



HAL
open science

A Generic Interactive Membership Function for Categorization of Quantities

Liqun Liu, Romain Vuillemot

► **To cite this version:**

Liqun Liu, Romain Vuillemot. A Generic Interactive Membership Function for Categorization of Quantities. IEEE Computer Graphics and Applications, In press, 43, pp.39 - 48. 10.1109/mcg.2023.3301449 . hal-04241000

HAL Id: hal-04241000

<https://hal.science/hal-04241000v1>

Submitted on 22 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A GENERIC INTERACTIVE MEMBERSHIP FUNCTION FOR CATEGORIZATION OF QUANTITIES

Liqun Liu

École Centrale de Lyon, SICAL, LIRIS
liuliqun.cs@gmail.com

Romain Vuillemot

École Centrale de Lyon, SICAL, LIRIS
romain.vuillemot@ec-lyon.fr

December 14, 2023

ABSTRACT

The membership function is to categorize quantities along with a confidence degree. This paper investigates a generic user interaction based on this function for categorizing various types of quantities without modification, which empowers users to articulate uncertainty categorization and enhance their visual data analysis significantly. We present the technique design and an online prototype, supplementing with insights from three case studies that highlight the technique’s efficacy among different types of quantities. Furthermore, we conduct a formal user study to scrutinize the process and reasoning users employ while utilizing our technique. The findings indicate that our technique can help users create customized categories. Both our code and the interactive prototype are made available as open-source resources, intended for application across varied domains as a generic tool.

1 Introduction

Zadeh proposed Fuzzy logic theory [1] in 1965, to represent imprecise or ambiguous information. Since then, it is applied to many domains [2], e. g., decision-making, control system, machine learning and computer vision, among others. In fuzzy logic, the *membership function* is the key concept to define the confidence of truth in the information. This function applies to any quantities and enables the reasoning over the real-world problem by generating partial, fuzzy categories that capture the impreciseness and ambiguity in information. In this paper, we argue it can also support humans’ reasoning and logic [3] if this function is made interactive.

The challenge we seek to tackle involves establishing an explicit mapping between quantities and categories with assured confidence. Our objective is to delve into how humans rationalize quantities across varied domains and data types. While most visualization techniques and tools usually address it during the visual encoding of data (e. g., using color), it remains internal—or at best using a legend—without providing a fully explicit set of properties for this mapping and explicit the humans’ perception of the internal categorization. Some approaches use automatic categorization methods for extracting and representing knowledge on quantitative values, e. g., KNN, K-Means, DBSCAN. Besides, a number of research introduced fuzzy logic into visualization, such as Disk diagram [4], FINGRAMS [5], or FuzzyRadar [6]. However, those methods do not provide enough control over the mapping process and the possibility for users to categorize quantities themselves. Some methodologies allow for the categorization of quantities with customizable boundaries for uncertainty analysis, but they typically concentrate on specific data types, such as medical volume data [7, 8].

In this paper, we present an enhanced generic interactive membership function for quantities from various domains, aiming at capturing human reasoning when defining parameters such as the number of categories and their span (shown in Fig. 1). Such parameters usually are based on data distribution or domain conventions, but using the technique will be made customizable in a few mouse clicks. We implement three case studies to affirm the efficacy of our interactive technique, which seamlessly integrates our interactive technique with visualizations from three distinct domains, each dealing with a unique type of quantity. Finally, we conduct a user study utilizing age and temperature datasets to

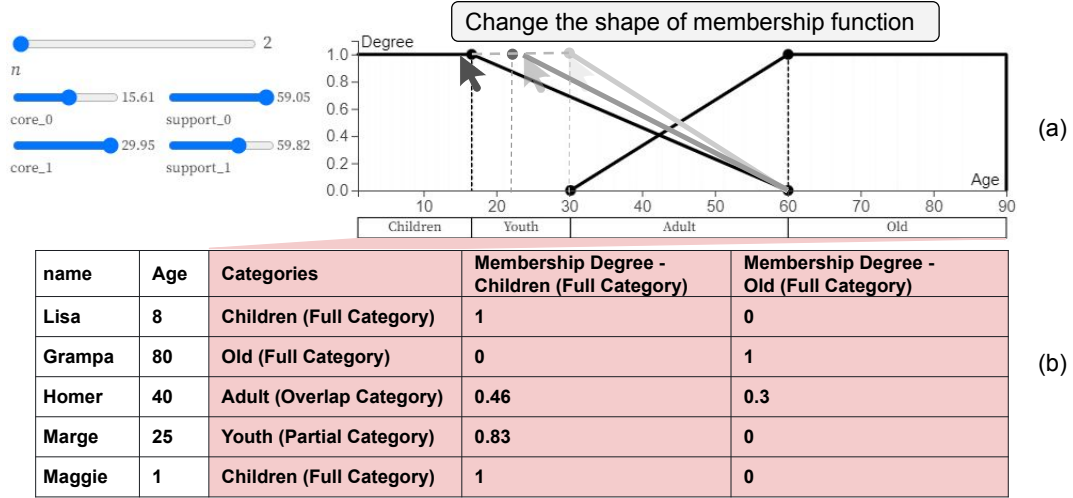


Figure 1: The illustration of an interactive membership function. Users can change the shape of the membership function by dragging the black points in (a) to adjust the range of the categories (Children, Youth, Adult, and Old). This interactive membership function helps users map the quantities (column *Age*) into categories (column *Categories*). The table in (b) shows the membership degrees derived from the membership function.

investigate the impact of our interactive membership function on the categorization process and results as performed by users. The main contributions are as follows:

- We introduce an enhanced interactive membership function that let users categorize quantitative values into categories and describe those generated categories with semantic expression.
- We introduce an interactive prototype that embodies this membership function within a web application. In addition, we offer users an API, extending them the flexibility to apply our technique across a wide types of quantities.
- We report on a user study to evaluate the interactive technique. This includes the meticulous collection and subsequent analysis of data pertaining to how users navigate and operate the prototype.

2 Related Works

This section reviews two aspects of works related to categorizing quantities that include fuzzy logic theory in visualization and uncertainty visualization.

2.1 Fuzzy Logic Theory in Visualization

Visualizations exist that employ the principles of fuzzy logic theory to encapsulate the ambiguity of categorization - a concept we refer to as *fuzzy visualization*. In this context, *fuzzy* signifies the possibility of a truth value fluctuating on a spectrum between absolute falsehood and absolute truth. For example, the Disk diagrams [4] proposed by Yeseul Park *et al.* can visualize fuzzy set (a class of objects with a continuum of grades of membership). It describes the complexity of fuzzy sets by showing the degree among sets with the layout of star coordinates.

