大規模言語モデルの数学的構造 II

enriched categoryによる言語モデル

Agenda

大規模言語モデルの数学的構造 II

enriched categoryによる言語モデル

Part 1　第一部のふりかえり

Part 2　言語の論理性と意味のモデルとしてのcopresheaf

Part 3　enriched category論入門

Part 4　enriched category論の言語理論への応用


Part 1
第一部のふりかえり
大規模言語モデルの数学的構造 II

言語のカテゴリー L を、preorder category として捉える

言語を構成する意味を持つ文字列である、語･フレーズ･文･文の連続 ... を「表現」とします。任意の表現 S, T について、表現Sの文字列が表現Tの部分文字列であるとき、$S \leq T$と順序を定義します。

この順序は、次の性質が確かめられますので preorder（前順序）です。

1. $S \leq S$（反射律）
2. $S \leq T$ かつ $T \leq U$ なら、$S \leq U$（推移律）

この時、言語の表現をオブジェクトとし、表現間の順序 $S \leq T$ を $S \to T$ という射(モルフィズム)を定義すると考えると、言語は、preorder category として捉えることができることがわかります。

射の合成の結合性についてです。
次のPとLの対応を見ればわかると思います。

P

$$
\begin{tikzcd}
S \arrow[r, phantom, "\leq"] \arrow[rr, bend left, phantom, "\leq"] & T \arrow[r, phantom, "\leq"] \arrow[rr, bend right, phantom, "\leq"'] & U \arrow[r, phantom, "\leq"] & V
\end{tikzcd}
$$

$$S \leq U \leq V \Longrightarrow S \leq V$$

$$S \leq T \leq V \Longrightarrow S \leq V$$

L

$$
\begin{tikzcd}
S \arrow[r, "f"'] \arrow[rr, bend left, "f \circ g"] & T \arrow[r, "g"'] \arrow[rr, bend right, "g \circ h"'] & U \arrow[r, "h"'] & V
\end{tikzcd}
$$

$$(f \circ g) \circ h = f \circ g \circ h$$

$$f \circ (g \circ h) = f \circ g \circ h$$

ここでは、次のことに注意しましょう。

- この言語のカテゴリー L では、二つの異なるオブジェクト間の射は、ただ一つ存在するか、あるいは、全く存在しないかのいずれかです。
- この言語のカテゴリー L は、ある言語が文字表現を持つ限り、日本語、英語、... といった違いを超えた言語の特徴を表現しています。

表現Sの「意味」を、カテゴリーL内の $S \rightarrow T$ なる射全体の集合と考える

Firth は、次のような言葉を残しています。

「我々は、ある語を、それが引きつれている仲間たちによって知ることになる。」

こうした意味の理解をベースにして、ある表現Sの意味を、それが「引きつれている仲間たち」の全体の集合で考えることにします。Sの「仲間たち」とは、$S \rightarrow T$ なる射が存在する表現Tのことです。

それらの全体を考えるということは、$S \rightarrow T$ なる射全体の集合を考えるということと同じことです。

意味のカテゴリーは、copresheaf Set^L で表現される

カテゴリー L 内の $S \rightarrow T$ なる射を、$L(S,T)$ と表します。また、$L(S,-)$ で、Sの意味を表現している、Sのすべての「仲間たち」への「射全体の集合」を表すことにしましょう。

SがLのオブジェクトである時、それに対応する意味のカテゴリーを構成しているオブジェクトは、射全体の集合$L(S,-)$だと考えることができます。この対応は、カテゴリーL から集合のカテゴリーSetへのfunctorを考えることに他なりません。

カテゴリー論では、カテゴリーCから集合のカテゴリーSetに値を持つfunctorで構成されるカテゴリーを、Set^Cで表し、C上のcopresheaf と呼びます。

意味のカテゴリーは、言語のカテゴリー L から、集合のカテゴリーSetに値を持つ、functor で構成される L上のcopresheaf $\mathrm{Set}^{\mathrm{L}}$ で表現されることになります。

集合に値を持つ表現可能なfunctorである意味のcategory Set^Lを、L上の**copresheaf**といいます。

意味の$category$ Set^L

copresheaf

$$Set^L$$

$L(x,-)$

$L(z,-)$ $L(a,-)$

$L(y,-)$

すべてのLのオブジェクトxについて
L(x, −)なるfunctorの集まり

言語のカテゴリーと意味のカテゴリーの対応は、Yoneda Embeddingで与えられる

言語のカテゴリー Lと意味のカテゴリー Set^Lの対応は、Lのオブジェクトに、Set^LのオブジェクトL(S, −)を対応付けるものです。

こうした対応づけをYoneda Embeddingと呼びます。

category L とcategory Set^Lの対応は、
red と$L(red, -)$ を対応づけるものです。

$$\begin{array}{ccc} L & \Longrightarrow & Set^L \\ \Cup & & \Cup \\ \mathrm{red} & \longmapsto & \mathrm{L(red,\ -)} \end{array}$$

より一般的な、次の値の割り当てを、Yoneda embedding
と言います。

$$\begin{array}{ccc} L & \Longrightarrow & Set^L \\ \Cup & & \Cup \\ \mathrm{x} & \longmapsto & \mathrm{L(x,\ -)} \end{array}$$

Part 2

言語の論理性と
意味のモデルとしてのcopresheaf

大規模言語モデルの数学的構造 II

Part 2

言語の論理性と
意味のモデルとしてのcopresheaf

- 意味のカテゴリー copresheaf について
- “red or blue” の意味を考える
- “red and blue” の意味を考える

意味のカテゴリー　copresheaf について

大規模言語モデルの数学的構造 II

$L(x,y)$と$L(x,-)$

Lの射 $x \to y$ を$L(x,y)$ と表す。

$L(x,y)$ なる全ての射の集合を$L(x,-)$ で表す。

Firth流の解釈をすれば、この集合$L(x,-)$ が「xの仲間たち」の全体となり、表現xの意味を表すことになる。

$L(x,y)$の性質

言語のカテゴリーLの定義に戻って考える。

Lで、$L(x,y)$ すなわち $x \rightarrow y$なる射が存在するのは、xがyの部分文字列になる時だけである。

よって、Lの任意のオブジェクトx，y について射$L(x,y)$は、ただ一つ存在するか、まったく存在しないかのいずれかである。

L(x, y)の性質

ただ一つの要素からなる集合を$*$で表し、空集合を$\emptyset$で表すと、

$$L(x, y) = \begin{cases} * & x \to y \text{なる射が存在する場合} \\ \emptyset & \text{それ以外の場合} \end{cases}$$

