Searching for types of goals in a Conceptual Space of goal characteristics

Felix Weber[1,2,*], Ahmed M. H. Abdelfattah[3] and Kai-Uwe Kühnberger[2]

[1]*Zentrum für Digitale Lehre, Campus-Management und Hochschuldidaktik (virtUOS) (Universität Osnabrück),Heger-Tor-Wall 12 (building 42), 49074 Osnabrück, Germany*

[2]*Institute of Cognitive Science, University of Osnabrück, Wachsbleiche 27 (building 50), 49090 Osnabrück, Germany*

[3]*Faculty of Science, Ain Shams University, Cairo, Egypt*

Abstract

This paper introduces an approach to combine ideas from Gärdenfors' Conceptual Spaces Framework (CSF) [1] and Clustering Techniques to the domain of Psychometrics in general and Motivational Psychology in particular. In the goal-setting literature of the last 50 years, scientists have postulated and empirically confirmed a wide variety of goal types with specific characteristics. We aim to reproduce such goal types, in line with CSF, as convex regions in quality dimensions. The data we use originates from an ongoing field study with a digital study assistant (DSA) for goal-setting in higher education and holds goals in natural language ($n = 637$), formulated by university students ($n = 38$), each related to scores for 32 goal characteristics, assessed with the Goal Characteristics Questionnaire (GCQ) [2]. The method we apply in this paper is searching for multi-peaked distributions by visual inspection of violin plots, scatter plots, and kernel density estimation plots (KDE) of single characteristics and two-dimensional permutations. If there are differences in data density in dimensions, applying clustering algorithms in these dimensions is worth the computation time. The results show multi-peaked distributions, while no non-overlapping convex clusters are evident by visual inspection. In the KDE plots, summits of high density exist, which are prospective candidates for convex regions, aka types. The findings encourage us to proceed in the endeavor and apply clustering algorithms in future studies, which may allow us to reproduce previous findings of goals and their characteristics with a novel method, apply and test the CSF on real-world data, and possibly derive new insights into the nature of educational goals.

Keywords

Conceptual Spaces, Goal Characteristics, Ontologies of Cognitive Phenomena, Symbol Grounding, Clustering algorithms

1. Introduction

The interdisciplinary research outlined in this paper brings together Conceptual Spaces, an approach from cognitively inspired Artificial Intelligence that promises to solve the symbol grounding problem, with data points from a digital goal-setting intervention in a high-dimensional psychometric space of goal characteristics, accessed with the Goal Characteristics Questionnaire (GCQ) [2] in a web-based field-study of a digital goal-setting tool for higher education [3, 4, 5].

In this paper, we outline the backgrounds of Conceptual Spaces and show commonalities with the psychometric method of factor analysis. We then outline the application of these concepts in an ongoing field study collecting student goals in natural language, structured as hierarchical goals systems (HGS). The latter are tree-shaped sets of synergistic goals, in which every goal except the root goal serves a superordinate goal. The results section shows a subset of a large set of violin-, scatter-, and kernel-density estimation plots that will be used for dimension selection. Scores for 32 goal characteristics describe each goal, representing the subjective qualities of the goal for its bearer, who generated the scores by answering the items of the GCQ self-assessment tool. The dimensions are equivalent to a multidimensional goal characteristics space with goals as points. This perspective allows the application of clustering algorithms to identify distinct types of goals that German university students typically have.

1.1. Conceptual Spaces

Peter Gärdenfors mainly developed the idea of Conceptual Spaces as a solution for the symbol grounding problem and introduced it in his multifaceted book "Conceptual Spaces - The Geometry of Thought" [1]. Theorists produced proposals on how to deploy Conceptual Spaces and ground concepts. However, to evaluate the actual merit, there is a need to implement proofs-of-concept in real-world applications [6]. Classic examples of similarity dimensions are hue, saturation, and brightness for colors, pitch and amplitude for sounds, and width, height, and depth for geometrical size. These dimensions are, in a sense, orthogonal to each other, so one dimension cannot even partially be expressed by another dimension.

Suppose types are represented as convex regions in a quality space. In that case, there are, in principle, ways how to identify such regions: Either, in a top-down approach, a region or centroid of a region is defined based on the semantics of the space, or, in a bottom-up approach, data about entities in the feature space is collected, and clustering algorithms yielded convex regions. Labeling the new type is straightforward in the former case because the semantics are known. In the latter, the labeling of the new type is non-trivial. In this paper, we chose a bottom-up approach where we aim to identify clusters from data.