Numerous combined methodologies exist for the visualization of fuzzy sets and fuzzy clustering, a process wherein each data point may be associated with multiple clusters. The typical one is the combination between radial coordinate and parallel coordinate plot, introduced by Zhou *et al.* [9]. This method reflects real-world clustering scenarios and improves the understanding of fuzzy clusters. Similarly, RadViz proposed by Sharko *et al.* [10] visualizes the fuzzy clustering of multiple dimensional datasets using radial visualization and the dimensional reduction method. Also, Rose Diagram proposed by Buck *et al.* [11] uses vectors of fuzzy attributes to visualize a fuzzy weighted graph. This method is an extension of the standard polar area diagram (a type of pie chart) and is helpful for decision-makers to choose a better way from several options while estimating the potential trade-offs.

With the development of visualization techniques, fuzzy visualization has become more abundant. Hall *et al.* [12] introduce parallel coordinates to fuzzy visualization and transform 3D or more than three-dimensional data into two dimensions without losing information. Similarly, Pham *et al.* [13] introduce the 3D parallel coordinates to visualize

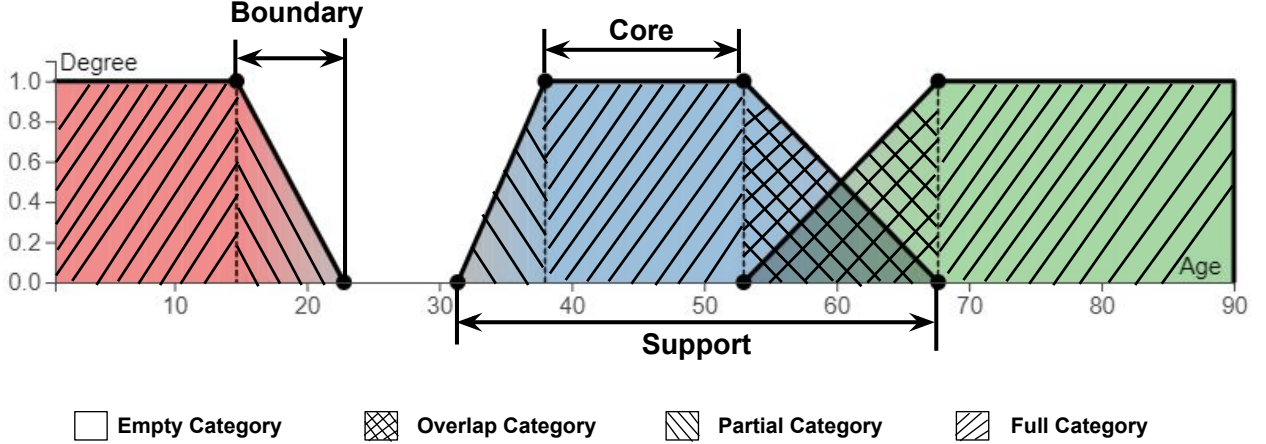


Figure 2: Parameters and categories of membership function. The parameters contain *Core*, *Support* and *Boundary*. Based on three parameters, there are four kinds of categories generated, *Full Category* with only one membership, *Partial Category* with one membership, *Empty Category* without any membership, and *Overlap Category* with two memberships.

fuzzy data. Its advantage is that it is easier to distinguish the core and support in fuzzy sets. Furthermore, visualization techniques use novel methods to render fuzzy relations. Berthold *et al.* [14] propose a model to visualize the collection of fuzzy points in parallel coordinates. The parallel coordinates visualize the degree of membership function with different degrees of shading.

Despite these works facilitating the communication of predefined categories, they do not empower users to generate their own categories to encapsulate uncertain information. Our work equips users with the ability to customize category boundaries, thus enabling a more tailored data representation.

2.2 Uncertainty Visualization

Research on uncertainty visualization also offers efficient methods to convey ambiguity during the categorization process. As introduced by Brodlié *et al.* [15], the uncertainty of visualization exists in any data type, e. g., point data, scalar data, multi-field scalar data, and vector data. In order to address the challenges in uncertainty visualization, Skeels *et al.* [16] propose a classification of uncertainty for information visualization. It includes five categories: disagreement uncertainty from the difference of multiple times measurements, completeness uncertainty from the missing values, inference uncertainty from the model and prediction, measurement precision uncertainty from imprecise measurements, and credibility uncertainty from the conflict of the different data sources. Much effort has been devoted to designing visualization techniques, such as what Dong *et al.* [17] did, an interactive tool to help users recognize the situations and comprehend the ambiguity. The field of uncertainty visualization techniques has made significant contributions to specific domains, notably medical data. This includes rendering volume data with a membership function as demonstrated by Lundström *et al.* [7] and design of interactive tool by Drebin *et al.* [8].

In conclusion, these works have presented several probabilistic uncertainty classifications with visual encoding methods. However, these works are primarily directed towards specific types of datasets rather than the development of universally applicable tools. Our research presents an interactive methodology that facilitates users in formulating adaptable categories. Importantly, this technique extends its applicability to a variety of quantity types, transcending the constraints of specific data types.

3 Enhanced Membership Function

Liu *et al.* have already proposed an interactive membership function [18] to categorize quantities into categories based on fuzzy logic theory, but with limited interaction capabilities. We augment the interactive user experience by enriching the interaction ways so users can categorize quantities in a more efficient and visual way. We also include the technique in a set of coordinated views to present the generated attributes and their impact on connected visualizations.

3.1 Parameters and Categories

The membership function is the mapping function for creating fuzzy sets with the confidence of truth. As shown in Fig. 2, the x -axis refers to the quantitative values, and the y -axis refers to the membership degree. The values of membership degrees hint at how confident an element belongs to a specific set.

The membership function generates the categories with different membership degrees. These categories come from different shapes of membership function, which is related to three parameters as shown in Fig. 2: 1) *Core*: it refers to the elements entirely or fully belonging to one membership, and the membership degree is equal to 1; 2) *Support*: it refers to the elements that belong to one membership with membership degrees are more than 0; and 3) *Boundary*: the categories that contain elements that have non-zero memberships and incomplete memberships.

Based on these parameters, there are four special categories generated, which are *Partial Category*, *Empty Category*, *Full Category* and *Overlap Category* as shown in Fig. 2. *Empty Category* refers to the categories that do not belong to any categories, and their membership degrees are equal to 0. *Overlap Category* refers to the categories that belong to two memberships at the same time and with membership degrees more than 0 and less than 1. *Partial Category* refers to the categories that only belong to one membership but with membership degrees less than 1 and more than 0. *Full Category* corresponds to parameter *Core*.