と表すことができる。

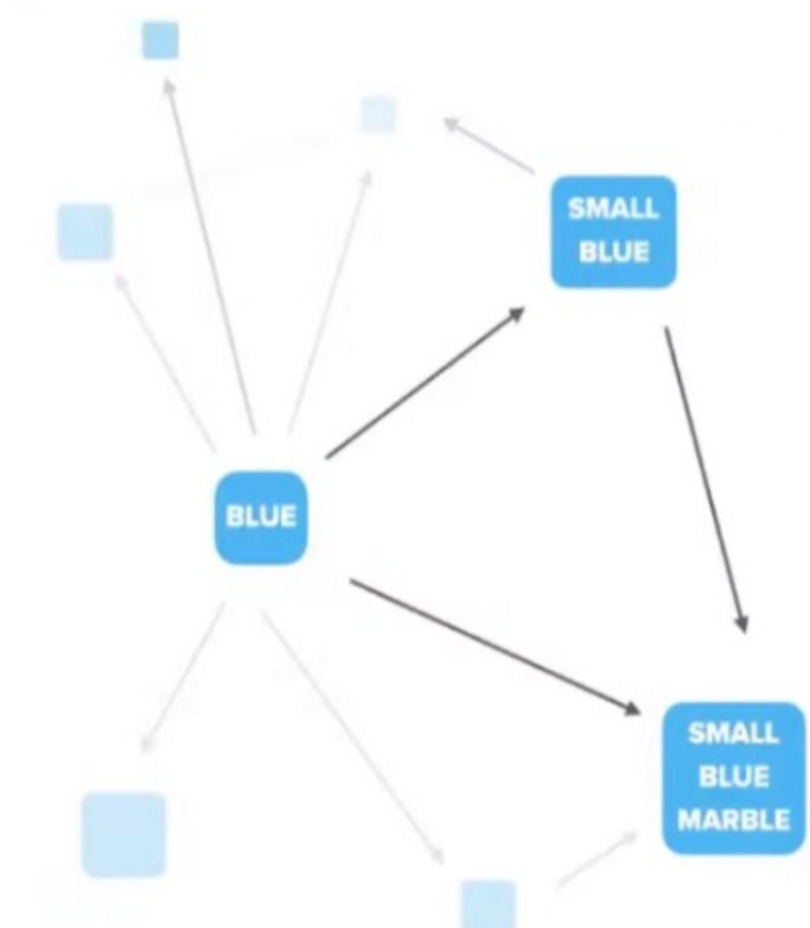
例えば、

$$L(red, red\ firetruck) \cong *$$
$$L(red, beautiful\ red\ ruby) \cong *$$
$$L(red, hot\ dog) = \emptyset$$
$$L(red, this\ is\ a\ pen) = \emptyset$$

$L(blue, -)$ の例

L の表現

ここで、* とマークされた
表現の全体の集合が、
blueの意味を与える。



意味のカテゴリー について

- Lのオブジェクトである言語の表現x を、集合$L(x,-)$に割り当てるfunctorを考える。
- xに応じた異なる集合$L(x,-)$の割り当ては、異なるfunctorによって生み出される。
- 意味のカテゴリーは、言語のカテゴリー Lから集合のカテゴリーSetへのfunctorからなるカテゴリー $L \to Set$ である。
- このfunctorカテゴリーを Set^L（あるいは$[\mathrm{L}, \mathrm{Set}]$）と表し、L上のcopresheafと呼ぶ。

“red or blue” の意味を考える

大規模言語モデルの数学的構造 II

集合の和”A∪B”と論理の”A or B”

一対の集合A, Bがあるとする。例えば、

集合A:　赤い色を持つすべてのものの集合
（赤い口紅、フランス国旗、一時停止標識など）。

集合B:　青い色を持つすべてのものの集合
（コマドリの卵、フランス国旗、ブルーベリーなど）。

赤か青のどちらかを持つすべてのものの集合とは何か？
それは別の集合、すなわちそれらの和 A∪B である。

上の例で言うと、集合 A∪B は、”red or blue”という概念に対応する集合である。

集合の和に対応するfunctor coproduct

集合の和を取ることができるように、functorに対しても同様の操作を行うことができる。

それを「集合の和」と呼ぶ代わりに、”coproduct” という別の名前をつけ、’⊔’ という記号で表す。

”red"と ”blue"を表す2つのfunctorのcoproductは、それらの論理和、つまり ”red or blue"という概念を表すもう1つのfunctorである。

coproductの定義

coproductは、カテゴリー論的には、次のように定義される。
説明は、別の機会に行う。

Let C be a category and let X_1 and X_2 be objects of C. An object is called the coproduct of X_1 and X_2, written $X_1 \sqcup X_2$, or $X_1 \oplus X_2$, or sometimes simply $X_1 + X_2$, if there exist morphisms $i_1 : X_1 \to X_1 \sqcup X_2$ and $i_2 : X_2 \to X_1 \sqcup X_2$ satisfying the following universal property: for any object Y and any morphisms $f_1 : X_1 \to Y$ and $f_2 : X_2 \to Y$, there exists a unique morphism $f : X_1 \sqcup X_2 \to Y$ such that $f_1 = f \circ i_1$ and $f_2 = f \circ i_2$. That is, the following diagram commutes:

$$\begin{tikzcd}
 & Y & \\
X_1 \arrow[ur, "f_1"] \arrow[r, "i_1"'] & X_1 \coprod X_2 \arrow[u, dashed, "f"] & X_2 \arrow[ul, "f_2"'] \arrow[l, "i_2"]
\end{tikzcd}$$

The unique arrow f making this diagram commute may be denoted $f_1 \sqcup f_2$, $f_1 \oplus f_2$, $f_1 + f_2$, or $[f_1, f_2]$. The morphisms i_1 and i_2 are called *canonical injections*, although they need not be injections or even monic.

