CATEGORICAL INFORMATION THEORY

MSS COLLOQUIUM AT UIUC
2011/11/29

DAVID I. SPIVAK

ABSTRACT. Classical information theory, as developed by Claude Shannon in 1948, studies how to optimize the quantity of data that can be transmitted across a noisy channel, but it ignores what this information "means". It is clear that information is intended to mean or signify something – this is its only purpose – but how can such a thing as meaning be formalized? In this talk I'll discuss how category theory may be useful in this endeavor. Going further, we may postulate that information is always intended as a communication from one party to another (perhaps from an entity to a later version of itself). In this case we can ask, "what is the relationship between the structure of the information being transferred and the structure of the two communicating parties?" I will outline a possible answer to this question in the form of a category-theoretic communication protocol for transferring information between parties.

No prior knowledge of category theory will be necessary to understand this talk.

I. Introduction
- A. What is Information?
  - 1. "Information theory" vs. Meaning
  - 2. Examples – dictionary, engineers schematic diagram, database
  - 3. Information vs. data
  - 4. Relational, not absolute
- B. Historical context
  - 1. Frege – Begriffsschrift
  - 2. Hilbert: "This formula game is carried out according to certain definite rules, in which the technique of our thinking is expressed. [...] The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds" (Hilbert, 1928, 475).
  - 3. Dan Kan: "Information is inherently a combinatorial affair." Facts are fit together in appropriate ways to create new facts.
- C. Category theory
  - 1. History
    - a. Invented to connect topology and algebra
    - b. Became unifying language for lots of math
    - c. Successfully applied in CS – programming languages
  - 2. Algebraic geometry
    - a. Grothendieck and Weil conjectures (proved power of CT)
    - b. Idea: sheaves (hard to talk about without categories)
    - c. Local sections (we'll talk more about this)
- D. Databases
  - 1. Why Databases? Information!
  - 2. Example – tables connected in a specific way
  - 3. I'll give a simple CT formulation.
- E. Outline of the talk
  - 1. Basic category theory
  - 2. Databases
  - 3. Change of schema
  - 4. Queries
  - 5. Communication (time permitting)

II. Basic category theory
- A. Categories

    1. A category $\mathcal{C}$ is a graph in which paths can be declared equivalent
        a. A set $\mathbf{Ob}(\mathcal{C})$
        b. A set of arrows $\mathbf{Arr}(\mathcal{C})$ each with a source and target object.
        c. An equivalence relation on paths ($\mathbf{Path}(\mathcal{C})$), respecting source and target, and pre-served under composition.
    2. Examples of categories
        a. Random thing.
        b. $E \rightrightarrows V$
        c. $\mathbf{Set}$ (saturated)
        d. Topological spaces, manifolds, rings, etc.
  B. Functors
    1. A functor $F\colon \mathcal{C} \to \mathcal{D}$ is a mapping between categories
        a. A map $\mathbf{Ob}(\mathcal{C}) \to \mathbf{Ob}(\mathcal{D})$
        b. A map $\mathbf{Arr}(\mathcal{C}) \to \mathbf{Path}(\mathcal{D})$
        (1) If $\mathcal{D}$ is saturated then we may assume $\mathbf{Arr}(\mathcal{C}) \to \mathbf{Arr}(\mathcal{D})$.
        c. Equivalent paths in $\mathcal{C}$ are sent to equivalent paths in $\mathcal{D}$.
    2. Examples
        a. $\mathbf{Rings} \to \mathbf{Set}$
        b. $\boxed{E \rightrightarrows V} \to \boxed{N \circlearrowleft}$
        c. $\boxed{E \rightrightarrows V} \to \mathbf{Set}$. (Draw tables)
        d. $\boxed{N \circlearrowleft} \to \mathbf{Set}$
III. Databases
 A.

(1)
$$\mathcal{C} :=$$
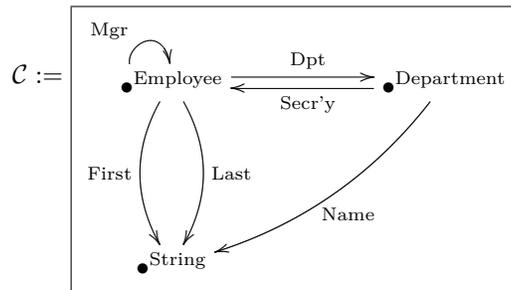
  B. $\mathcal{C}-\mathbf{Set}$ is a topos
    1. Morphisms are called natural transformations
    2. Subobjects, unions, products, quotients
    3. Logic (typed lambda calculus)
  C. The graph example $\boxed{E \rightrightarrows V}$ in terms of databases
    1. Morphisms of graphs
IV. Morphism of schemas
  A. Example: $\boxed{E \rightrightarrows V} \to \boxed{N \circlearrowleft}$
    1. Pulling back data
    2. Pushing forward data?
  B. For every functor $F\colon \mathcal{C} \to \mathcal{D}$ exists
    1. $\Delta_F\colon \mathcal{D}-\mathbf{Set} \to \mathcal{C}-\mathbf{Set}$,
    2. $\Sigma_F\colon \mathcal{C}-\mathbf{Set} \to \mathcal{D}-\mathbf{Set}$, and
    3. $\Pi_F\colon \mathcal{C}-\mathbf{Set} \to \mathcal{D}-\mathbf{Set}$.
  C. DB example

1.