1.2. Psychometrics in general and factor analyses in particular

Psychometrics is a sub-discipline of psychology that aims to measure latent constructs, such as intelligence or conscientiousness. The psychometric method of factor analysis is used to identify a set of ideally uncorrelated unobserved factors that can explain the variance in observed variables. It is used, amongst others, to develop questionnaires with items as observed variables that measure latent variables. Ideally, those latent variables are not correlated and can be interpreted geometrically as orthogonal vectors spanning a feature space. So, what psychometric questionnaires and Conceptual Spaces have in common is that they are multidimensional, and the dimensions are intended to be non-redundant and orthogonal to each other.

1.3. Educational goals' characteristics

Goals have been described as internal representations of desired states [7], and Cognitive Scientists from various sub-disciplines have investigated them for decades. The result is a rich

toolbox of empirical methods for assessing goal characteristics relevant to goal pursuit and achievement. The goal characteristics Questionnaire (GCQ) [2] is a relatively young instrument that measures 32 dimensions of goal characteristics by self-assessment with Likert scale items. With its' 32 dimensions, the GCQ allows researchers to get a global view of the characteristics of goals. Among these dimensions are, for instance, *social support*, which describes how far others support the striving for a specific goal, or *self-congruence*, which describes how far a goal is in line with the identity and the lifestyle of its bearer.

1.4. Clustering in goal characteristics space

Goals and their GCQ scores can be interpreted as points in a high-dimensional (32d) space, which allows us to combine such data with concepts and methods from the Conceptual Spaces Framework. Suppose types of goals exist not only in scientific papers, hypothetical in nature, and invoked by empirical tricks, such as median splits. In that case, it should be possible to identify their –possibly fuzzy– borders in the goal characteristics space. Examples of such goal types from previous research are, for instance, *S.M.A.R.T goals*, which are specific, measurable, achievable, realistic, and time-bound [8], or *performance goals*, which are pursued to fulfill measurable, and visible external benchmarks. If clustering algorithms could identify clusters of goals that are semantically congruent with goal types from goal-setting research in a Conceptual Space, that would harden previous findings by novel methods.

1.5. Research questions

Can previously found concepts, such as goal types, be reproduced as clusters in relevant GCQ dimensions? If so, this empirical evidence would confirm their concordance with reality. If not, their practical value would be at least partially in question. Suppose borders between types are fuzzy, and the empirical methods applied so far rely on median splits. In that case, the concepts were relative to the data sets in question, and their semantics were not as absolute as they were in the case of reproduction by clustering methods.

In this paper, we take the first step by searching for areas of high density in one-dimensional violin plots, two-dimensional scatter plots, and kernel density estimation (KDE) plots of GCQ data. *Kernel density estimation* is a smoothing technique that makes it easier for observers to find patterns in a plot. The KDE algorithm is also used for smoothing the violin plots. The results can serve as the foundation for a dimension selection of the conceptual GCQ space before clustering algorithms, such as the distance-based k-Means algorithm and the density-based DBSCAN algorithm, will be applied to the data. The core benefit of the selection of a subset of dimensions is decreasing required computational resources and more optimal conditions for identifying clusters.

The central questions we address in this study are: Which dimensions of the GCQ are most suitable for clustering methods because they show more than one peak (visible in violin plots)? Which dimensions are redundant and hence can be excluded from clustering methods because they highly correlate with another dimension (visible in scatter plots)? Which dimension combinations are most promising for successful clustering because there is more than one peak in their bi-variate distribution (visible in kernel density estimation plots)?

2. Methods

The data was collected in an online study and analyzed with Python scripts, as described in the following subsections. It consists of goals in natural language, for which participants answered the GCQ, which results in scores for 32 dimensions. So far, the method applied to the data is the visual inspection of univariate and bivariate distributions of goal characteristics with a particular focus on multi-peaked distributions. We interpret multi-peaked distributions as indicators for distinct clusters in the respective dimensions. We will include those dimensions in future clustering computations and exclude dimensions with single-peaked distributions. Additionally, strong correlations between dimensions, showing up in two-dimensional plots as high data density along the main diagonal, indicate that one of the dimensions can be excluded from clustering computations to decrease the resource requirements.