基本的なアイデア

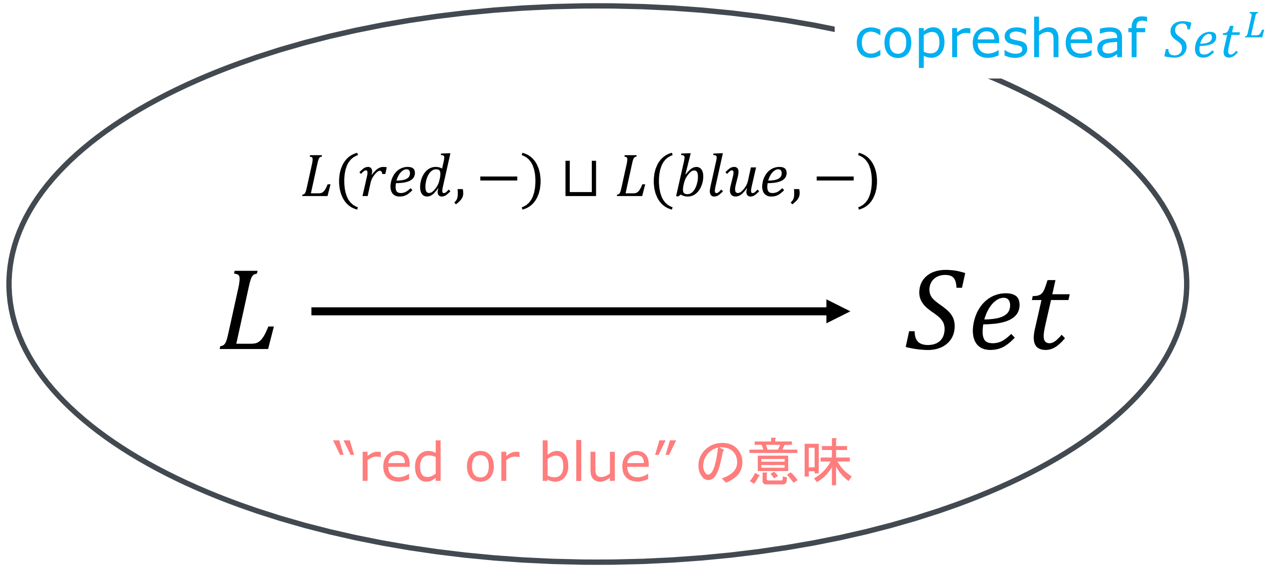
functor $L(red, -)$は、意味 “red”を表し、
functor $L(blue, -)$は、意味 “blue”を表す。

それでは、”red or blue” の意味を表すfunctor は何であろうか？

答えは、functor $L(red, -)$ と functor $L(blue, -)$ の coproduct $L(red, -) \sqcup L(blue, -)$ である。

意味のカテゴリー copresheaf で考える

“red or blue” の意味は、意味のカテゴリー copresheaf Set^L のオブジェクトである functor $L(red, -) \sqcup L(blue, -)$ で表される。



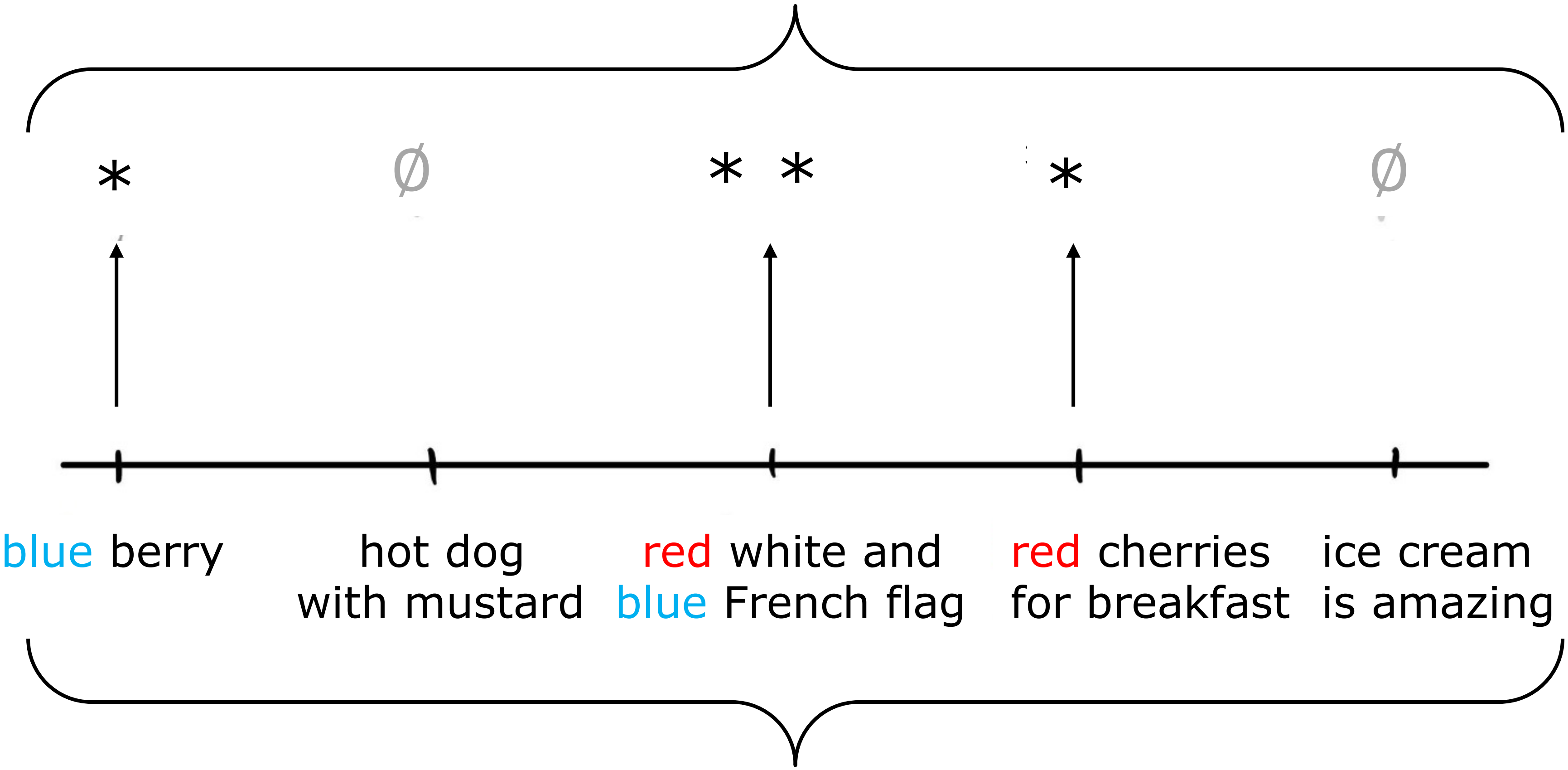
言語のカテゴリー Lと意味のカテゴリーSet^Lを結ぶ
Yoneda embedding

$$
\begin{array}{ccc}
L & \longrightarrow & Set^L \\
\rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\
y & \longmapsto & L(red, y) \sqcup L(blue, y)
\end{array}
$$

