An Elephant in the Dark

Creating Semantic Representations of Perceived Data with Conceptual Spaces

Hadi Banaee, Erik Schaffernicht, and Amy Loutfi

Centre for Applied Autonomous Sensor Systems, School of Science and Technology, Örebro University, SE-701 82 Örebro, Sweden
{hadi.banaee,erik.schaffernicht,amy.loutfi}@oru.se

Abstract. This paper discusses the task of creating semantic representations to describe numerical observations using conceptual spaces. The theory of conceptual spaces is considered as a semantic representation to conceptualise the perceived numerical information and to infer linguistic descriptions. We propose a data-driven approach to construct conceptual spaces from numerical data automatically. First, the elements of a conceptual space are derived based on a set of numerical observations in order to semantically represent the concepts of a given data set. This data-driven conceptual space is then employed for the task of semantic inference, in order to linguistically describe unknown perceived observations.

Keywords: Semantic representation · Conceptual spaces.

1 Motivation

Some Hindus bring an elephant to be exhibited in a dark room. A number of men touch and feel the elephant in the dark and, depending upon where they touch it, they believe the elephant to be like a water spout (trunk), a fan (ear), a pillar (leg) and a throne (back)... [1].

This is the beginning of an ancient parable, called *The Elephant in the Dark*, to demonstrate the problem of perception limitations. In this story, the individuals have their own perceptions of the elephant (an unknown concept for them) and therefore use their own *inference* to explain it. This is the problem of describing a concept based on the perceived information. The men sought to map or categorise the perceived information according to similar concepts that were known to them. However, their failure to successfully describe the concept of *Elephant* was due to the limitations of their sensory perceptions.

Describing unknown observations in natural language appears to be an easy task for humans. Both speakers and hearers have a great deal of common sense understanding of the concepts and properties that enable them to describe such observations. An example is the description of "Hippogriffs" in J.K. Rowling's Harry Potter books as: *Hippogriffs have the bodies, hind legs, and tails of horses, but the front legs, wings, and heads of eagles, with cruel beaks and large orange*

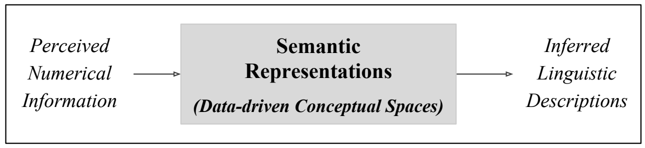

Fig. 1: Semantic representation for describing numerical data.

eyes [2]. This description uses familiar concepts that are most similar (eagle and horse), together with perceivable features (orange, large, etc.) that are understandable for humans. However, deriving descriptions for unknown concepts is no trivial task in *artificial intelligence* (AI). This task is especially crucial if the information given to the system is in the form of numeric or non-symbolic measurements (e.g., sensor data).

One goal of cognitive science is to construct artificial systems that can understand and model the cognitive activities of humans, such as concept learning and semantic inference [3]. However, a critical issue is how the given information is to be modelled in knowledge representation frameworks [4,5]. Concerning the task of the semantic description of concepts by means of perceived data, two aspects need to be considered: *induction* and *semantics*. Inductive inference performs a generalisation from a number of observations, which infers the characteristics of the concepts. Semantic inference is the process of inferring meaningful descriptions or truth conditions from semantically enriched information represented in logical or natural sentences. Neither symbolic, nor sub-symbolic approaches satisfactorily address these two AI problems simultaneously.

Consequently, the theory of *conceptual spaces* was introduced by Gärdenfors [6] as a mid-level representation to addressing both concept learning and semantic inference problems [3]. A conceptual space consists of a set of *quality dimensions* in various *domains*. These are placed within a geometrical structure in order to model, categorise, and represent the *concepts* [6]. This paper considers the task of semantic representation in describing the numerical observations. The semantic representation task investigates representational models in order to be able to bind perceived numerical data as input into a set of linguistic characterisations as output (See Fig. 1). Our claim here is that the conceptual spaces can be considered as a semantic representation to conceptualise the perceived numerical information and to be utilised to infer linguistic descriptions.

Conceptual spaces are principally derived in a *knowledge-driven* manner, on the assumption that there is prior knowledge from perceptual mechanisms or experts that manually initialise the elements of the conceptual space (i.e., domains, quality dimensions, and concepts' regions) [7]. However, the challenge discussed here is how to automatically construct a conceptual space from given information [8] to perform concept learning and semantic inference tasks. This is an important motivation, due to a growing class of problems that involves more complicated observations that have little or no prior knowledge concerning their semantic significance [9].

