A Category Theoretical Investigation of the Type Hierarchy for Heterogeneous Sensor Integration

CLIFF JOSLYN, EMILIE PURVINE, MICHAEL ROBINSON

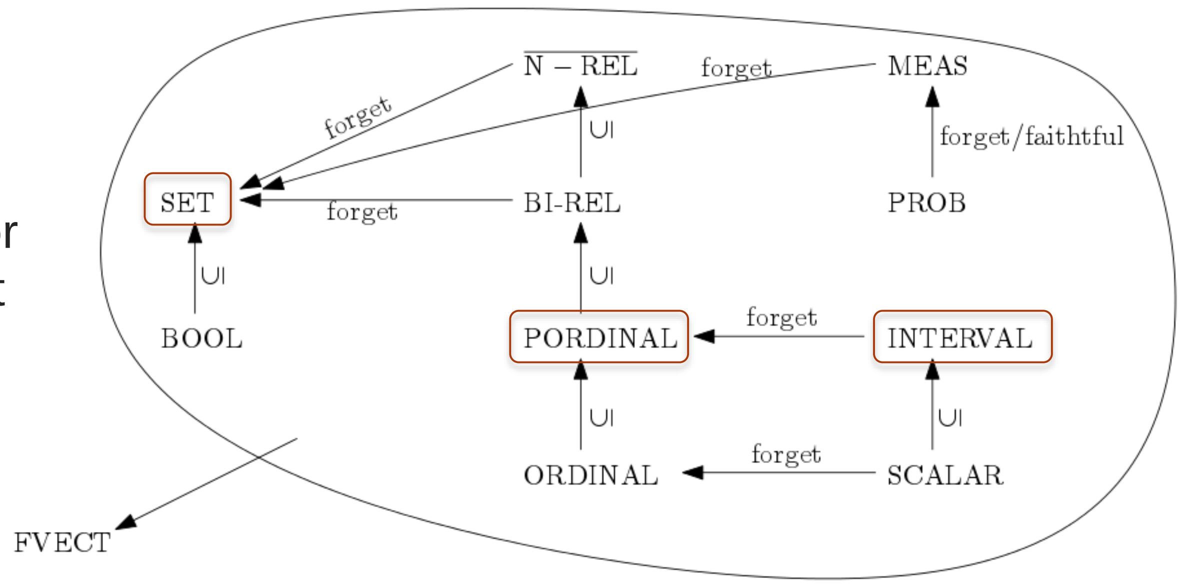
Pacific Northwest National Laboratory American University

Outline

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Setting the stage – information integration scenario [Michael Robinson's flash talk]
- Our strategy to encode data into a vector space
- Examples
  - SET
  - PORDINAL
  - INTERVAL – a work in progress

Data type hierarchy



- Data types fall into a few categories – literally!
  - Can create categories for each type of data so that stalks are objects in that category
- Transform between data types using *functors*
  - Lose some information e.g., SCALAR → ORDINAL we lose addition
- Transformation to FVECT allows us to turn any data type into a finite dimensional vector space, keeping structure

Emilie Purvine, Cliff Joslyn, Michael Robinson. *A Category Theoretical Investigation of the Type Hierarchy for Heterogeneous Sensor Integration*. https://arxiv.org/abs/1609.02883

Our strategy

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Given a single stalk, or data type, e.g., set, boolean, partial ordinal, ordinal, interval, scalar, probability, binary relation…
  - 1. Create a category, C, where the stalk is an object.
    - E.g., if a stalk is a totally ordered set then it is an object in the category ORDINAL

Our strategy

- Given a single stalk, or data type, e.g., set, boolean, partial ordinal, ordinal, interval, scalar, probability, binary relation…
  - 1. Create a category, C, where the stalk is an object.
    - E.g., if a stalk is a totally ordered set then it is an object in the category ORDINAL
  - 2. Given a single object $c \in C$ (your stalk), find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$

Our strategy

- Given a single stalk, or data type, e.g., set, boolean, partial ordinal, ordinal, interval, scalar, probability, binary relation…
  - 1. Create a category, C, where the stalk is an object.
    - E.g., if a stalk is a totally ordered set then it is an object in the category ORDINAL
  - 2. Given a single object $c \in C$ (your stalk), find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - 3. Create a vector space V with basis elements from $Decat(D)$

