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

CLIFF JOSLYN, EMILIE PURVINE, MICHAEL ROBINSON

Pacific Northwest National Laboratory  American University
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 $\rightarrow$ ORDINAL we lose addition

- Transformation to FVECT allows us to turn any data type into a finite dimensional vector space, keeping structure

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 \( D \), \( \text{Decat}(D) \)
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 $D$, $\text{Decat}(D)$

3. Create a vector space $V$ with basis elements from $\text{Decat}(D)$

```
c  \xrightarrow{\text{surjection}}  \text{Decat}(D)  \xrightarrow{\text{surjection}}  V
```

(Probably should be “vector space-ification”)

"categorification"
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 $D$, $\text{Decat}(D)$
3. Create a vector space $V$ with basis elements from $\text{Decat}(D)$

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 $\text{Decat}(D)$
Example – categorical data

\[ c = \{ \text{police officer, protester, bystander} \} \]

1. Create a category, \( C \), where stalk \( c \) is an object.
   - \( C = \text{SET} \) as defined previously, clearly \( c \) is an object in \( \text{SET} \)

2. Find a category \( D \) such that there is a surjection from \( c \) to the isomorphism classes of \( D \), \( \text{Decat}(D) \)

3. Create a vector space \( V \) with basis elements from \( \text{Decat}(D) \)
Example – categorical data

\[ C = \{ \text{police officer, protester, bystander} \} \]

1. Create a category, \( C \), where stalk \( c \) is an object.
   - \( C = \text{SET as defined previously, clearly } c \) is an object in SET

2. Find a category \( D \) such that there is a surjection from \( c \) to the isomorphism classes of \( D, Decat(D) \)
   - \( Ob(D) = \{ \text{police officer, protester, bystander} \} \)
   - \( Hom_{SET}(x,y) = \begin{cases} \{id_x\} & x = y \\ \emptyset & \text{else} \end{cases} \)

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

\[ D = \begin{array}{c}
\text{police officer} \\
\text{protester} \\
\text{bystander}
\end{array} \]
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**) = \[
   \begin{cases} 
   \{id_x\} & \text{if } x = y \\
   \emptyset & \text{else} 
   \end{cases}
   \]
   - Then Decat(**D**) = { {police officer}, {protester}, {bystander} } and there is clearly a surjection (bijection in this case) from **c** to Decat(**D**)  

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.
   - \( C = \text{SET as defined previously, clearly } c \) is an object in \( \text{SET} \)

2. Find a category \( D \) such that there is a surjection from \( c \) to the isomorphism classes of \( D \), \( \text{Decat}(D) \)
   - \( \text{Ob}(D) = \{ \text{police officer, protester, bystander} \} \)
   - \( \text{Hom}_{\text{SET}}(x, y) = \begin{cases} \{ \text{id}_x \} & x = y \\ \emptyset & \text{else} \end{cases} \)
   - Then \( \text{Decat}(D) = \{ \{ \text{police officer} \}, \{ \text{protester} \}, \{ \text{bystander} \} \} \) and there is clearly a surjection (bijection in this case) from \( c \) to \( \text{Decat}(D) \)

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


**Example – categorical data**

\[ c = \{ \text{police officer, protester, bystander} \} \]

1. **Create a category, \( C \), where stalk \( c \) is an object.**
   - \( C = \text{SET} \) as defined previously, clearly \( c \) is an object in \( \text{SET} \)

2. **Find a category \( D \) such that there is a surjection from \( c \) to the isomorphism classes of \( D \), \( \text{Decat}(D) \)**
   - \( \text{Ob}(D) = \{ \text{police officer, protester, bystander} \} \)
   - \( \text{Hom}_{\text{SET}}(x, y) = \begin{cases} \text{id}_x & x = y \\ \emptyset & \text{else} \end{cases} \)
   - Then \( \text{Decat}(D) = \{ \{ \text{police officer} \}, \{ \text{protester} \}, \{ \text{bystander} \} \} \) and there is clearly a surjection (bijection in this case) from \( c \) to \( \text{Decat}(D) \)

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

**Bonus:** There is no structure on \( c \) (or any categorical) which is seen by the absence of morphisms other than identities in \( D \). Map \( V \to V \) is just the identity matrix in this case
Example –
categorical data (cont.)

