What Machine Learning Tells Us About the Mathematical Structures of Concepts

Jun Otsuka (jotsuka@bun.kyoto-u.ac.jp / Kyoto University & RIKEN Center for Advanced Intelligence Project)

PDF & more

Question: what kind of mathematical structure is needed to model *concepts*?

"What are concepts?" is one of the fundamental questions in philosophy, where Aristotle, Kant, Wittgenstein, etc. have proposed different models of concepts. Meanwhile, machine learning literature has developed powerful methods to learn *representation* from data, as well as mathematical models to analyze such representations. This presentation aims to bridge between these two traditions by identifying mathematical models & machine-learning counterparts of the (1) Aristotelian, (2) Wittgensteinian, (3) Functional, and (4) Symmetry-based theories of concept.

Aristotelian Abstractionism

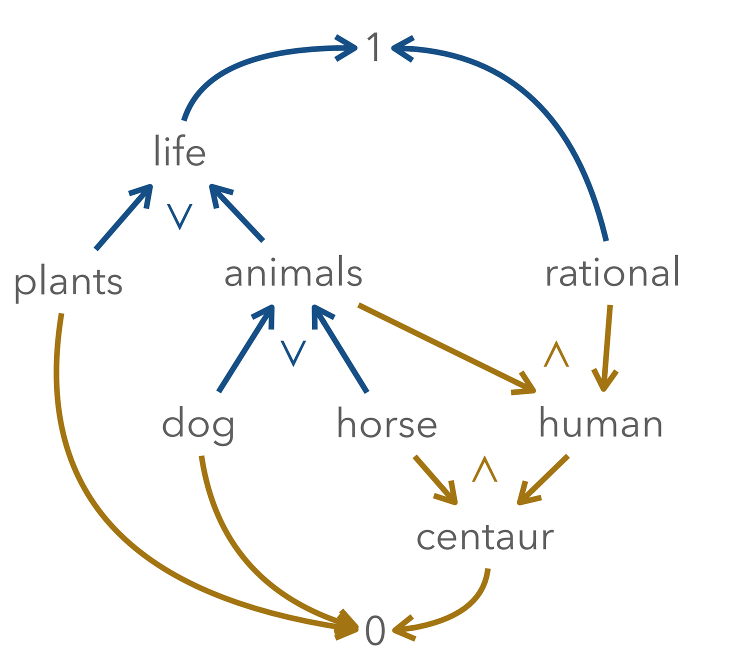
Idea: concepts are formed by extracting common properties of individual things.

Cf. Aristotle, Locke, etc. (see Heis 2008).

Modeled by...

Boolean Algebra

with abstraction ∨ and instantiation ∧ operators



Pros & Cons

- ✓ Intuitive
- ✓ Well-studied syntax & semantic relationship (Stone duality)
- ✗ Allows arbitrary abstraction (e.g. beef ∨ cherry = red juicy food?)
- ✗ Difficult to learn from data (GOFAI)

Cluster Concept Theory

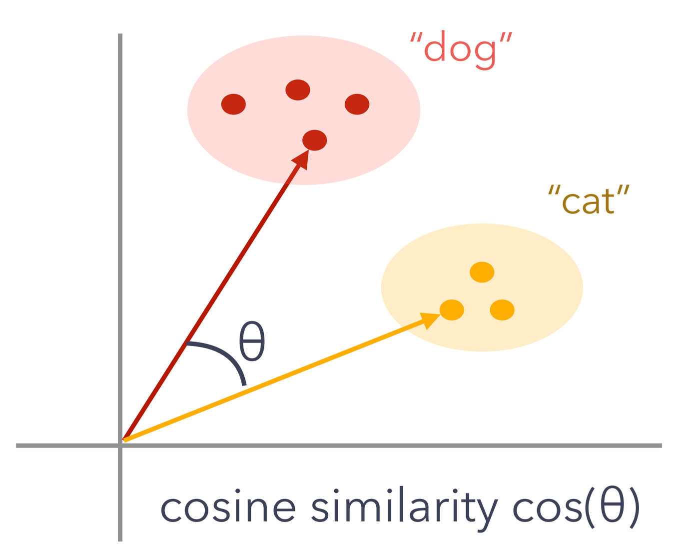
Idea: concepts are groups of similar things or impressions.