2 On the notion of Semantic Representation

The notion of a *semantic representation* has been used in a variety of ways in different areas such as knowledge representation in AI, cognitive science, and philosophy of language. Two prominent traditions for semantic representations exist [10]. One is to study the semantics of words by representing the relations of the words in natural language. For such representations, also called *amodal* approaches [11], the input is linguistic information. Another tradition focuses on conceptual structures for the representation of meanings, which considers the relations between concepts and percepts to model the semantics. In this case of semantic representations, also called *experiential* [12], the input is a set of perceptual information. The origin of this kind of semantic representation is the study of cognitive semantics, wherein the focus is on the meaning of the concepts as a cognitive phenomenon [13]. Cognitive semantics considers the meaning of linguistic expressions as mental entities coming from our perceptions. The perceptual information is then formed as concepts in our mind. This point of view is opposed by the *realist* approaches that define semantics as something out in the world [6]. Here semantics can be represented using e.g., abstract propositions and description logic, and can be modelled and verified by truth conditions. Within the cognitive semantics, however, the meaning is a conceptual structure that comes before the truth [6]. Semantic representations, from the cognitive point of view, should be a conceptual structure which represents both perceptual and linguistic information. In this work, the notion of a semantic representation follows the latter definition, by first constructing a conceptual representation using perceptual information, and then inferring semantically enriched descriptions. Therefore, a semantic representation of knowledge provides a conceptual structure for the meaning of perceived concepts [10]. This kind of representation eases the task of semantic reasoning of the perceived information.

3 Data-driven Conceptual Spaces as Semantic Representations

The main contribution of this study is to investigate the possibility of inferring human understandable semantics for any given data set through the data-driven conceptual spaces. This section explains how a conceptual space can be automatically constructed using the observed information, and then how such data-driven conceptual space can be utilised to infer semantic descriptions for any unseen observation. This process is then assessed by applying the approach on a data set of *leaf* examples. The formal definitions, the proposed algorithms, and the technical aspects of the framework are elaborated in detail in [14].

Identifying Quality Dimensions The origin of quality dimensions is still an open question in the field of conceptual spaces [6]. Once the process of constructing a conceptual space starts, as Quine noted in [15], some innate quality dimensions are needed to make *concept learning* possible. However, there is no unique

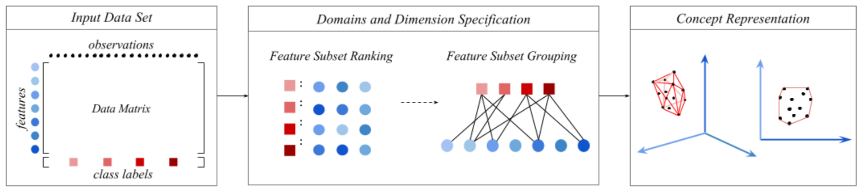

Fig. 2: Steps for constructing a conceptual space from numerical data.

way to specify which set of dimensions is sufficient to characterise the concepts. In many developed examples of conceptual spaces, determining the quality dimensions relies on the background knowledge. Phenomenal (human perceptual) quality dimensions are usually *chosen* by the experts, and the scientific (sensory) quality dimensions are usually *inferred* from the perceived behaviours [6]. However, this issue is more challenging when dealing with systems where there is no prior knowledge to explain the semantics of dimensions. An agreed point in the literature of conceptual spaces is that it is almost impossible to provide a complete list of human perceptual quality dimensions [16].

Construction of Conceptual Spaces The framework we propose provides a procedure to utilise machine learning algorithms for the task of identifying relevant features and concepts in a numerical data set, to specify the domains and quality dimensions of a conceptual space in a data-driven manner. Our underlying assumption for the use of machine learning techniques is that highly discriminative and distinctive features are adequate choices for quality dimensions and domains, since they allow clear separation of the different concepts.

To identify those discriminative and distinctive features, we use information-theoretic measures like joint mutual information to rank the relevance of a feature in relation to each concept. We represent the feature-concept associations in a weighted bipartite graph and use a heuristic search, based on finding maximum bi-cliques, to group high-ranked features, which are then chosen as domains for the different concepts. After determining the quality dimensions and domains, the concept representation is constructed from the available instances. Thereto, two properties are estimated: the concept's convex regions and the concept's salient weights in relation to the quality dimensions. This calculation is formulated based on the associated observations to the concepts, without involving the external knowledge. Fig. 2 illustrates the steps of constructing a conceptual space from a set of numerical data in a data-driven manner.