この集合は、yがredを含む時 *
そうではない時は、空集合 ∅ である。

この集合は、yがblueを含む時 *
そうではない時は、空集合 ∅ である。

Set^Lでの $L(red, y) \sqcup L(blue, y)$ の値



Lでの表現 y

"red or blue" の意味を表す

$$L(red, -) \sqcup L(blue, -)$$

$$\mathrm{L}(\mathbf{red}, -) \sqcup \mathrm{L}(\mathbf{blue}, -) \overset{\text{SORT OF}}{=} \begin{bmatrix} * \\ * \\ * \\ \varnothing \\ * \, * \\ \varnothing \\ \vdots \end{bmatrix} \begin{array}{l} \textbf{deep red Bing cherries} \\ \textbf{small blue marble} \\ \textbf{beautiful blue ocean} \\ \textbf{did you put the kettle on} \\ \textbf{red and blue fireworks} \\ \textbf{Sencha green tea} \\ \vdots \end{array}$$

“red and blue” の意味を考える

大規模言語モデルの数学的構造 II

集合の交わり”A∩B”と論理の”A and B”

前回のセッションと議論はパラレルに進む。

一対の集合A, Bがあるとする。例えば、

集合A: 赤い色を持つすべてのものの集合
（赤い口紅、フランス国旗、一時停止標識など）。

集合B: 青い色を持つすべてのものの集合
（コマドリの卵、フランス国旗、ブルーベリーなど）。

赤と青のどちらをも持つすべてのものの集合とは何か？
それは別の集合、すなわちそれらの交わり A∩B である。

上の例で言うと、集合 A∩B は、”red and blue”という概念に対応する集合である。

集合の交わりに対応するfunctor product

集合の交わりを取ることができるように、functorに対しても同様の操作を行うことができる。

それを「集合の交わり」と呼ぶ代わりに、”product” という別の名前をつけ、’×’ という記号で表す。

”red"と ”blue"を表す2つのfunctorのproductは、それらの論理積、つまり ”red and blue"という概念を表すもう1つのfunctorである。

productの定義

productは、カテゴリー論的には、次のように定義される。
説明は、別の機会に行う。

Fix a category C. Let X_1 and X_2 be objects of C. A product of X_1 and X_2 is an object X, typically denoted $X_1 \times X_2$, equipped with a pair of morphisms $\pi_1 : X \to X_1$, $\pi_2 : X \to X_2$ satisfying the following universal property:

- For every object Y and every pair of morphisms $f_1 : Y \to X_1$, $f_2 : Y \to X_2$, there exists a unique morphism $f : Y \to X_1 \times X_2$ such that the following diagram commutes:

$$\begin{tikzcd}
& Y \arrow[dl, "f_1"'] \arrow[d, dashed, "f"] \arrow[dr, "f_2"] & \\
X_1 & X_1 \times X_2 \arrow[l, "\pi_1"] \arrow[r, "\pi_2"'] & X_2
\end{tikzcd}$$

Whether a product exists may depend on C or on X_1 and X_2. If it does exist, it is unique up to canonical isomorphism, because of the universal property, so one may speak of *the* product. This has the following meaning: let X', π'_1, π'_2 be another cartesian product, there exists a unique isomorphism $h : X' \to X_1 \times X_2$ such that $\pi'_1 = \pi_1 \circ h$ and $\pi'_2 = \pi_2 \circ h$.

coproductの定義
前回見たもの

coproductは、カテゴリー論的には、次のように定義される。
説明は、別の機会に行う。

Let C be a category and let X_1 and X_2 be objects of C. An object is called the coproduct of X_1 and X_2, written $X_1 \sqcup X_2$, or $X_1 \oplus X_2$, or sometimes simply $X_1 + X_2$, if there exist morphisms $i_1 : X_1 \to X_1 \sqcup X_2$ and $i_2 : X_2 \to X_1 \sqcup X_2$ satisfying the following universal property: for any object Y and any morphisms $f_1 : X_1 \to Y$ and $f_2 : X_2 \to Y$, there exists a unique morphism $f : X_1 \sqcup X_2 \to Y$ such that $f_1 = f \circ i_1$ and $f_2 = f \circ i_2$. That is, the following diagram commutes:

$$\begin{tikzcd}
& Y & \\
X_1 \arrow[ur, "f_1"] \arrow[r, "i_1"'] & X_1 \coprod X_2 \arrow[u, dashed, "f"] & X_2 \arrow[ul, "f_2"'] \arrow[l, "i_2"]
\end{tikzcd}$$

The unique arrow f making this diagram commute may be denoted $f_1 \sqcup f_2$, $f_1 \oplus f_2$, $f_1 + f_2$, or $[f_1, f_2]$. The morphisms i_1 and i_2 are called *canonical injections*, although they need not be injections or even monic.

product

$$
\begin{tikzcd}
 & & Y \arrow[dll, "f_1"'] \arrow[d, dashed, "f"] \arrow[drr, "f_2"] & & \\
X_1 & & X_1 \times X_2 \arrow[ll, "\pi_1"] \arrow[rr, "\pi_2"'] & & X_2
\end{tikzcd}
$$

矢印の向きを
反対にしたもの

coproduct

$$
\begin{tikzcd}
 & & Y & & \\
X_1 \arrow[urr, "f_1"] \arrow[rr, "i_1"'] & & X_1 \coprod X_2 \arrow[u, dashed, "f"] & & X_2 \arrow[ull, "f_2"'] \arrow[ll, "i_2"]
\end{tikzcd}
$$