Cf. Wittgenstein's family resemblance; Rosch's prototype theory.

Modeled by...

Vector Spaces

Concepts are vectors in a high-dim vector space, with their similarity being measured by some metric, e.g. cos(θ).



Pros & Cons

- ✓ Easy to learn from data (e.g. word-embeddings such as word2vec)
- ✓ Allows some algebraic operations (e.g. king - man + women = queen)
- ✗ Difficult to deal with other logical connectives, such as negation or disjunction.
- ✗ Difficult to learn complex expressions (e.g. sentences, etc.)

Functional Model

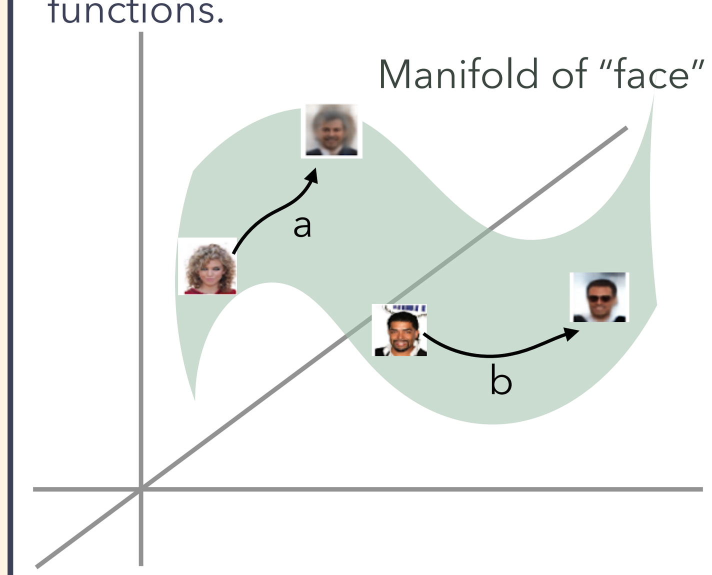
Idea: concepts are a set of properties having a certain functional relationship.

Cf. Lotze & Cassirer's Funktionsbegriff; Gärdenfors's topological concept.

Modeled by...

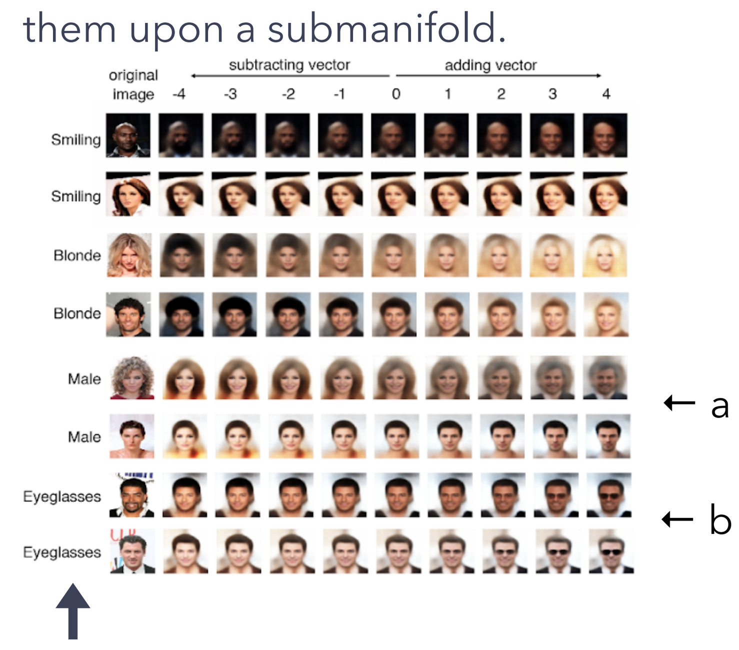
Manifolds

Possible data points "live" on a submanifold in a non-euclidean space, where the submanifold (=concept) structure is determined by a set of functions.



