determining the optimal number of simultaneous contributors for multi-user CAD parts, Comput Aided Des Appl. 14 (2017) 1–12. https://doi.org/10.1080/16864360.2016.1273578.

[5] M. Johnson, E. Ozturk, B. Yalvac, L. Valverde, X. Peng, K. Liu, Examining Adaptive Expertise: A Novel Comparison of Student and Practicing Engineer CAD Modeling Performance, in: 2015: p. V005T05A007. https://doi.org/10.1115/IMECE2015-50296.

[6] R. Celjak, N. Horvat, S. Skec, Exploring the Potential of Tracking CAD Actions in Project-based Courses, in: CAD Solutions, LLC, 2022: pp. 302–307. https://doi.org/10.14733/cadconfp.2022.302-307.