Semantic Inference in Conceptual Spaces The semantic inference process is introduced to linguistically represent a new observation within the built conceptual space. First, a symbol space is introduced which includes the semantics of the corresponding concepts and quality dimensions. Then, the inference of linguistic descriptions for an unknown observation is performed in two phases: (1)

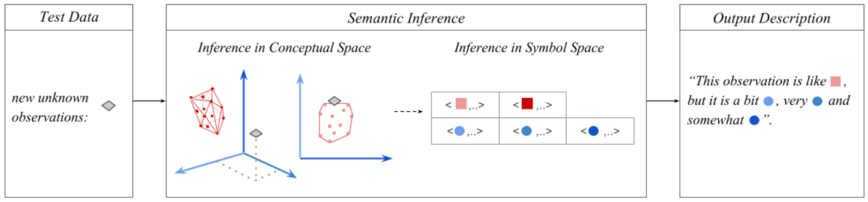

Fig. 3: Steps of semantic inference in constructed conceptual spaces.

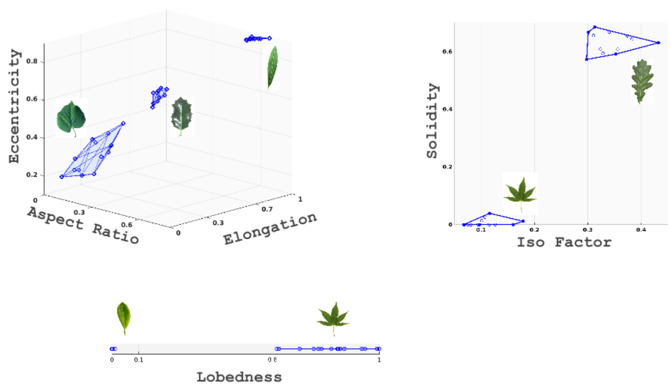

Fig. 4: The conceptual space of leaf data set.

Inclusion: the new instance is localised into the built conceptual space to determine its associated concepts and quality dimensions. This determination is done by considering the inclusion of the instance and the use of similarity measures in such space. (2) Realisation: then, the lexicalisation of the instance is induced by extracting the semantic labels of the associated concepts and quality dimensions. Fig. 3 illustrates the steps of semantic inference for a new observation within a constructed conceptual space.

A Case Study: Leaf Data Set The plausibility of the proposed approach is tested using a set of leaf samples. The output conceptual space is then used to infer linguistic descriptions for a set of new leaves. Fig. 4 shows the derived conceptual space of six leaf concepts. The approach has specified six quality dimensions that are grouped in three domains within the data-driven constructed conceptual space. By applying the semantic inference on the conceptual space of leaves, an unknown perceived leaf can be located within the space, and then be characterised by its associated concepts and quality dimensions. For example, a new leaf can be linguistically described as: "*This unknown leaf is like **Japanese Maple** leaves, but it is **oval** with a **lobed** margin.*"

4 Related Work on Conceptual Spaces and AI

The aim of representing knowledge in a conceptual space is to develop an intuitive interpretation of the relationship between symbolic and sub-symbolic information [6,3]. Gärdenfors has discussed thoroughly the role of conceptual spaces as a knowledge representation framework in AI systems [5], focusing on the tasks of *induction* and *reasoning* [17]. Concept formation tightly connects the theory of conceptual spaces to the *learning* problem. Many approaches for learning are typically performed by *connectionist* approaches (i.e., *neural network architectures* [18]). But such solutions neglect the explainability of the involved concepts or the learned model itself. In addition to the theoretical AI problems, the feasibility of using conceptual spaces has been studied in various application domains of AI, such as geographical measurement [19], cognitive robotics [20,21], and visual perception [22]. Using data mining approaches in the process of deriving conceptual spaces has been studied in a few isolated works. Keßler [8] outlined the idea of using conceptual spaces to describe data, with some discussions on the possibility of automatically generating such spaces from databases. Lee [23] proposed a data mining method coupled with conceptual spaces, which addresses cognitive tasks such as concept formation using clustering techniques. The main drawback of these approaches is that they rely on knowing about the semantics of an application area beforehand in order to directly determine the domains and the quality dimensions.