Our strategy

- Given a single stalk, or data type, e.g., set, boolean, partial ordinal, ordinal, interval, scalar, probability, binary relation…
  - 1. Create a category, C, where the stalk is an object.
    - E.g., if a stalk is a totally ordered set then it is an object in the category ORDINAL
  - 2. Given a single object $c \in C$ (your stalk), find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - 3. Create a vector space V with basis elements from $Decat(D)$

$$\begin{tikzcd} c \arrow[r, "\text{surjection}"] \arrow[rr, bend right, "\text{``categorification''}"'] & Decat(D) \arrow[r] & V \end{tikzcd}$$

(Probably should be “vector space-ification”)

Our strategy

- Given a single stalk, or data type, e.g., set, boolean, partial ordinal, ordinal, interval, scalar, probability, binary relation…
  - 1. Create a category, C, where the stalk is an object.
    - E.g., if a stalk is a totally ordered set then it is an object in the category ORDINAL
  - 2. Given a single object $c \in C$ (your stalk), find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - 3. Create a vector space V with basis elements from $Decat(D)$

$$
\begin{tikzcd}
c \arrow[r, "\text{surjection}"] \arrow[rr, bend right, "\text{“categorification”}"'] & Decat(D) \arrow[r] & V
\end{tikzcd}
$$

(Probably should be “vector space-ification”)

- Bonus: if we put structure of c into D via morphisms then we can get a map $V \to V$ which encodes this structure – the adjacency matrix of morphisms in Decat(D)

Example – categorical data

c = {police officer, protester, bystander}

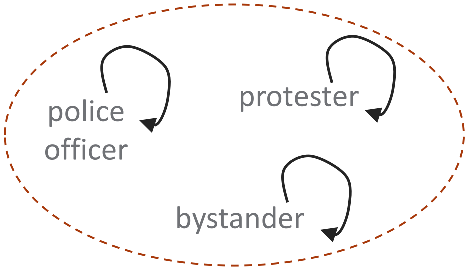
- 1. Create a category, C, where stalk c is an object.
  - C = SET as defined previously, clearly c is an object in SET
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
- 3. Create a vector space V with basis elements from $Decat(D)$

Example – categorical data

c = {police officer, protester, bystander}

- 1. Create a category, C, where stalk c is an object.
  - C = SET as defined previously, clearly c is an object in SET
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - $Ob(D)$ = {police officer, protester, bystander}
  - $Hom_{SET}(x, y) = \begin{cases} \{id_x\} & x = y \\ \emptyset & else \end{cases}$
  -

$D =$



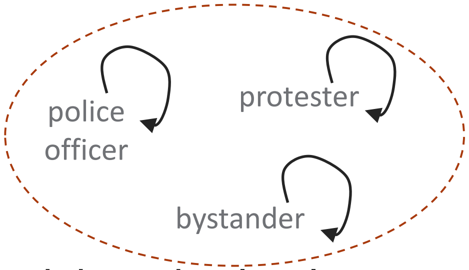
- 3. Create a vector space V with basis elements from $Decat(D)$

Example – categorical data

$$c = \{\text{police officer, protester, bystander}\}$$

- 1. Create a category, C, where stalk c is an object.
  - C = SET as defined previously, clearly c is an object in SET
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - $Ob(D)$ = {police officer, protester, bystander}
  - $Hom_{SET}(x, y) = \begin{cases} \{id_x\} & x = y \\ \emptyset & else \end{cases}$
  - Then $Decat(D)$ = { {police officer}, {protester}, {bystander} } and there is clearly a surjection (bijection in this case) from c to $Decat(D)$

$D =$



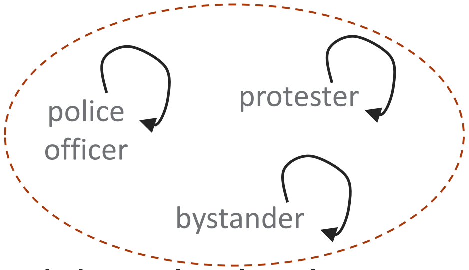
- 3. Create a vector space V with basis elements from $Decat(D)$

Example – categorical data

c = {police officer, protester, bystander}

- 1. Create a category, C, where stalk c is an object.
  - C = SET as defined previously, clearly c is an object in SET
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - $Ob(D)$ = {police officer, protester, bystander}
  - $Hom_{SET}(x,y) = \begin{cases} \{id_x\} & x = y \\ \emptyset & else \end{cases}$
  - Then $Decat(D)$ = { {police officer}, {protester}, {bystander} } and there is clearly a surjection (bijection in this case) from c to $Decat(D)$