2.1. Data collection

We implemented the technical setup of our web-based study with the Django web framework, a PostgreSQL database, and HTML, JavaScript, and CSS for the graphical user interface. We invited participants via mailing lists of the Institute of Cognitive Science and the Institute of Psychology, University of Osnabrück, and compensated with test-person hours and participation in a tombola for three Amazon vouchers of 10, 20, and 50 Euros. Participants were asked to formulate at least three educational goals. They chose one goal from the list and derived subgoals in the next step. Finally, the participants answered the Goal Characteristics Questionnaire for all goals. The preliminary data presented here includes 637 goals from 38 students. Data analysis scripts were implemented with the Python programming language.

3. Results

The resulting plots are 32 violin plots (one per GCQ dimension), 496 two-dimensional scatter and KDE plots [1], for each two-dimensional goal characteristic combination.

3.1. Violin plots for single GCQ dimensions

Violin plots are rotated kernel density estimation plots mirrored such that the shapes are amplified. Figure 1 shows violin plots for all 32 GCQ dimensions, ordered by subscales in each row. The *Content Specificity* plot (row 1, plot 1) is an example of a prospective candidate for clustering because it shows peaks at one and approximately 0.5. On the other side, the *Measurability* plot (row 1, plot 6) shows only one prominent peak, lowering the probability of successful clustering in this dimension.

---

[1] $496 = \sum_{n=1}^{n=32} n - 1 = 496$, as each of the 32 characteristics was plotted against each of the others and not itself.
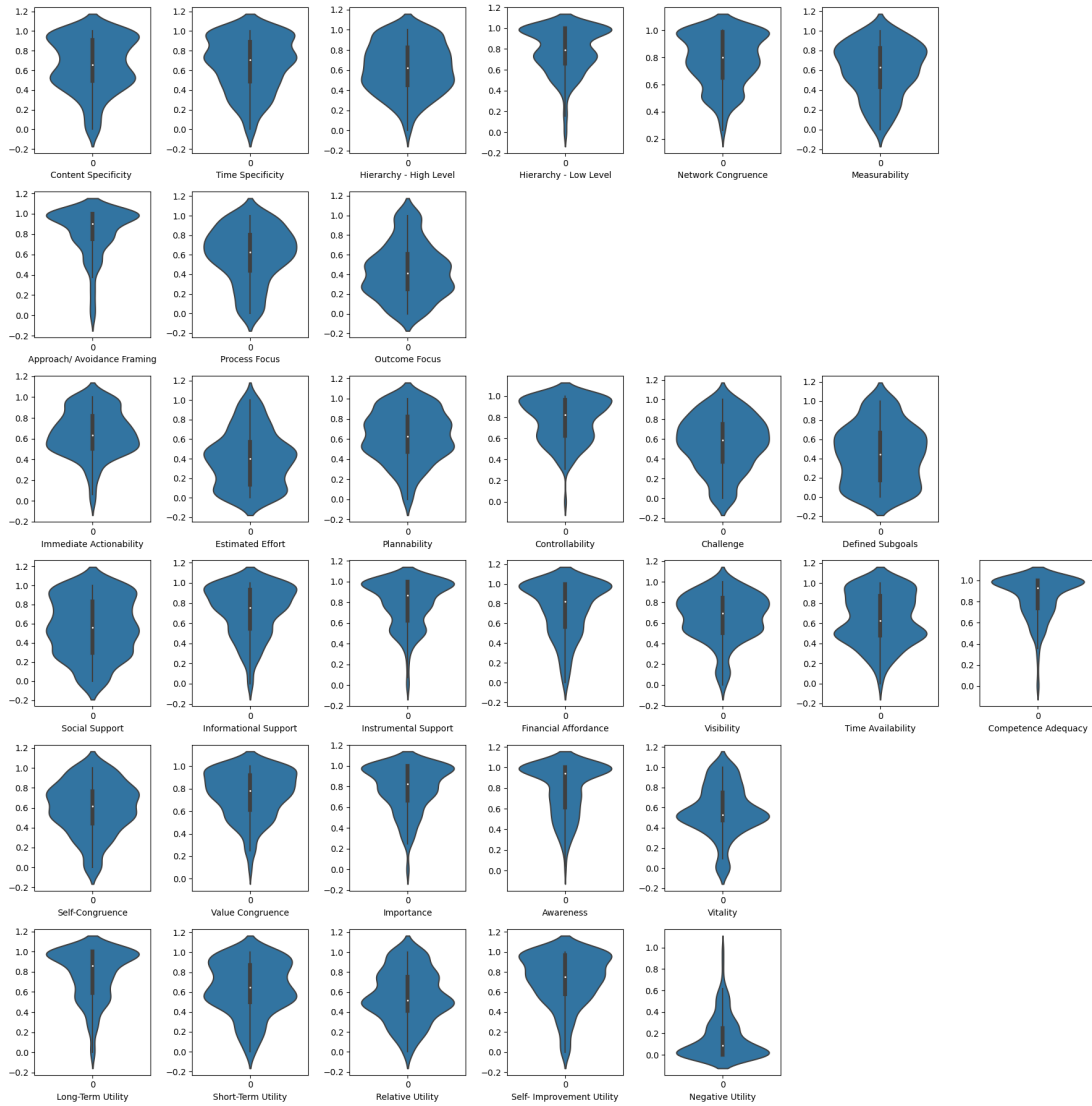
Figure 1: Violin-plots for all GCQ dimensions, ordered by rows according to the semantically ordered subscales of the GCQ. Violin plots are rotated and mirrored kernel density estimation plots. Multi-peaked plots indicate that distinct types of goals can be separated within the dimension