3.2 Interactions

This paper builds on an existed interactive membership functions [18] and enriches the types of interactions. In the membership function, as shown in Fig. 1, a user can adjust these parameters to generate different shapes of the membership functions by dragging the black points on the membership function or dragging the sliders on the left side. Besides, membership function also offers the naming function. The membership function would change the categories' names by inputting texts in the input fields. Therefore, users can derive new attributes from raw data by interacting with the technique.

4 Implementation

We implemented the enhanced interactive membership function with library D3 [19] and JavaScript. It is released in Observable¹, which is able to be used by coding `import {fuzzyCut} from '0820d2ad 9cfa734d'` in other notebooks. An online prototype² is available on Observable, along with a supplemental material³. This interactive prototype inputs the quantitative data such as the *age* and outputs the categories of age such as *children* as shown in Fig. 1(b). This prototype allows users to operate it to derive values of new attributes by changing the shape of the membership function. The quantitative data can be binned into categories with membership degrees and customized names through the interactive membership function, which expands the applications of fuzzy theory in the interactive visualization fields. The current technique is based on the web. In the future, we will release an R and Python package to expand its application in different programming environments.

Moreover, we calculate the color values in the membership function using HSL model. As shown in Fig. 2, which is based on the Eq. 1:

$$Color = HSL(MAX(Hue(d), Hue(r)), MAX(\mu_d(x), \mu_r(x)), l) \quad (1)$$

where $Hue(d)$ and $Hue(r)$ represent the *Hue* value of d_{th} trapezoid and r_{th} trapezoid, $\mu_d(x)$ and $\mu_r(x)$ for the membership degree of elements x in the d_{th} trapezoid and r_{th} trapezoid, which are equal to the value of saturation. The value of lightness l is equal to 50 in this color model of the membership function. Thus, the less the membership degree is, the smaller the saturation is. It would not render any color when the element does not belong to any trapezoid.

5 Case Studies

In this section, we demystify the mechanics of the generic interactive membership function and exemplify its deployment across diverse types of quantities, which are substantiated through three case studies (A detailed representation of the data and their inherent characteristics utilised in these cases is provided in Table 1): displacement with uncertain force

¹<https://observablehq.com>

²<https://observablehq.com/d/0820d2ad9cfa734d>

³<https://observablehq.com/d/9479ec50c448978d>

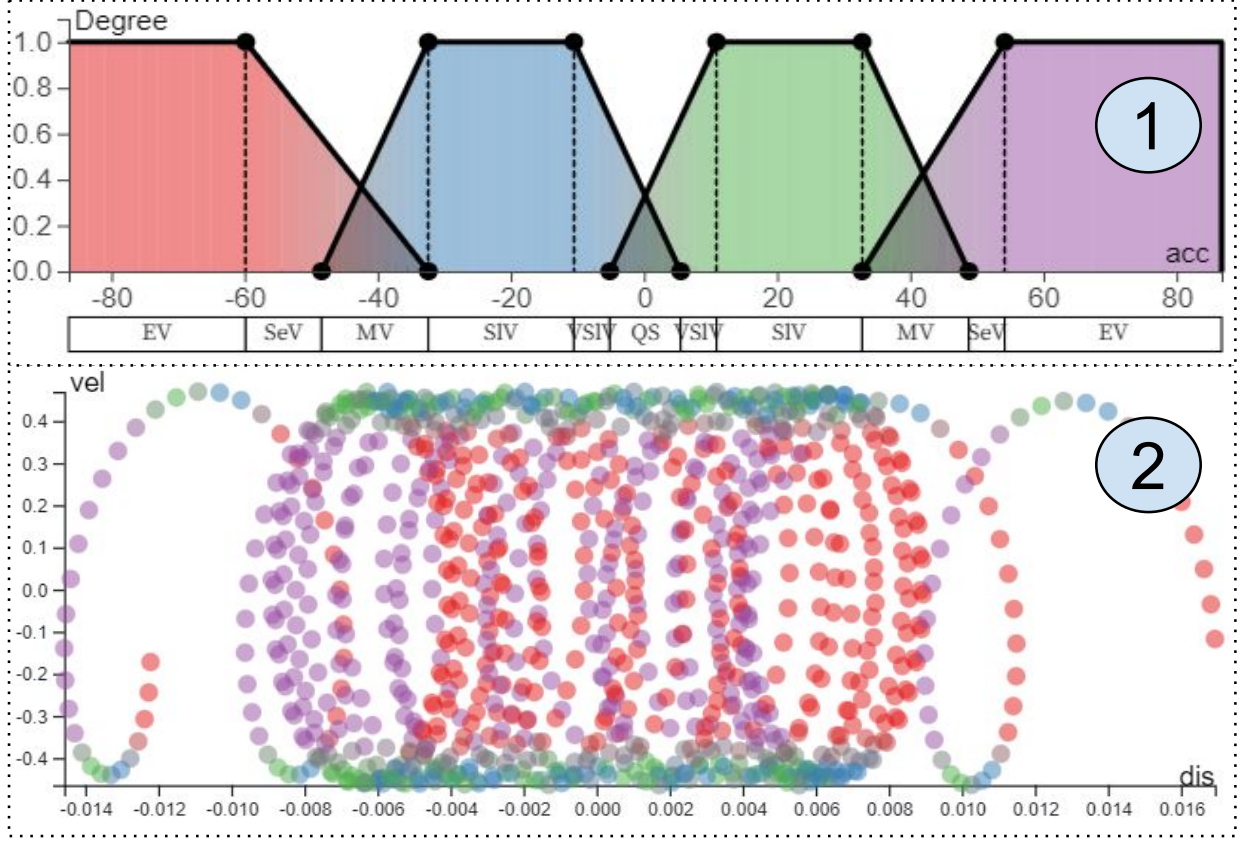


Figure 3: Categorize acceleration. The membership function creates 11 categories, as shown in ①, where the x axis refers to displacement, and y axis refers to membership degree. A scatterplot combines with it ②, where the x axis refers to displacement, and the y axis refers to velocity. The scatterplot reflects the vibration status and the colors also correspond to the colors in ①.

(case one), status of carbon dioxide (case two) and computational time of optimization algorithm (case three). We implemented three prototypes to help users categorize specific types of data interactively.

5.1 Case 1: Displacement with Uncertainty Force

In the vibration control field, control strategy greatly influences performance. Compared with traditional control strategies, the fuzzy control strategy does not need explicit mathematical modeling, making it more applicable in complex modeling, in which assumptions and approximations are often made. An interactive visual technique to observe the changes in vibration status is essential. In this case, domain experts can gain insight into fuzzy control strategy and reveal the influence of parameters by adjusting parameters easily, which is significantly meaningful for vibration control engineering. The dataset includes 1024 items with attributes *acc* (acceleration), *vel* (velocity), and *dis* (displacement).