$D =$

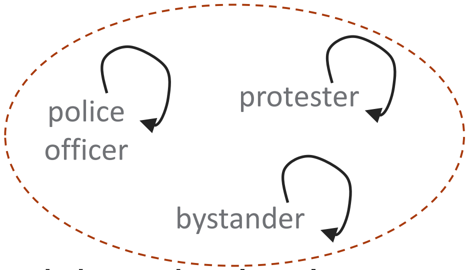


- 3. Create a vector space V with basis elements from $Decat(D)$
  - V = $\mathbb{R}$[{ {police officer}, {protester}, {bystander} }]
    = { α ·{police officer} $+\beta$ ·{protester} $+\gamma$ ·{bystander} : $\alpha, \beta, \gamma \in \mathbb{R}$ }

Example – categorical data

c = {police officer, protester, bystander}

- 1. Create a category, C, where stalk c is an object.
  - C = SET as defined previously, clearly c is an object in SET
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - $Ob(D)$ = {police officer, protester, bystander}
  - $Hom_{SET}(x,y) = \begin{cases} \{id_x\} & x = y \\ \emptyset & else \end{cases}$
  - Then $Decat(D)$ = { {police officer}, {protester}, {bystander} } and there is clearly a surjection (bijection in this case) from c to $Decat(D)$

$D =$ 

- 3. Create a vector space V with basis elements from $Decat(D)$
  - V = ℝ[{ {police officer}, {protester}, {bystander} }]
    = { α ·{police officer} +β ·{protester} +γ ·{bystander} : $\alpha, \beta, \gamma \in \mathbb{R}$ }
- Bonus: There is no structure on c (or any categorical) which is seen by the absence of morphisms other than identities in D. Map $V \rightarrow V$ is just the identity matrix in this case

Example – categorical data (cont.)

- Replace stalk {police officer, protester, bystander} with new stalk $\mathbb{R}$[{ {police officer}, {protester}, {bystander} }]
- Now stalk is a vector space!

Example – categorical data (cont.)

- Replace stalk {police officer, protester, bystander} with new stalk $\mathbb{R}$[{ {police officer}, {protester}, {bystander} }]
- Now stalk is a vector space!

- Problem: what if our analysis indicates that
  $2\cdot${police officer} $+3\cdot${protester} $+1.9\cdot${bystander}
  is part of a globally consistent assignment? How do we interpret this?

Example – categorical data (cont.)

- Replace stalk {police officer, protester, bystander} with new stalk $\mathbb{R}$[{ {police officer}, {protester}, {bystander} }]
- Now stalk is a vector space!

- Problem: what if our analysis indicates that
  $2\cdot${police officer} $+3\cdot${protester} $+1.9\cdot${bystander}
  is part of a globally consistent assignment? How do we interpret this?
  - There can't be any globally consistent assignments

Example – categorical data (cont.)

- Replace stalk {police officer, protester, bystander} with new stalk $\mathbb{R}$[{ {police officer}, {protester}, {bystander} }]
- Now stalk is a vector space!

- Problem: what if our analysis indicates that
  $2 \cdot$ {police officer} $+3 \cdot$ {protester} $+1.9 \cdot$ {bystander}
  is part of a globally consistent assignment? How do we interpret this?
  - There can’t be any globally consistent assignments
  - There are 2 police officers, 3 protesters, and … 1.9 (???) bystanders

Example – categorical data (cont.)

- Replace stalk {police officer, protester, bystander} with new stalk $\mathbb{R}$[{ {police officer}, {protester}, {bystander} }]
- Now stalk is a vector space!

- Problem: what if our analysis indicates that $2\cdot$ {police officer} $+3\cdot$ {protester} $+1.9\cdot$ {bystander} is part of a globally consistent assignment? How do we interpret this?
  - There can't be any globally consistent assignments
  - There are 2 police officers, 3 protesters, and … 1.9 (???) bystanders
  - I saw somebody. It's a protester with probability 3/6.9, a police officer with probability 2/6.9, and a bystander with probability 1.9/6.9.
  - …

Example – categorical data (cont.)

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

- Replace stalk {police officer, protester, bystander} with new stalk $\mathbb{R}$[{ {police officer}, {protester}, {bystander} }]
- Now stalk is a vector space!