Pros & Cons

- ✓ Learnable from data (e.g. generative models such as VAE)
- ✓ One can create new images by sliding them upon a submanifold.



each dimension of the submanifold represents distinct characteristics?

- ✗ These characteristics are not always distinct but are sometimes "entangled"

Symmetry

Idea: concepts are a set of properties invariant under a certain group of transformations.

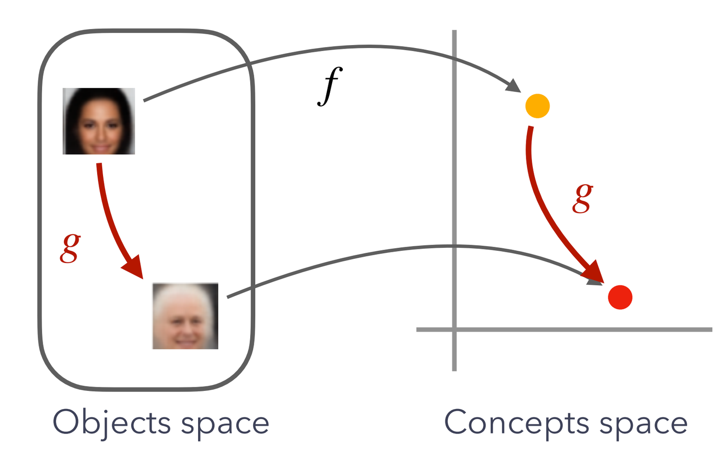
Cf. Berkeley, van Fraassen, Jantzen

Modeled by...

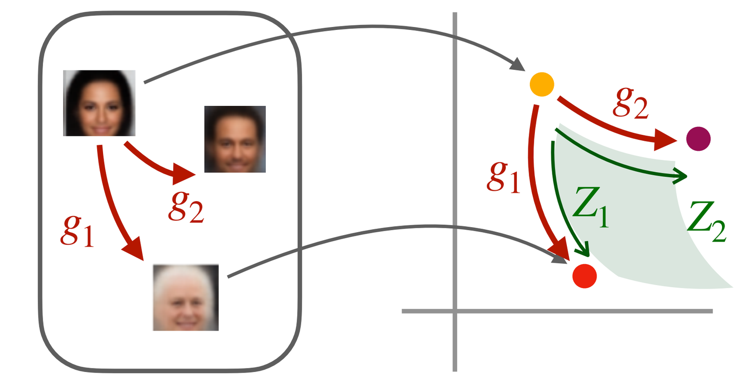
Groups

Representations transform in accordance with changes in objects, i.e. the representation map f *commutes* with object transformations/groups g, s.t.

$$g \cdot f(x) = f(g \cdot x)$$



Moreover, representations must be *disentangled*, so that one can tweak one characteristic without affecting others (each dimension encodes a distinct characteristic).



- There is a decomposition $Z = Z_1 \times \ldots \times Z_n$ of the latent space corresponding to the decomposition $G = G_1 \times \ldots \times G_n$ of transformations
- s.t. G_i acts on Z_i but not other Z_j, $(i \neq j)$

(Higgins+ 2018)

Geometrical vs. Algebraic sides of concepts

Geometrical

- "Similarity" measured by a metric on the conceptual space (Wittgenstein)
- A "concept" must form a connected region (Gärdenfors)
- The space of possible concepts must be constrained (Lotze, Cassirer)

Algebraic

- Concepts are related to each other by algebraic operations (abstraction, addition, negation, transformation, etc.)
- Each concept embodies a set of rules that prescribe how its properties are transformed.

An adequate mathematical framework for modeling concepts must combine their geometrical & algebraic aspects.

References

- Gärdenfors, P. (2000), *Conceptual Spaces*, MIT Press
- Heis, J. (2008), *The Fact of Modern Mathematics: Geometry, Logic, and Concept Formation in Kant and Cassirer,* PhD diss.
- Higgins, I. et al. Towards a Definition of Disentangled Representations. arXiv [cs.LG] (2018)
- Jantzen, BC. (2015), Projection, symmetry, and natural kinds, *Synthese* 192(11)
- van Fraassen, BC. (1989). *Laws and Symmetry*, OUP

Advertisement