すべてのYと

$f_1 : Y \to X_1,$

$f_2 : Y \to X_2$

について

右の図が可換となる

ユニークな射f

$f : Y \to X_1 \times X_2$

が存在する。

Universal property

すべてのYと

$f_1 : X_1 \to Y,$

$f_2 : X_2 \to Y$

について

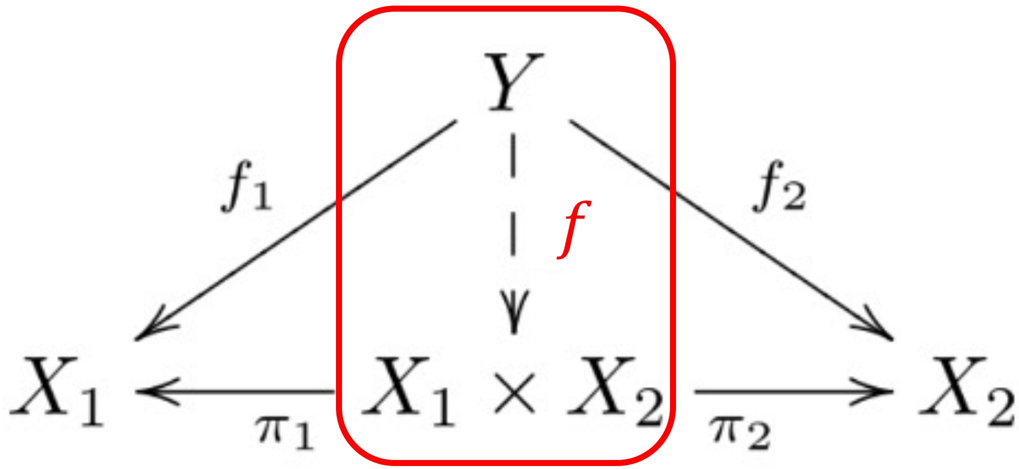
右の図が可換となる

ユニークな射f

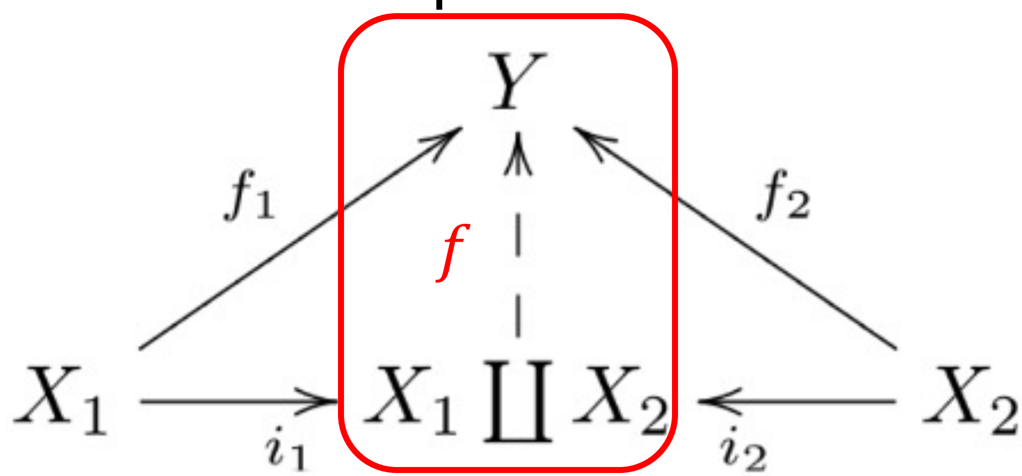
$f : X_1 \sqcup X_2 \to Y$

が存在する。

product



coproduct



基本的なアイデア

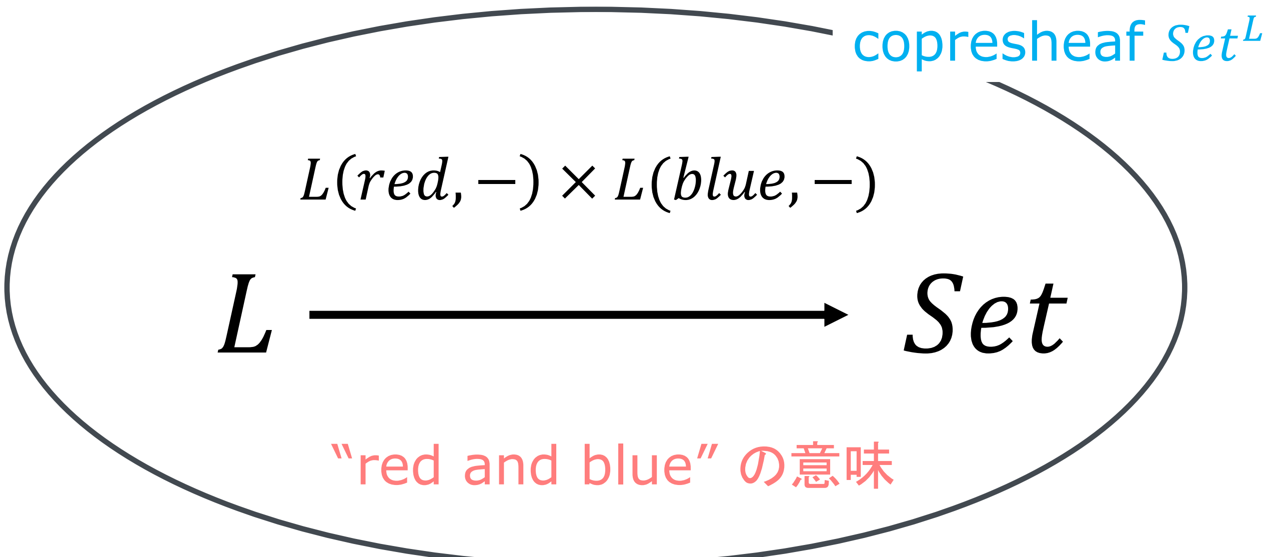
functor $L(red, -)$は、意味 “red”を表し、
functor $L(blue, -)$は、意味 “blue”を表す。

それでは、”red and blue” の意味を表すfunctor は何であろうか？

答えは、functor $L(red, -)$ と functor $L(blue, -)$ のproduct の functor $L(red, -) \times L(blue, -)$ である。

意味のカテゴリー copresheaf で考える

“red and blue” の意味は、意味のカテゴリー copresheaf Set^L のオブジェクトである functor $L(red,-) \times L(blue,-)$ で表される。