- Problem: what if our analysis indicates that
  2 ·{police officer} +3 ·{protester} +1.9 ·{bystander}
  is part of a globally consistent assignment? How do we interpret this?
  - There can't be any globally consistent assignments
  - There are 2 police officers, 3 protesters, and … 1.9 (???) bystanders
  - I saw somebody. It's a protester with probability 3/6.9, a police officer with probability 2/6.9, and a bystander with probability 1.9/6.9.
  - …
- Interpretation must be situation dependent

Example – partial order data

c = [{cat, dog, mammal}, {cat ≤ mammal, dog ≤ mammal}]

- 1. Create a category, C, where stalk c is an object.
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
- 3. Create a vector space V with basis elements from $Decat(D)$

Example – partial order data

c = [{cat, dog, mammal}, {cat ≤ mammal, dog ≤ mammal}]

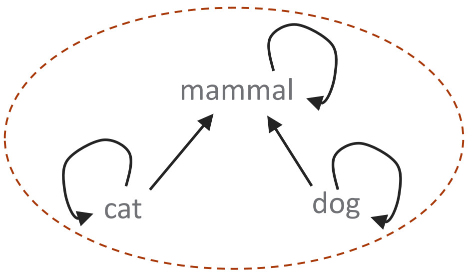
- 1. Create a category, C, where stalk c is an object.
  - $Ob(C) = \{(P, \mathcal{L}): P \in Ob(SET), \mathcal{L} \subseteq P \times P$ is reflexive, transitive, antisymmetric$\}$
  - $Hom_C\big((P, \mathcal{L}), (R, \mathcal{M})\big) = \{$order preserving functions $P \rightarrow R\}$
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
- 3. Create a vector space V with basis elements from $Decat(D)$

Example – partial order data

c = [{cat, dog, mammal}, {cat ≤ mammal, dog ≤ mammal}]

- 1. Create a category, C, where stalk c is an object.
  - $Ob(C) = \{(P, \mathcal{L}): P \in Ob(SET), \mathcal{L} \subseteq P \times P$ is reflexive, transitive, antisymmetric$\}$
  - $Hom_C\big((P, \mathcal{L}), (R, \mathcal{M})\big) = \{$order preserving functions $P \to R\}$
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - $Ob(D)$ = {cat, dog, mammal}
  - Hom_{SET} = {cat → mammal, dog → mammal} ∪ {identities}
  - Then $Decat(D)$ = { {cat}, {dog}, {mammal} } and there is clearly a surjection (bijection again) from c to $Decat(D)$

$D =$

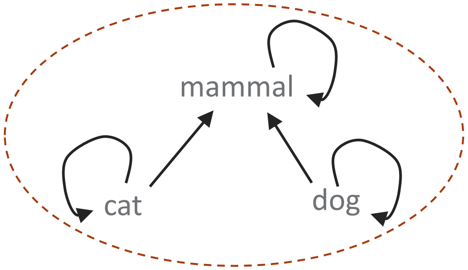


- 3. Create a vector space V with basis elements from $Decat(D)$

Example – partial order data

c = [{cat, dog, mammal}, {cat ≤ mammal, dog ≤ mammal}]

- 1. Create a category, C, where stalk c is an object.
  - $Ob(C) = \{(P, \mathcal{L}) : P \in Ob(SET), \mathcal{L} \subseteq P \times P$ is reflexive, transitive, antisymmetric}
  - $Hom_C\big((P, \mathcal{L}), (R, \mathcal{M})\big) =$ {order preserving functions $P \to R$}
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - $Ob(D)$ = {cat, dog, mammal}
  - Hom_{SET} = {cat → mammal, dog → mammal} ∪ {identities}
  - Then $Decat(D)$ = { {cat}, {dog}, {mammal} } and there is clearly a surjection (bijection again) from c to $Decat(D)$

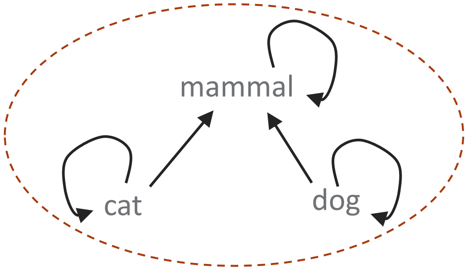
$D =$ 

- 3. Create a vector space V with basis elements from $Decat(D)$
  - V = ℝ[{ {cat}, {dog}, {mammal} }] = { α ·{cat} +β ·{dog} +γ ·{mammal} : $\alpha, \beta, \gamma \in \mathbb{R}$ }