In order to help domain experts observe vibration status, we designed a prototype by combining a scatterplot (Fig. 3(②)) with our technique (Fig. 3(①)). The new attributes including *Categories*, *Membership Degree*, and *Color*. The x -axis in the scatterplot represents displacement, and the y -axis refers to velocity, which can reflect the vibration status. As shown in ①, we created 4 trapezoids (*Full Category*). Besides, there are also three *Overlap Category* generated. Using the interactive membership function makes it easy to categorize acceleration and monitor which parts are unstable and quasi-static.

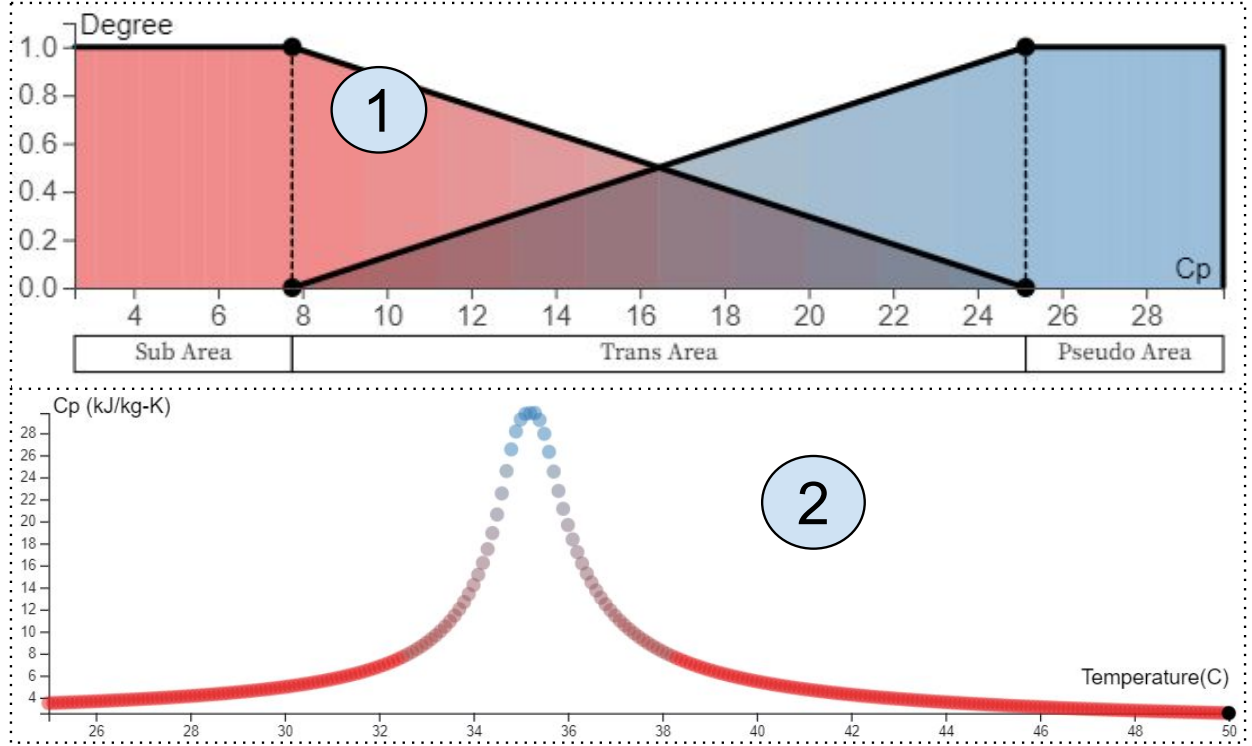


Figure 4: Categorize **Specific Heat**(C_p). In ①, the x axis shows C_p and y axis represents the membership degree, which generates two *Full Categories*. ② shows the relationship between temperature(x axis) and C_p (y axis) with the scatter plots.

5.2 Case 2: Status of Carbon Dioxide

Supercritical carbon dioxide is a special carbon dioxide status, which usually happens at specific temperatures and pressure. Carbon dioxide usually exists as gas at standard temperature and pressure, or it could be a solid called dry ice when the temperature is frigid or pressure increases efficiently. However, if the temperature and pressure reach a point, carbon dioxide exists in a critical status, which is midway between liquid and gas, called supercritical carbon dioxide. Nevertheless, there is no clear line to judge this special carbon dioxide, which is usually a temperature range and pressure. In this case, our proposed technique can help domain experts observe the changes around a critical point and interactively generate the confidence of belonging to this point. The dataset includes 251 items and with attributes *Temperature*, *Pressure*, *Density* and C_p (Critical Heat).

As showed in Fig. 4, we created 2 trapezoids (*Full Category*), and represented with 2 colors: red and blue. It generated two *Full Categories* that are *Pseudo Area* from 25.7 to 29.8 and *Sub Area* from 2.6 to 7.67. Besides, it also generated one *Overlap Category* named *Trans Area* from 7.67 to 25.7. Fig. 4 ② shows the relationship between temperature and C_p . The supercritical carbon dioxide is around the peak point in ②, C_p is a crucial parameter to judge if it is supercritical carbon dioxide. The blue points are close to supercritical carbon dioxide, and the red points are not. Thus, using this technique helps users understand and analyze the dataset.

5.3 Case 3: Computational Time

Computational time plays a significant role in the efficiency of an algorithm for solving combinatorial optimization problems. This case study categorizes the computational time for solving the Stochastic Container Relocation Problem (SCRCP) that aims to retrieve all containers from a bay with the minimum number of relocations, which is a typical combinatorial optimization problem arising in container port operations. We collected data from [20], which includes the computational times for solving the SCRCP by an exact algorithm on 288 instances over 24 problem sets. A problem set is characterized by two parameters/attributes: S and T . S represents the number of stacks of a bay, and T represents the number of tiers of a bay. Due to the curse of dimensionality, the computational time increases dramatically with the scale of the problem, and as a result, the algorithm cannot compute some problems within one hour. However, it can

compute some problems in several milliseconds. Thus, evaluating the samples based on categorized computational times can help researchers choose the correct way to solve this problem fast.

As shown in Fig. 5, ① illustrates that we created three main categories, which correspond to small-scale problems, medium-scale problems, and large-scale problems, respectively. Usually, we define the small-scale problem that can be solved to optimality within five minutes, while we define those that cannot be solved to optimality within one hour as large-scale problems. However, for the SCRP, we cannot simply differentiate the problem scales by the computational time because their computational times can vary significantly for the instances in the same problem set. Therefore, to better understand the computational complexities of these problem sets, we would like to make a more detailed classification of the instance scales. This case study classifies the instances into seven categories with membership degrees based on their computational times. Making such detailed classifications is because the researcher is also concerned with the extent to which the instances can be solved quickly or slowly.