言語のカテゴリー Lと意味のカテゴリーSet^Lを結ぶ Yoneda embedding

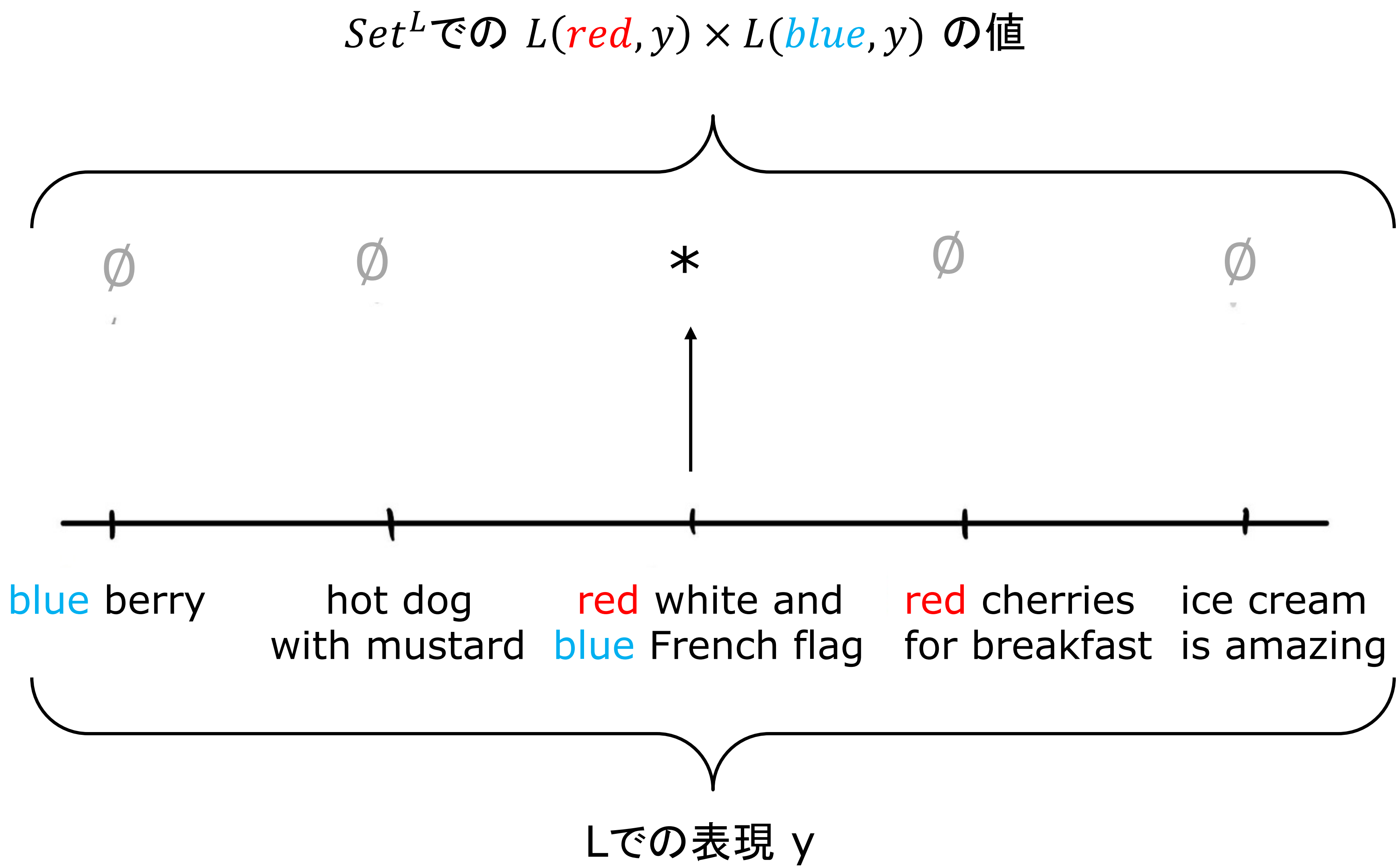
この集合は、
yがredを含み、かつ
yがblueを含む時 ＊
そうではない時は、
空集合 ∅ である。

$(*,*) \cong *$

$$L \longrightarrow Set^L$$

$$\rotatebox{90}{$\in$} \qquad\qquad \rotatebox{90}{$\in$}$$

$$y \longmapsto L(red, y) \times L(blue, y) = \begin{cases} * \\ \emptyset \end{cases}$$



Set^Lでの $L(red, y) \times L(blue, y)$ の値
∅
∅
*
∅
∅
blue berry
hot dog with mustard
red white and blue French flag
red cherries for breakfast
ice cream is amazing
Lでの表現 y

“red and blue” の意味を表す

$$L(red, -) \times L(blue, -)$$

$$\mathrm{L}(\mathbf{red}, -) \times \mathrm{L}(\mathbf{blue}, -) \overset{\text{SORT OF}}{=} \begin{bmatrix} \varnothing \\ \varnothing \\ \varnothing \\ \varnothing \\ * \\ \varnothing \\ \vdots \end{bmatrix} \begin{array}{l} \textbf{deep red Bing cherries} \\ \textbf{small blue marble} \\ \textbf{beautiful blue ocean} \\ \textbf{did you put the kettle on} \\ \textbf{red and blue fireworks} \\ \textbf{Sencha green tea} \\ \quad\quad\vdots \end{array}$$

Part 3

enriched category論入門

大規模言語モデルの数学的構造 II

Part 3

enriched category論入門

- 言語のカテゴリーに確率を導入する
- enriched category とは何か？
- enrich以前のカテゴリーとenrich化されたカテゴリーを比較する
- commutative monoidal preorder

言語のカテゴリーに確率を導入する

第一部のふりかえりと第二部の課題

大規模言語モデルの数学的構造 II

言語のcategoryに確率を導入する

これまでみてきた言語のcategory L では、二つの表現SとTがある時、SがTの部分文字列である時、S → T という射が存在します。

例えば、次のような射が category L には存在します。

red → red firetruck
red → red idea

S → T という射を、単なる部分文字列の関係としてではなく、表現Sが表現Tを「連想させる」という関係として考えると、普通の言語使用の局面では、red → red firetruck の方が、red → red idea よりたくさん出現するような気がします。

「連想」というのが曖昧だというなら、表現の「継続」あるいは表現の「連続」と考えて構いません。

こうした違いを、数値的に次のように表現することにします。

$$\text{red} \xrightarrow{0.12} \text{red firetruck}$$

$$\text{red} \xrightarrow{0.003} \text{red idea}$$

この例は仮のものですが、ここでのポイントは、射 red → red firetruck に割り当てられた 0.12という数字が、射 red → red ideaに割り当てられた 0.003という数字より大きいということです。

このことが、「普通の言語使用の局面では、red → red firetruck の方が、red → red idea よりたくさん出現するような気がする」ということを表現していると考えましょう。

もう少しきちんと定義すれば、これらの数字は、

表現Sが現れた時、表現Sの「継続」として表現Tが現れる
条件付き確率π(T|S)