Example – partial order data

c = [{cat, dog, mammal}, {cat ≤ mammal, dog ≤ mammal}]

- 1. Create a category, C, where stalk c is an object.
  - $Ob(C) = \{(P, \mathcal{L}) : P \in Ob(SET), \mathcal{L} \subseteq P \times P$ is reflexive, transitive, antisymmetric}
  - $Hom_C\big((P,\mathcal{L}),(R,\mathcal{M})\big) = \{$order preserving functions $P \to R\}$
- 2. Find a category D such that there is a surjection from c to the isomorphism classes of D, $Decat(D)$
  - $Ob(D)$ = {cat, dog, mammal}
  - Hom_{SET} = {cat → mammal, dog → mammal} ∪ {identities}
  - Then $Decat(D)$ = { {cat}, {dog}, {mammal} } and there is clearly a surjection (bijection again) from c to $Decat(D)$

$D =$ 

- 3. Create a vector space V with basis elements from $Decat(D)$
  - V = ℝ[{ {cat}, {dog}, {mammal} }] = { α ·{cat} +β ·{dog} +γ ·{mammal} : $\alpha, \beta, \gamma \in \mathbb{R}$ }
- Bonus: There is structure on c (and all partial orders) which is seen through the morphisms in D encoding the partial order relation.
  - Basis elements mapped to their principle filter

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Example in progress – real interval valued data

$c = \{[a, b] : a, b \in \mathbb{R}, a \leq b\}$ with structure of a semiring (addition, multiplication, identities are defined, but not inverses) and a partial order relation $[a, b] \leq [c, d]$ if and only if $a \leq c, b \leq d$

- 1. Create a category, C, where stalk c is an object.

Example in progress – real interval valued data

$c = \{[a, b] : a, b \in \mathbb{R}, a \leq b\}$ with structure of a semiring (addition, multiplication, identities are defined, but not inverses) and a partial order relation $[a, b] \leq [c, d]$ if and only if $a \leq c, b \leq d$

- 1. Create a category, C, where stalk c is an object.
  - $Ob(C) = \{$ partially ordered semirings $\}$
  - $Hom_C(R_1, R_2) = \{$ order preserving semiring homomorphisms $\}$

Example in progress – real interval valued data

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

$c = \{[a,b] : a,b \in \mathbb{R}, a \le b\}$ with structure of a semiring (addition, multiplication, identities are defined, but not inverses) and a partial order relation $[a,b] \le [c,d]$ if and only if $a \le c, b \le d$

- 1. Create a category, C, where stalk c is an object.
  - $Ob(C) =$ { partially ordered semirings }
  - $Hom_C(R_1, R_2) =$ { order preserving semiring homomorphisms }
- In this case, when *c* already has a lot of structure, going through an intermediate category *D* is probably more trouble than is needed
  - A ring can be encoded as a preadditive category – a category in which all Hom sets have the structure of an abelian group – with one object
  - …

Example in progress – real interval valued data

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

$c = \{[a, b] : a, b \in \mathbb{R}, a \leq b\}$ with structure of a semiring (addition, multiplication, identities are defined, but not inverses) and a partial order relation $[a, b] \leq [c, d]$ if and only if $a \leq c, b \leq d$

- 1. Create a category, C, where stalk c is an object.
  - $Ob(C) =$ { partially ordered semirings }
  - $Hom_C(R_1, R_2) =$ { order preserving semiring homomorphisms }
- In this case, when *c* already has a lot of structure, going through an intermediate category *D* is probably more trouble than is needed
  - A ring can be encoded as a preadditive category – a category in which all Hom sets have the structure of an abelian group – with one object
  - …
- Instead, starting with fields – closest to vector spaces – and working down to semirings
  - For field *F* “categorify” as vector space F^1
  - Work in progress to “categorify” rings and then semirings

Summary & Future Work

- Real-world data integration tasks need a way to put different data types on the same footing
- Finite dimensional vector spaces give that common footing and allow topological invariants to be calculated
- We developed a three step process and showed examples for three different data types
- Future work:
  - Implementing this into existing sheaf code packages
  - Showing examples using real data
  - Exploring how distances in the original sets get transformed