► Replace stalk \{\text{police officer, protester, bystander}\} with new stalk
\(\mathbb{R}^{|\{\{\text{police officer}\}, \{\text{protester}\}, \{\text{bystander}\}\}|}\)

► Now stalk is a vector space!
Example – categorical data (cont.)

- Replace stalk \{\text{police officer, protester, bystander}\} with new stalk 
  \( \mathbb{R}[\{\text{police officer}\}, \{\text{protester}\}, \{\text{bystander}\}] \)

- Now stalk is a vector space!

- Problem: what if our analysis indicates that 
  \(2 \cdot \{\text{police officer}\} + 3 \cdot \{\text{protester}\} + 1.9 \cdot \{\text{bystander}\}\) 
  is part of a globally consistent assignment? How do we interpret this?
Example — categorical data (cont.)

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

- Problem: what if our analysis indicates that 
  \[2 \cdot \{\text{police officer}\} + 3 \cdot \{\text{protester}\} + 1.9 \cdot \{\text{bystander}\}\]
  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 \{\text{police officer}, \text{protester}, \text{bystander}\} with new stalk \mathbb{R}[\{\text{police officer}\}, \{\text{protester}\}, \{\text{bystander}\}]

- Now stalk is a vector space!

- Problem: what if our analysis indicates that
  \[2 \cdot \{\text{police officer}\} + 3 \cdot \{\text{protester}\} + 1.9 \cdot \{\text{bystander}\}\]
  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 \{\text{police officer}, \text{protester}, \text{bystander}\} with new stalk \mathbb{R}[\{\text{police officer}\}, \{\text{protester}\}, \{\text{bystander}\}]]

► Now stalk is a vector space!

► Problem: what if our analysis indicates that
\[2 \cdot \{\text{police officer}\} + 3 \cdot \{\text{protester}\} + 1.9 \cdot \{\text{bystander}\}\]
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.)

Replace stalk \{\text{police officer, protester, bystander}\} with new stalk 
\[ \mathbb{R}^3 = \{ \{\text{police officer}\}, \{\text{protester}\}, \{\text{bystander}\} \} \]

Now stalk is a vector space!

Problem: what if our analysis indicates that 
\[ 2 \cdot \{\text{police officer}\} + 3 \cdot \{\text{protester}\} + 1.9 \cdot \{\text{bystander}\} \]
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

\[ c = \{ \{\text{cat, dog, mammal}\}, \{\text{cat} \leq \text{mammal}, \text{dog} \leq \text{mammal}\} \} \]

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 \( D, \text{Decat}(D) \)

3. Create a vector space \( V \) with basis elements from \( \text{Decat}(D) \)
Example – partial order data

\[ c = \{\{\text{cat, dog, mammal}\}, \{\text{cat} \leq \text{mammal}, \text{dog} \leq \text{mammal}\}\} \]

1. Create a category, \( C \), where stalk \( c \) is an object.
   - \( Ob(C) = \{(P, \mathcal{L}) : P \in Ob(SET), \mathcal{L} \subseteq P \times P \text{ is reflexive, transitive, antisymmetric}\} \)
   - \( Hom_C((P, \mathcal{L}), (R, \mathcal{M})) = \{\text{order preserving functions } P \to R\} \)

2. Find a category \( D \) such that there is a surjection from \( c \) to the isomorphism classes of \( D, Decat(D) \)

3. Create a vector space \( V \) with basis elements from \( Decat(D) \)
Example – partial order data

\[ c = \{\{\text{cat, dog, mammal}\}, \{\text{cat} \leq \text{mammal, dog} \leq \text{mammal}\}\} \]

1. Create a category, \( C \), where stalk \( c \) is an object.
   - \( \text{Ob}(C) = \{(P, L) : P \in \text{Ob}(\text{SET}), L \subseteq P \times P \text{ is reflexive, transitive, antisymmetric}\} \)
   - \( \text{Hom}_C((P, L), (R, M)) = \{\text{order preserving functions } P \to R\} \)