5 Conclusions

This paper presents the notion of data-driven conceptual spaces as a tool for creating semantic representations in order to linguistically describing numerical data. The proposed approach holds for certain classes of problems. It explores applications wherein the input data is difficult to interpret at first glance. Within such applications, the task of specifying the interpretable domains and dimensions based on human perceptions is non-trivial. These classes of problems usually deal with raw sensor data (sometimes multi-variate data) with little or no prior knowledge about their semantics [9]. One issue of constructing conceptual spaces in a data-driven way is the semantics of domains. Feature grouping method is based on how well a subset of the features distinctly represents the various concepts. However, there still exists the problem of verifying the semantic dependency of the quality dimensions within a domain. Regarding this problem, Gärdenfors in [6] suggests that the verification of deciding whether two quality dimensions are integral or not can be done by empirical testing based on the expert judgements, and not necessarily using analytical techniques. It is seemingly difficult to realise the semantic dependency of the features analytically. For example, values of RGB as the dimensions of the colour domain do not indicate their semantic relations. Indeed, solving the issue of domain specification can lead to forming a general solution to the problem of determining an evaluation criterion to choose between competing conceptual spaces, an issue raised by Gärdenfors in [17].

References

1. Rumi, J.a.D., Arberry, A.J.: Tales from the Masnavi. Curzon paperbacks. Curzon Press (1993)
2. Rowling, J.K.: Harry Potter And The Prisoner Of Azkaban. New York: Arthur A. Levine Books (1999)
3. Aisbett, J., Gibbon, G.: A general formulation of conceptual spaces as a meso level representation. Artificial Intelligence **133**(1) (2001) 189–232
4. Gärdenfors, P.: Symbolic, conceptual and subconceptual representations. In: Human and Machine Perception. Springer (1997) 255–270
5. Gärdenfors, P.: Conceptual spaces as a framework for knowledge representation. Mind and Matter **2**(2) (2004) 9–27
6. Gärdenfors, P.: Conceptual spaces: The geometry of thought. MIT press (2000)
7. Rickard, J.T., Aisbett, J., Gibbon, G.: Reformulation of the theory of conceptual spaces. Information Sciences **177**(21) (2007) 4539–4565
8. Keßler, C.: Conceptual spaces for data descriptions. In: The cognitive approach to modeling environments (CAME), workshop at GIScience. (2006) 29–35
9. Rickard, J.T.: A concept geometry for conceptual spaces. Fuzzy optimization and decision making **5**(4) (2006) 311–329
10. Griffiths, T.L., Steyvers, M., Tenenbaum, J.B.: Topics in semantic representation. Psychological review **114**(2) (2007) 211
11. Feng, Y., Lapata, M.: Visual information in semantic representation, Association for Computational Linguistics (2010) 91–99
12. Vigliocco, G., Meteyard, L., Andrews, M., Kousta, S.: Toward a theory of semantic representation. Language and Cognition **1**(2) (2009) 219–247
13. Allwood, J.S., Gärdenfors, P.: Cognitive semantics: Meaning and cognition. Volume 55. John Benjamins Publishing (1999)
14. Banaee, H.: From Numerical Sensor Data to Semantic Representations : A Data-driven Approach for Generating Linguistic Descriptions. PhD thesis, Örebro University, School of Science and Technology (2018)
15. Quine, W.V.O.: Ontological relativity and other essays. Number 1. Columbia University Press (1969)
16. Gärdenfors, P.: Semantics based on conceptual spaces. In: Indian Conference on Logic and Its Applications, Springer (2011) 1–11
17. Gärdenfors, P.: Induction, conceptual spaces and ai. The Dynamics of Thought (2005) 109–124
18. Sun, R.: Artificial intelligence: Connectionist and symbolic approaches. (1999)
19. Adams, B., Raubal, M.: Conceptual space markup language (csml): Towards the cognitive semantic web. In: ICSC. (2009) 253–260
20. LeBlanc, K.: Cooperative anchoring: sharing information about objects in multi-robot systems. (2010)
21. Chella, A., Frixione, M., Gaglio, S.: Anchoring symbols to conceptual spaces: the case of dynamic scenarios. Robotics and Autonomous Systems **43**(2) (2003) 175–188
22. Chella, A., Frixione, M., Gaglio, S.: A cognitive architecture for artificial vision. Artificial Intelligence **89**(1-2) (1997) 73–111
23. Lee, I.: Data mining coupled conceptual spaces for intelligent agents in data-rich environments. In: International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Springer (2005) 42–48