This technique is beneficial in two ways. Firstly, the researcher has a clear picture of the distribution of the computational times of each problem set, as shown in Fig. 5(②), which enables users to have a more accurate understanding of the computational complexities of different problem sets. Secondly, the researcher can quickly identify the problem sets that are very difficult to solve and thus turn to heuristic algorithms to obtain a near-optimal solution in a reasonable time.

6 User Study

In order to verify the effectiveness and limitation of the interactive technique, we conducted a user study. Our main goal is to investigate how easily users can use the technique and how effectively users can categorize quantities with different data types. By interviewing some users who had the experience of categorizing quantities and exploring the habits humans resonate with quantities, we summarized three important hypotheses in this user study:

- H1** The interactive technique would help people categorize quantities and affect their categorization results.
- H2** The number of data items would affect the categorization results, which means that the same user will create different categorizations with different amounts of data items.
- H3** Users would express the uncertainty categories with comparative and descriptive words.

6.1 User Study Setup

In order to test if the hypotheses are TRUE or FALSE, we designed the user study in three steps. We first recruited participants who were in the computer science field or not, and then we invited participants to categorize the quantities. To ensure we recruit participants who are familiar with the datasets used, we employed two widely recognized datasets: *Age* and *Temperature* data. Finally, we measured the categorizing quantities results.

Participants: We recruited 8 participants to take part in this study, and their ages range from 26 to 32, with 5 males and 3 females. Two of them have computer science field backgrounds and have rich experience reasoning on quantities. All of them have not used our technique before. We invited all participants to fill in a pre-study questionnaire relating to essential information, such as if they are similar to visualization and have a computer science background. Besides, in this pre-study questionnaire, we invited the participants to categorize the *Age* data and *Temperature* data based on their background knowledge.

Tasks: We invited all participants to operate the interactive technique at least once, and we collected their operating activities and the categorization results. We designed a web page⁴ to let users operate the technique following the introduction. In this web page, we first explained the technique with examples, and then we implemented two prototypes using *age dataset* and *temperature dataset*. Through users' activities on this web page, we recorded the operating results and saved these results to a remote server. This includes three steps. In **Step. 0**, there is a brief introduction to this interactive technique, which includes some basic information and workflow introduced by a video. In **Step. 1**, we invite participants to create a specific visualization to make it similar to the target visualization (the visualization we pre-supplied) in aspects of names range of categories. This helps participants to understand how to manipulate the technique quickly. After that, we invite the participants to categorize two datasets *Age* and *Temperature* with different amount of data items in **Step. 2**. Firstly, participants categorize the data with a small number of data items and then with a greater amount of data items if the participants think that the categories should be updated when the amount of the data items increases. After finishing this operation, participants need to fill in another questionnaire that includes qualitative feedback about this technique. We encouraged participants to 'think aloud' about what they are doing, how

⁴observablehq.com/d/45001390e4f1b08f

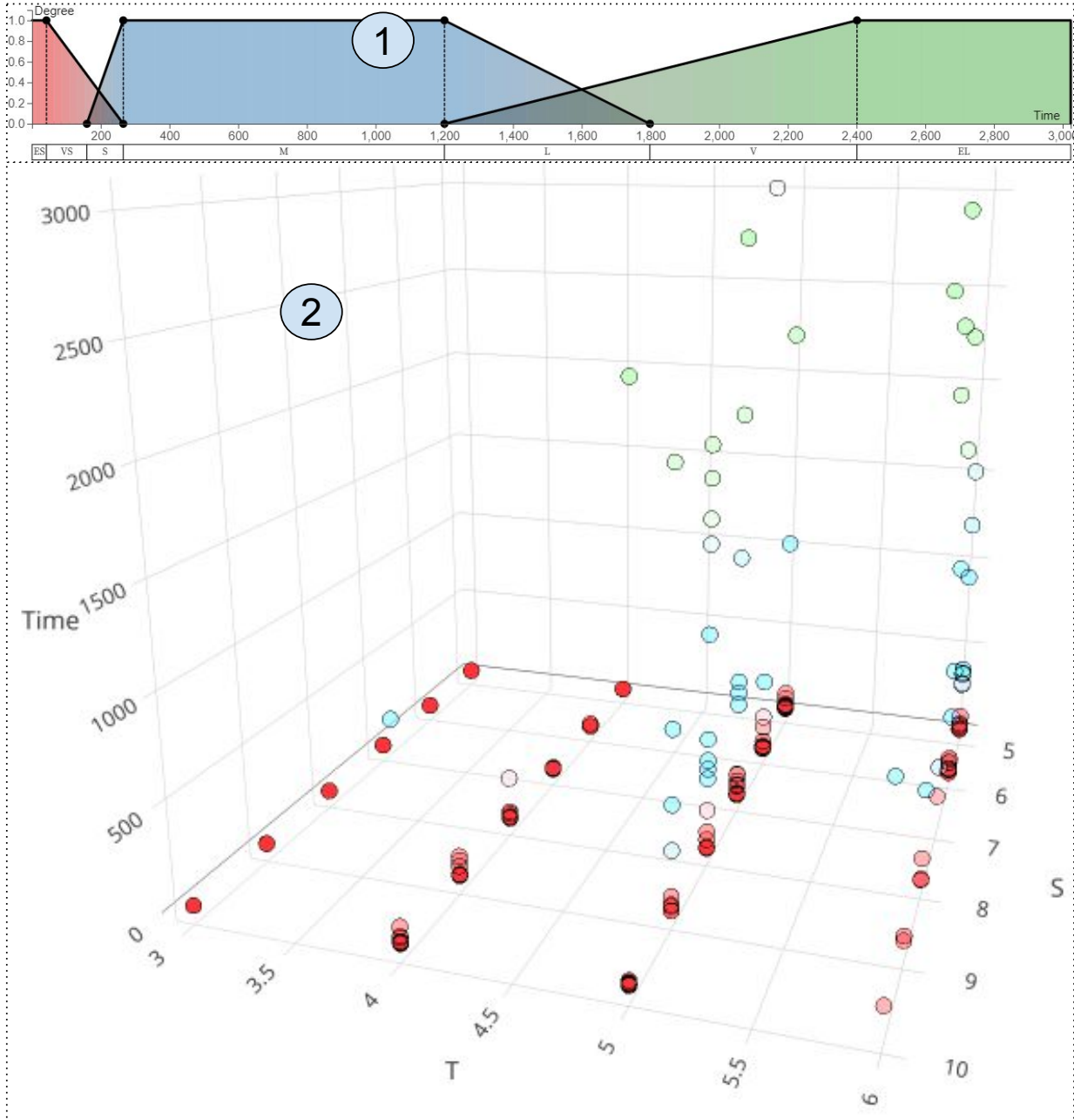
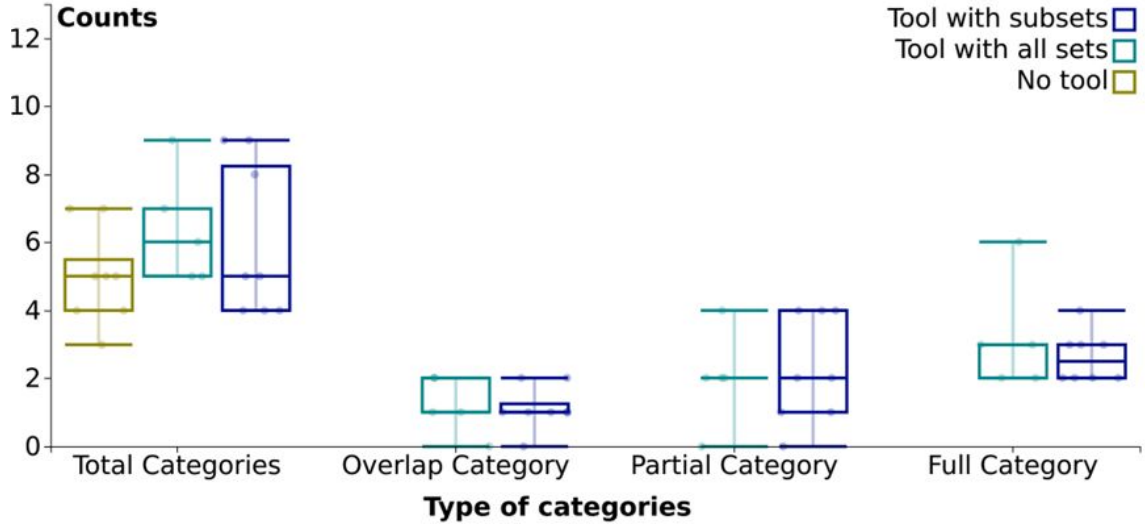