(2)
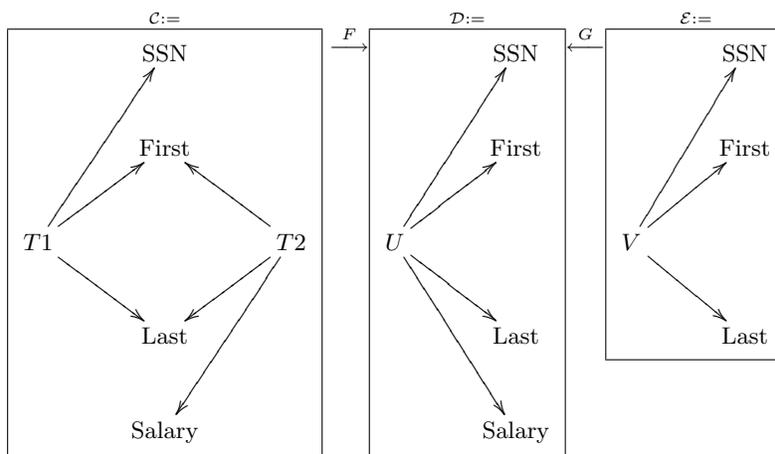
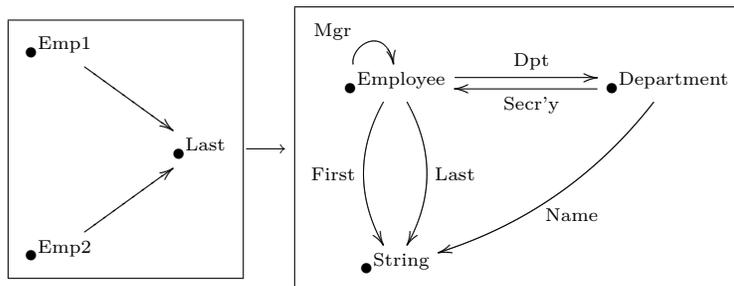
V. Queries
- A. Grothendieck construction
  1. Given $\delta\colon \mathcal{C} \to \mathbf{Set}$, get $\int \delta \to \mathcal{C}$
  2. Objects: $\{(c,x) | c \in \mathbf{Ob}(\mathcal{C}), x \in \delta(c)\}$
  3. An arrow $(c,x) \to (c',x')$ for every $f\colon c \to c'$ such that $f(x) = x'$.
  4. In other words, objects are id-cells, arrows are other cells.
  5. The fiber over each object is its set of instances.
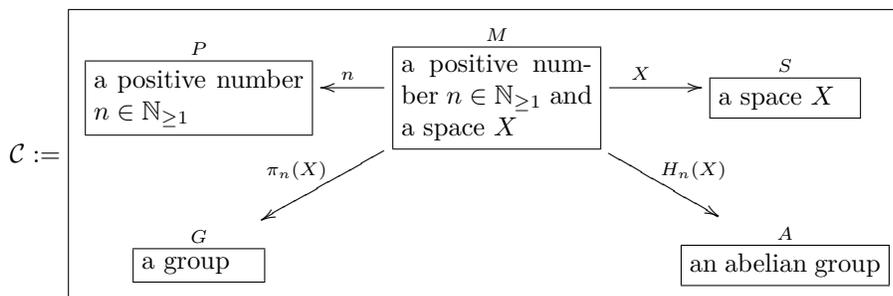- B. Local sections
  1. E.g.

  2. A lift is the same as two employees with same last name.
  3. Find two employees, one named Bob, one named Sue, that have the same manager:
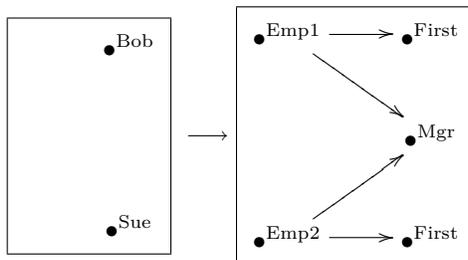
  4. Same idea: I have a topological space X. I know that $\pi_1(X) = pi_2(X) = 0, \pi_3(X) = \mathbb{Z}$ and $H_1(X) = 0$.

      5. Like integer sequence database.

VI. Communication

    A. Basic idea:

        1. Everyone has a schema and data

        2. Interacting groups have their own schema and data mapping to each subgroup

        3. A simplicial complex of interactions.

    B. There is a nice model for communicating new information

        1. Can be done within category theory.

        2. See me afterwards if interested.