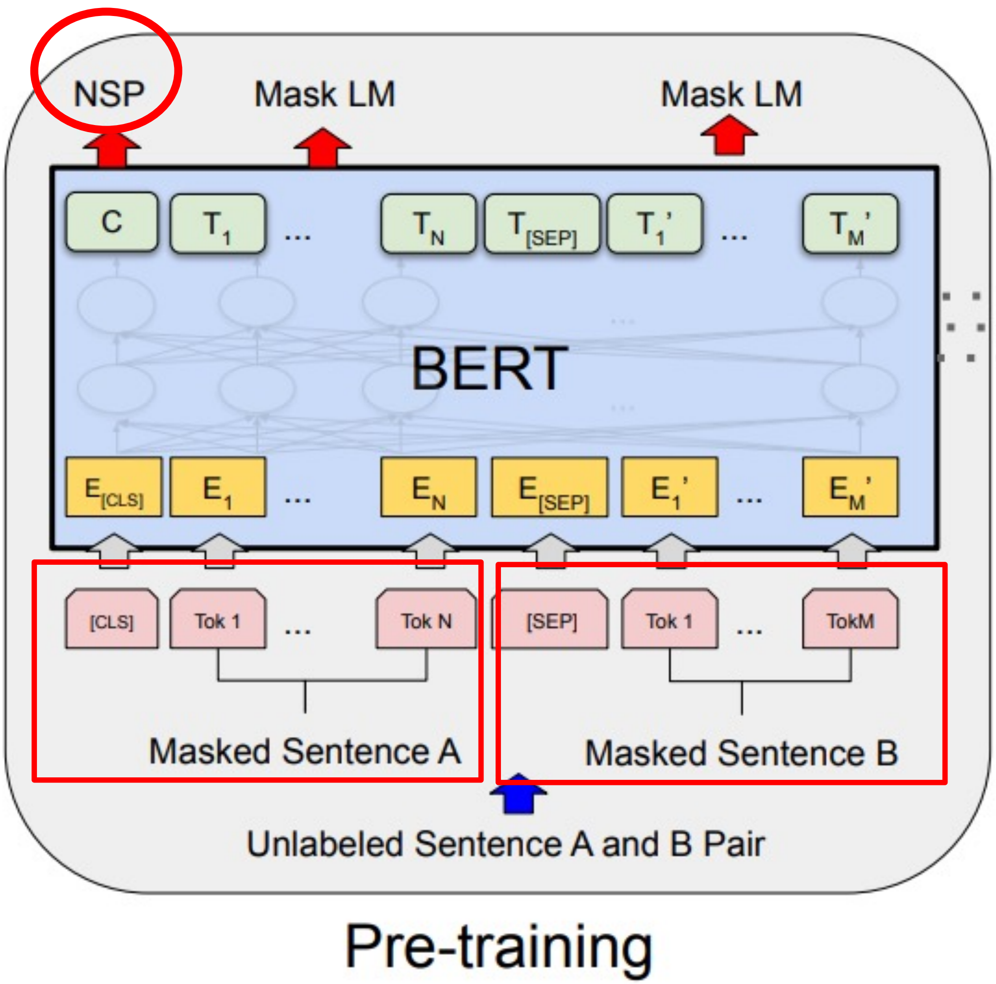
だと考えることができます。

大規模言語モデルでの Next Sentence Prediction 導入の意味

僕は、翻訳モデルから大規模言語モデルへの飛躍をもたらした最大のものは、文と文の「継続」あるいは「連続」を学習することを可能とする能力の獲得にあったと考えています。

もっともプリミティブなものは、おそらくBERTのpretrainingで導入されたNext Sentence Prediction のフラグだったと思います。

Pre-training BERTのタスク
Next Sentence Prediction (NSP)



- **NSP**: Next Sentence Prediction

  IsNext/NotNext

NSPの入力:Sentence Pair

Sentence BがSentence Aとつながることを判定する

BERTの例でいえば、category L で次のような射が、ある確率を割り当てられて存在することを意味しています。

the man went to convenience store → the man went to convenience store he bought a gallon of milk

ただ、次のような射は存在しません。あるいは、割り当てられる確率はゼロです。

the man went to convenience store → the man went to convenience store penguin are flight less birds

このように、category L に確率を導入することで、大規模言語モデルの表現の継続の振る舞いを、より具体的にシミレートできます。

確率導入のもう一つの意味
文法性の付与

それだけではありません。言語のcategoryへの確率の導入は、言語に文法性を与えることに繋がります。

英語だと、射 cat → black cat は、射 cat → cat black より出現確率は高いのですが、フランス語だと、逆に、射 chat → chat noir の方が射 chat → noir chatより出現確率が高くなります。

この確率の違いは、英語では形容詞の後ろに名詞が続くのに対して、フランス語では名詞の後ろに形容詞が続くという、二つの言語の文法の違いを反映しています。

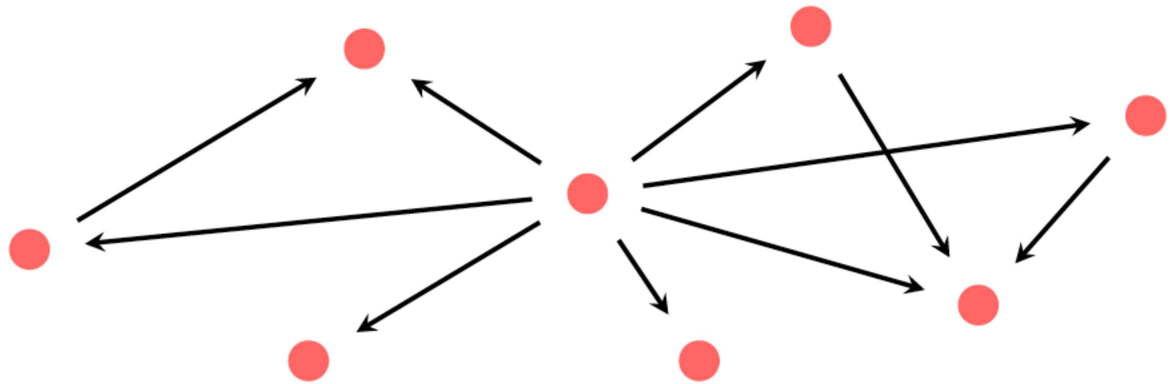
こうして、「形容詞は名詞の前にくる」（あるいは、形容詞は名詞の後ろにくる」）といった、文法ルールの形は取らないのですが、射に割り当てる確率を、その代理として、文法にあたるものを表現できます。

これまでは、言語からできるだけ構造を排除して、残った preorder の性質からだけ、言語のcategory L を作ってきたのですが、pregroupのような強い性質を前提にしなくとも、確率の導入だけで、言語のcategory Lに文法性を回復できるのです。