Figure 5: Categorize computational time. The membership function creates 7 categories as shown in ①, where x axis represents the time, and the y axis represents the membership degree. A 3D scatterplot ② combines to the membership function, where x -axis, y -axis, and z -axis are T (the number of tiers of a bay), S (the number of stacks of a bay), and Time respectively.

they think, and why they operate the parameters during the study. We took notes about these questions and assisted if it was necessary.

Measures: We recorded the information about the participants' operating activities and the categories generated by users to a remote server. The categories include maximum values, minimum values, names, and membership degrees. These attributes were collected and analyzed to calculate the participants' categorizing behaviors and reasoning. The first time categorization and second time categorization can reflect whether participants are disturbed by the number of data items.

(a) Categories of age dataset



(b) Categories of temperature dataset

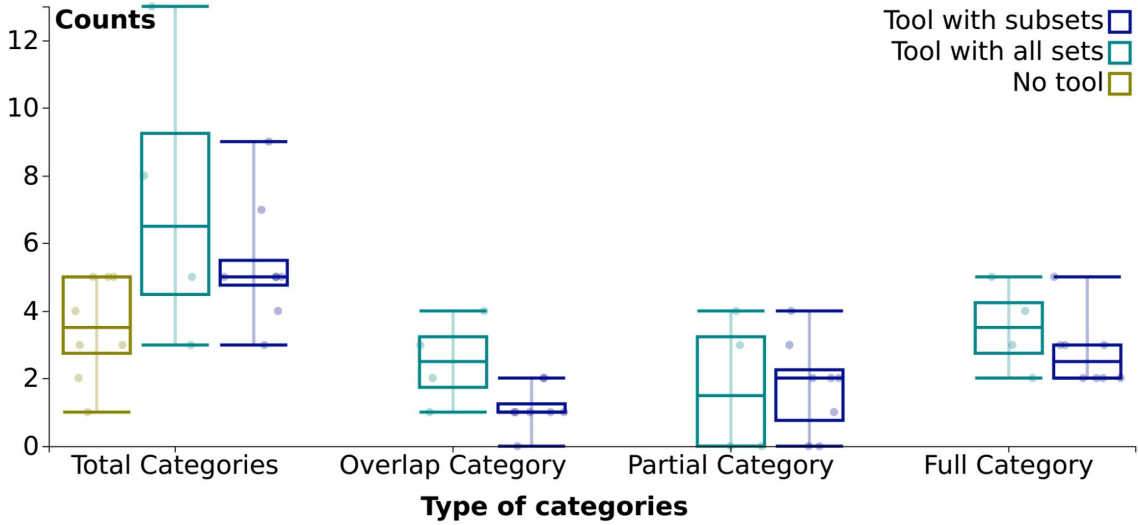


Figure 6: The statistic of categories was generated from 8 participants with an age dataset in (a) and a temperature dataset in (b). The x -axis represents the type of categories such as *Overlap Category*, *Full Category* and *Partial Category* as shown in Fig. 2; *Total categories* refers to all the generated categories; the y -axis represents the number of each type of categories.

6.2 Results

We visualized the results of categories for *age* dataset and *temperature* dataset in Fig. 6. The x -axis represents the type of categories, which contains: *Total Categories* for all categories generated, *Overlap Category* for the categories having two memberships at the same time, *Partial Category* for the categories only having one incomplete membership (the membership degree is more than 0 but less than 1) and *Full Category* for the categories only having one complete membership (the membership degree is equal to 1). We have illustrated these concepts of these categories in Fig. 2. The different colors of boxes showed the categorization by different ways, where □ represents the categorization without using any tool, □ represents the first-time categorization using our technique with the sub dataset and □ represents the

second-time categorization using our technique with the entire dataset. The y -axis represents the number of each type of category. The points in the boxes represent the individual participants.

As shown in both (a) and (b) in Fig. 6, the categorizations without using our tool did not generate *Overlap Category*, *Full Category* and *Partial Category* categories. The *Total Categories* showed that, compared with the categorization without using our tool, the categorizations by using our tool with either the sub dataset or the entire dataset generated more categories. This indicated that the interactive technique has a specific effect on participants while categorizing quantitative values because the number of categories increased significantly with the use of our tool, which verified H1. It also indicated that participants could understand the data deeply after using the tool. Besides, compared with the twice categorizations by using our tool, we found that the second time categorization generally had more counts in *Total Categories* and *Full Category*, which means that users created more categories in the second time categorization while using the entire data items. Therefore, the amount of the data items affect the categorization results, which verified H2.