3.2. Scatter plots and kernel density estimation plots for all GCQ dimension combinations

Because of the high number of 496 scatter plots and 496 KDE plots, we include a representative exemplary subset of 10 scatter plots and 10 KDE plots illustrating the kind of derivable insights (see Figure 2). The KDE plots at the lower left are partially multi-peaked; for instance, the three plots of the *Social Support* Characteristic (row 3, plot 2; row 4, plot 2; row 5, plot 2). So this dimension should be kept because the probability of cluster detection is high. On the other
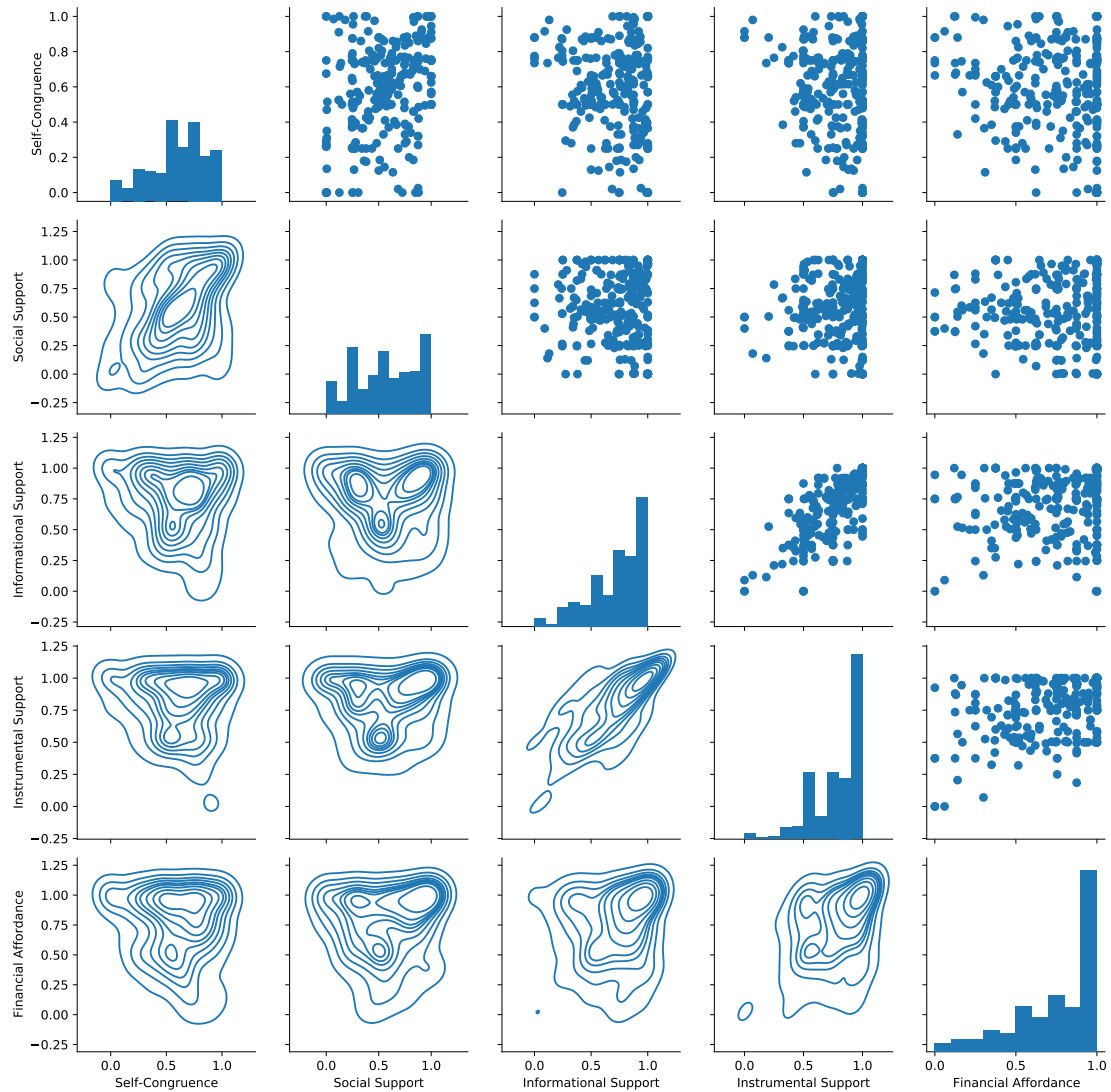
Figure 2: Pairwise scatter-plots and kernel density estimation (KDE) plots for five exemplary goal characteristics. *Informational Support* and *Instrumental Support* appear to have a linear relationship, as the KDE plot in row 4, plot 3, shows. Characteristic combinations with kernel density estimation plots showing more than one area with the highest density are prospective candidates for clustering algorithms.