2. Find a category \( D \) such that there is a surjection from \( c \) to the isomorphism classes of \( D, \text{Decat}(D) \)
   - \( \text{Ob}(D) = \{\text{cat, dog, mammal}\} \)
   - \( \text{Hom}_{\text{SET}} = \{\text{cat} \to \text{mammal, dog} \to \text{mammal}\} \cup \{\text{identities}\} \)
   - Then \( \text{Decat}(D) = \{\{\text{cat}\}, \{\text{dog}\}, \{\text{mammal}\} \} \) and there is clearly a surjection (bijection again) from \( c \) to \( \text{Decat}(D) \)

3. Create a vector space \( V \) with basis elements from \( \text{Decat}(D) \)
Example – partial order data

\[ c = \{ \{ \text{cat, dog, mammal} \}, \{ \text{cat} \leq \text{mammal, dog} \leq \text{mammal} \} \} \]

1. Create a category, \( C \), where stalk \( c \) is an object.
   - \( \text{Ob}(C) = \{(P, L) : P \in \text{Ob}\left(\text{SET}\right), L \subseteq P \times P \text{ is reflexive, transitive, antisymmetric} \} \)
   - \( \text{Hom}_C((P, L), (R, M)) = \{ \text{order preserving functions } P \to R \} \)

2. Find a category \( D \) such that there is a surjection from \( c \) to the isomorphism classes of \( D \), \( \text{Decat}(D) \)
   - \( \text{Ob}(D) = \{ \text{cat, dog, mammal} \} \)
   - \( \text{Hom}_{\text{SET}} = \{ \text{cat} \to \text{mammal, dog} \to \text{mammal} \} \cup \{ \text{identities} \} \)
   - Then \( \text{Decat}(D) = \{ \{ \text{cat} \}, \{ \text{dog} \}, \{ \text{mammal} \} \} \) and there is clearly a surjection (bijection again) from \( c \) to \( \text{Decat}(D) \)

3. Create a vector space \( V \) with basis elements from \( \text{Decat}(D) \)
   - \( V = \mathbb{R}[\{ \{ \text{cat} \}, \{ \text{dog} \}, \{ \text{mammal} \} \}] = \{ \alpha \cdot \{ \text{cat} \} + \beta \cdot \{ \text{dog} \} + \gamma \cdot \{ \text{mammal} \} : \alpha, \beta, \gamma \in \mathbb{R} \} \)
1. Create a category, \( C \), where stalk \( c \) is an object.
   - \( \text{Ob}(C) = \{(P, \mathcal{L}) : P \in \text{Ob}(SET), \mathcal{L} \subseteq P \times P \text{ is reflexive, transitive, antisymmetric}\} \)
   - \( \text{Hom}_C((P, \mathcal{L}), (R, \mathcal{M})) = \{\text{order preserving functions } P \rightarrow R\} \)

2. Find a category \( D \) such that there is a surjection from \( c \) to the isomorphism classes of \( D \), \( \text{Decat}(D) \)
   - \( \text{Ob}(D) = \{\text{cat, dog, mammal}\} \)
   - \( \text{Hom}_{\text{SET}} = \{\text{cat} \rightarrow \text{mammal}, \text{dog} \rightarrow \text{mammal}\} \cup \{\text{identities}\} \)
   - Then \( \text{Decat}(D) = \{\{\text{cat}\}, \{\text{dog}\}, \{\text{mammal}\}\} \) and there is clearly a surjection (bijection again) from \( c \) to \( \text{Decat}(D) \)

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

**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.
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.

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

- \( \text{Ob}(C) = \{ \text{partially ordered semirings} \} \)
- \( \text{Hom}_C(R_1, R_2) = \{ \text{order preserving semiring homomorphisms} \} \)

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

\[ c = \{ [a, b] : a, b \in \mathbb{R}, a \leq b \} \] 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.
   - \(Ob(C) = \{ \text{partially ordered semirings} \}\)
   - \(Hom_C(R_1, R_2) = \{ \text{order preserving semiring homomorphisms} \}\)

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
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