We collected and analyzed the results of the naming methods. Fig. 7 (a) showed all the labels used to name the categories by the participants and how frequently they used them each time. The words *Hot*, *Children* and *Old* showed up more frequently than other words, which indicates that the participants usually prefer to use some common words as labels for new categories, which verified H3. Besides, words *Very*, *Large*, *Small* and *Middle* also appeared more frequently, from which we can conclude that the participants usually prefer to use adjectives or adverbs combined with nouns to represent the degree of the data.

The post-study questionnaire collected some qualitative feedback from the participants, and the results are shown in Fig. 7 (b). The y -axis represents the 6 questions related to the performance of this technique and the x -axis represents the scores that the participants give for each question. The score is a Likert scale from 1 to 5, which means very negative to very positive. Most of the scores are more than 2. The question related to *Easy to interact* had the most positive feedback from the participants, which reflected that the participants could engage with the interactive functions quickly, and thus the interactive technique is effective. However, the question related to *Creating categories* and *Parameters understanding* had a relatively lower performance compared with other aspects. There might be because these participants are not familiar with the fuzzy theory, so it took time to understand the parameters of membership function and categorize.

6.3 Other Feedback

We also asked all participants their feedback while using the interactive technique, which is overall positive. All participants think this interactive technique is easy to understand and manipulate. The explicit and interactive view of the membership function facilitates the refinement of created fuzzy categories. This enhancement, in turn, fosters more accurate categorization, thereby aligning more closely with real-world scenarios. The participants are from a range of domains. Some of them thought this technique was beneficial in many domains, especially those with no clear mathematics relation. In that case, people decide on their own experience, so an interactive technique is essential to help people understand the information in the data.

Moreover, there are some issues with the current version. One of them is the calculation method. They thought it would be better if the distribution of the dataset could calculate the membership degree. In addition, there are operational issues. The names of categories will be overlapped when there are many categories, or these generated names will disappear when they zoom in or zoom out on this web page. In the parameters aspect, some participants said they could not quickly check the values of *core* and *support*.

Table 1: Overview of datasets incorporated into the interactive membership function. Each dataset consists of quantitative values with varying characteristics, including both continuous and integer types.

Datasets	Data types	Data characterization
Displacement	Quantitative values	Unbounded interval ($x \in [-\infty, +\infty]$), continuous
Critical heat	Quantitative values	Half-bounded interval ($x \in [0, +\infty]$), continuous
Computational time	Quantitative values	Half-bounded interval ($x \in [0, \infty]$), continuous
Temperature	Quantitative values	Unbounded interval ($x \in [-\infty, +\infty]$), continuous
Age	Quantitative values	Half-bounded interval ($x \in [0, \infty]$), integer

6.4 Limitations

The current version of our tool has a limitation—it can only categorize single quantities in a dataset. For instance, users can categorize based on the speed of taxis, but there are often multiple quantities to consider, such as the duration

(a) The frequency of categories' names



(b) The results of the post-study questionnaire

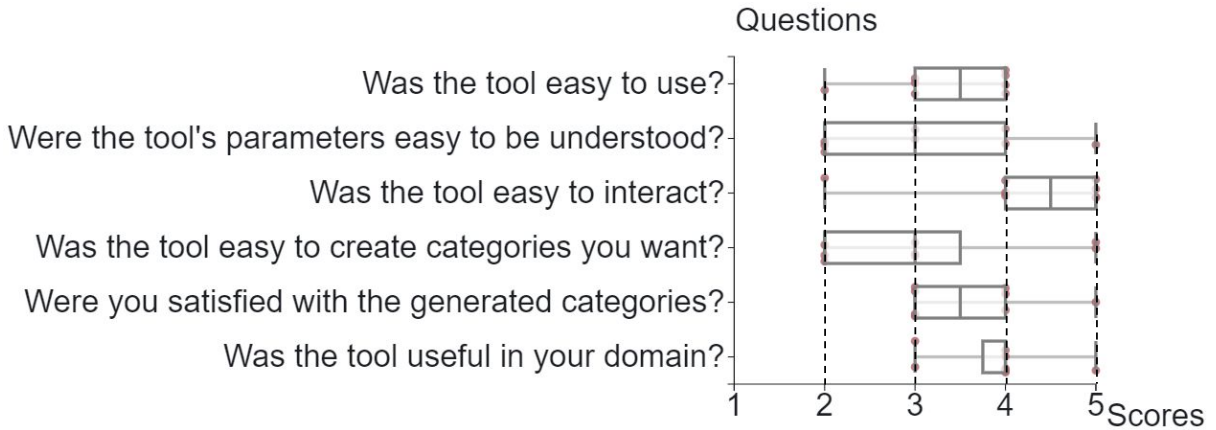


Figure 7: The name of fuzzy categories and the results of the questionnaire. A word cloud shows the used words for fuzzy categories in (a). The bigger size of the text, the more frequently participants use it. A Likert scale displays the analytic results of the questionnaire in (b). The box plot represents a Likert scale for 8 participants answering 6 questions. The scores are from 1 to 5, with 5 representing the most positive feedback and 1 representing the most negative feedback.

of trajectories or driving directions. In other words, our tool currently lacks the capability to classify datasets by considering these multiple features simultaneously.

Additionally, we have approached all quantitative values in a general way—as continuous data. In this approach, the boundary of categorization could be any intermediate value between two categories. However, this methodology encounters limitations when dealing with integer data, such as age. For instance, we might end up with a boundary value like 18.3, which is impractical in the context of age. Thus, addressing this limitation in handling integer data types is an area for future refinement in our work.

7 Discussion and Future Work

As we have introduced, our generic interactive technique enables users to categorize thereby articulating the obscured uncertainty inherent in human cognition, using a fuzzy membership function.

The presented case studies affirm that this interactive membership function enhances users' capacity to categorize quantities bearing diverse characteristics. This can also be conveniently integrated as a legend when combined with other systems. As depicted in Table 1, our technique is applicable to not only continuous quantitative values, but also to integer values. Additionally, it can handle various interval data, including half-bounded and unbounded intervals. Contrasting with earlier visualization approaches, our method furnishes users with a more generic interactive membership function, capable of easily accommodating a variety of quantity types.

The user study conducted underscores the accessibility and ease-of-use of our technique, which aids users in categorizing quantities within customizable boundaries. Particularly, the explicit and interactive visualization of the membership function is advantageous for users, enabling them to refine categories effectively.

In the future, we will continue working on uncertainty visualization by expanding the applications of the interactive technique. We have two potential directions: (1) combining the interactive technique with Rule-based Fuzzy Systems to make uncertainty decisions and (2) expanding the type of membership functions to fit into more types of data distribution so, contributing to the uncertainty visualization.