hand, the dimensions *Informational Support* and *Instrumental Support* in row 4, plot 3 seem to be highly correlated, indicating that we should eliminate those. Figure 2 stands representative for the complete set of KDE plots and scatter plots insofar that there are not visually separable clusters in the entire data set but multi-peaked distributions in a subset.

These exemplary results show that there are, in fact, multi-peaked distributions in single characteristics and characteristic combinations. Therefore, we will pursue the outlined approach

and take consecutive steps towards dimension selection and goal type identification in the condensed goal characteristics space.

4. Discussion

In this paper, we have outlined the application of the Conceptual Spaces Framework to a real-world application, namely a high-dimensional data set of educational goals. We have outlined how to use clustering in a bottom-up approach to identify convex regions defining goal concepts in the goal characteristics space. Paradigmatic plots illustrated how the upcoming dimension selection procedure will be implemented and gave evidence for the existence of clusters in the data. In the next steps, a more robust statistical method for dimension selection will be established and applied.

Acknowledgments

References

[1] P. Gärdenfors, Conceptual Spaces: The Geometry of Thought, Cambridge, MA, 2000. URL: https://linkinghub.elsevier.com/retrieve/pii/S0001691800000676.

[2] G. Y. Iwama, F. Weber, M. Prentice, F. Lieder, Development and Validation of a Goal Characteristics Questionnaire (2021). doi:10.31219/osf.io/dzhwv.

[3] F. Weber, Goal Trees as Structuring Element in a Digital Data-Driven Study Assistant, in: D. G. Sampson, D. Ifenthaler, I. Pedro, M. L. Mascia (Eds.), 16th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2019), 19, IADIS Press, Cagliari, Italy, 2019, pp. 413–416. doi:10.33965/celda2019_201911c053.

[4] F. Weber, J. Schrumpf, T. Thelen, Development of a Digital Goal Setting Companion for Higher Education, in: A. Kienle, A. Harrer, J. M. Haake, A. Lingnau (Eds.), DeLFI 2021 - Die 19. Fachtagung Bildungstechnologien, Gesellschaft für Informatik e.V., Bonn, 2021, pp. 109–114. URL: https://dl.gi.de/handle/20.500.12116/36996.

[5] F. Weber, J. Kernos, M. Grenz, J. Lee, Towards A Web-Based Hierarchical Goal Setting Intervention for Higher Education, in: D. G. Sampson, D. Ifenthaler, P. Isaías (Eds.), 18th International Conference on Cognition and Exploratory Learning in the Digital Age, CELDA 2021, IADIS Press, Lisbon, 2021.

[6] L. Bechberger, Towards conceptual logic tensor networks, CEUR Workshop Proceedings 2969 (2021).

[7] J. B. Vancouver, J. T. Austin, Goal constructs in psychology: Structure, process, and content., Psychological Bulletin 120 (1996) 338–375.

[8] G. T. Doran, There's a S.M.A.R.T way to write management's goals and objectives, Management Review 70 (1981) 35–36.