Application in decision making under uncertainty. As we have introduced in previous sections, we can easily create the fuzzy categories interactively with our technique, and individuals can name the generated categories. By connecting the interactive membership function with the complex Fuzzy Rule-based Systems, it can make the decision-making process visually with membership function and textual interpretation. The textual explanation comes from the knowledge of multiple experts. We can record more information about how experts make decisions based on their own acknowledge through the textual information.

Expand the types of membership function for various uncertainty data distribution. The Trapezoidal membership function calculates the membership degrees based on the linear relationships. However, the categorization mechanism is not always linear in the real world. The Trapezoidal membership function cannot fit all distributions of quantity data. In the future, we will implement other membership functions in our technique to fit into different categorization mechanisms, such as the Gaussian membership function, Triangular membership function (a special status of Trapezoidal membership function), and Generalized bell-shaped membership function.

8 Conclusion

We introduced an enhanced generic interactive membership function, to support users in their quantity analysis process, by mapping them into one or multiple categories, along with a confidence degree. We report on an evaluation that shows it is helpful for engineering domains with different types of quantities, especially for the evaluation without clear mathematics relations. It can help users interactively understand the quantities since they can observe how the categories change when they create different shapes of membership functions. The number of data items also affects users' decision-making in the categorization process because the number of data items might change the data distribution. This work opened new research areas at the intersection of interactive visualization and fuzzy logic domains, which is currently under-explored. Having released our prototype and code under an open-source license, users are enabled to engage with it through an API. This not only advances further research in this intersection but also fosters the application in a broader range of use cases.

References

- [1] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.
- [2] J. Alcalá-Fdez and J. M. Alonso, "A survey of fuzzy systems software: Taxonomy, current research trends, and prospects," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 1, pp. 40–56, 2015.
- [3] H. T. Clifford, W. Stephenson *et al.*, *An introduction to numerical classification*. Academic Press New York, 1975, vol. 240.
- [4] Y. Park and J. Park, "Interactive visualization of fuzzy set operations," in *Visualization and Data Analysis 2010*, vol. 7530. International Society for Optics and Photonics, 2010, p. 753002.
- [5] D. P. Pancho, J. M. Alonso, O. Córdón, A. Quirin, and L. Magdalena, "Fingrams: visual representations of fuzzy rule-based inference for expert analysis of comprehensibility," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 6, pp. 1133–1149, 2013.
- [6] F. Zhou, B. Bai, Y. Wu, M. Chen, Z. Zhong, R. Zhu, Y. Chen, and Y. Zhao, "Fuzzyradar: visualization for understanding fuzzy clusters," *Journal of Visualization*, vol. 22, no. 5, pp. 913–926, 2019.
- [7] C. Lundström, P. Ljung, A. Persson, and A. Ynnerman, "Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1648–1655, Nov. 2007. [Online]. Available: <http://ieeexplore.ieee.org/document/4376198/>
- [8] R. A. Drebin, L. Carpenter, and P. Hanrahan, "Volume rendering," *ACM SIGGRAPH Computer Graphics*, vol. 22, no. 4, pp. 65–74, Aug. 1988. [Online]. Available: <https://dl.acm.org/doi/10.1145/378456.378484>

- [9] F. Zhou, B. Bai, Y. Wu, M. Chen, Z. Zhong, R. Zhu, Y. Chen, and Y. Zhao, “FuzzyRadar: visualization for understanding fuzzy clusters,” vol. 22, no. 5, pp. 913–926. [Online]. Available: <http://link.springer.com/10.1007/s12650-019-00577-2>
- [10] J. Sharko and G. Grinstein, “Visualizing Fuzzy Clusters Using RadViz,” in *2009 13th International Conference Information Visualisation*, Jul. 2009, pp. 307–316, iSSN: 2375-0138.
- [11] A. R. Buck and J. M. Keller, “Visualizing uncertainty with fuzzy rose diagrams,” in *2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*, Dec. 2014, pp. 30–36.
- [12] L. Hall and M. Berthold, “Fuzzy parallel coordinates,” in *PeachFuzz 2000. 19th International Conference of the North American Fuzzy Information Processing Society - NAFIPS (Cat. No.00TH8500)*, Jul. 2000, pp. 74–78.
- [13] B. Pham and R. Brown, “Visualisation of fuzzy systems: requirements, techniques and framework,” *Future Generation Computer Systems*, vol. 21, no. 7, pp. 1199–1212, 2005. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167739X04000512>
- [14] M. R. Berthold and L. O. Hall, “Visualizing fuzzy points in parallel coordinates,” *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 3, pp. 369–374, Jun. 2003.
- [15] K. Brodlié, R. Allendes Osorio, and A. Lopes, “A Review of Uncertainty in Data Visualization,” in *Expanding the Frontiers of Visual Analytics and Visualization*, J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, Eds. London: Springer, 2012, pp. 81–109. [Online]. Available: https://doi.org/10.1007/978-1-4471-2804-5_6
- [16] M. Skeels, B. Lee, G. Smith, and G. G. Robertson, “Revealing Uncertainty for Information Visualization,” *Information Visualization*, vol. 9, no. 1, pp. 70–81, Jan. 2010, publisher: SAGE Publications. [Online]. Available: <https://doi.org/10.1057/ivs.2009.1>
- [17] X. Dong and C. C. Hayes, “Uncertainty Visualizations: Helping Decision Makers Become More Aware of Uncertainty and Its Implications,” *Journal of Cognitive Engineering and Decision Making*, vol. 6, no. 1, pp. 30–56, Mar. 2012, publisher: SAGE Publications. [Online]. Available: <https://doi.org/10.1177/1555343411432338>
- [18] L. Liu and R. Vuillemot, “Categorizing quantities using an interactive fuzzy membership function,” in *The 12th International Conference on Information Visualisation Theory and Applications*, 2021.
- [19] M. Bostock, V. Ogievetsky, and J. Heer, “D³ data-driven documents,” *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.
- [20] V. Galle, V. H. Manshadi, S. B. Boroujeni, C. Barnhart, and P. Jaillet, “The stochastic container relocation problem,” *Transportation Science*, vol. 52, no. 5, pp. 1035–1058, 2018.

Liqun Liu is a PhD graduate at École Centrale de Lyon, Université de Lyon, and Research Affiliate at LIRIS, France. His research interests lie at the intersection of visualization and transportation systems. He holds a B.Eng in Marine Engineering, Wuhan University of Technology.

Romain Vuillemot is an assistant professor of computer science at École Centrale de Lyon, Université de Lyon, France. His research interests include information visualization, visual analytics, and human-computer interactions, with a focus on advanced interaction techniques to include humans in the loop of complex predictive analytics processes. He has a PhD in computer science from INSA de Lyon.