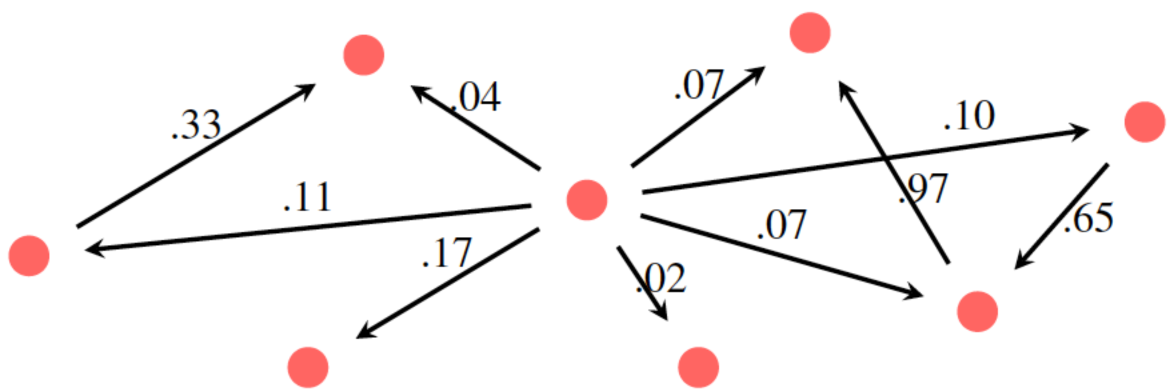
こうして拡張された言語のcategory を L′ としましょう。

確率を付与された言語のcategory L′のイメージ

これまでの
言語のcategory L



拡張された
言語のcategory L′



enriched category

このように、あるcategoryをベースにしながら、その上に新しい特徴を与えて作られるcategoryをenriched category と言います。

今回のセミナーが取り上げようとしているTai-Danae らの論文、“An Enriched Category Theory Of Language : From Syntax to Semantics” は、まさに、このenriched categoryを用いて、大規模言語モデルと言語の数学的構造に迫ろうとしたものです。

第一部で構成した代数的言語理論のふりかえり

第一部で構成した代数的（確率論なし）言語理論をふりかえってみましょう。

1. 表現の部分文字列の包含関係を射とするpreorder categoryとして、言語のカテゴリーLを特徴づける。
2. 言語のカテゴリーLから集合のカテゴリーSetへのfunctor categoryであるcopresheaf Set^L として、意味のカテゴリーを特長づける。copresheaf Set^Lのオブジェクトは、$L \to Set$ であるfunctorである。
3. 言語のカテゴリーLのオブジェクトを意味のカテゴリーcopresheaf Set^L のオブジェクトと結びつけるには、Yoneda embedding $x \mapsto L(x, -)$ を利用する。
4. Lの論理的構造 OR, ANDを表現するのに、functorのcoproduct, product を利用する。

重要な事は、第二部でも、第一部の次のような copresheaf 意味論の枠組みを基本的に踏襲すると言う事である

言語のカテゴリー
preorder

意味のカテゴリー
copresheaf

$$
\begin{tikzcd}
L \arrow[rr] & & Set^{L} \\
x \arrow[u, phantom, sloped, "\in"] \arrow[rr, mapsto, "\text{Yoneda embedding}"'] & & \mathrm{L}(x,-) \arrow[u, phantom, sloped, "\in"]
\end{tikzcd}
$$

問題は、Lの射L(x,y) が、次の様に確率値を持つように"enriche"化された時、この構成がどの様に変化するかということである。

$$\mathcal{L}(x,y) := \pi(y|x)$$

$$\mathcal{L}\,(red, red\ firetruck) = 0.02$$
$$\mathcal{L}\,(red, red\ idea) = 10^{-5}$$
$$\mathcal{L}\,(red, blue\ sky) = 0.$$

こうして拡張されたcategory Lを、[0, 1]上で enrich化されたカテゴリーという。

[0, 1]は、確率が値を取る 0から1までの実数(0,1の両端を含む)を表し、単位区間とよぶ。

第二部の課題

copresheaf意味論を継承するためには、次の様な問題に答えていかなければならない。

- Lの[0,1]上でのenrich化は、どの様に可能か？
- copresheafの[0,1]上でのenrich化は、どの様に可能か？
- Yoneda Embeddingの[0,1]上でのenrich化は、どの様に可能か？
- coproductの[0,1]上でのenrich化は、どの様に可能か？
- productの[0,1]上でのenrich化は、どの様に可能か？
- ……

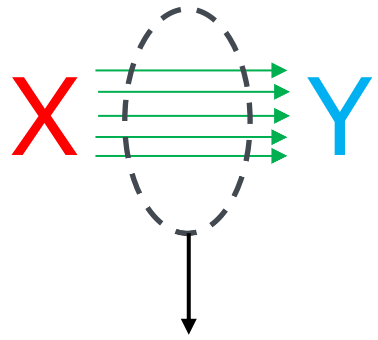
enriched category とは何か？

大規模言語モデルの数学的構造 II

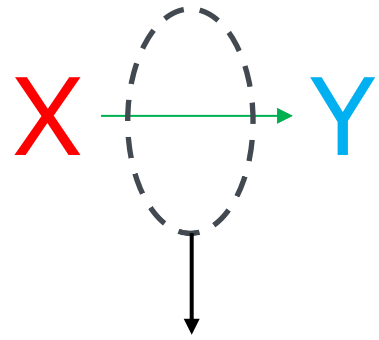
カテゴリーCの射の集まりを考える hom-set hom(X,Y)

カテゴリーCのオブジェクトXとオブジェクトYの間の射 $X \to Y$は、一般には、複数存在します。こうした射の集まりをhom(X,Y)と表し、hom-set と呼びます。

hom(X,Y)は、一般には複数の射を要素とする集合になります。



hom(X,Y)は集合



hom(X,Y)={*}

preorder categoryのように、オブジェクト間にたかだか一つの射しか存在しない場合、それは一つの要素からなる集合になります。

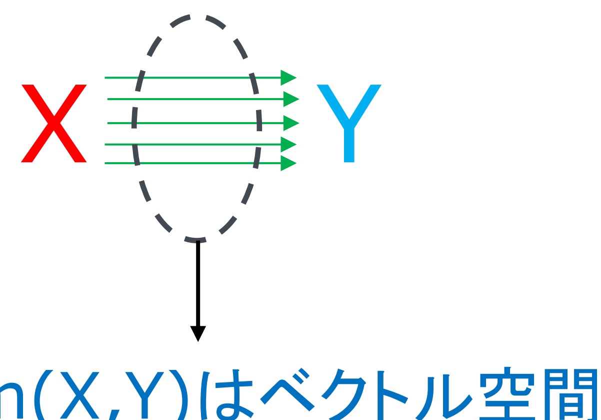
hom-setは、単なる集合以上の構造を持つことがあります。

ベクトル空間をオブジェクトとし、ベクトル空間の線形写像を射とするカテゴリーVectを考えます。

f,gがベクトル空間Xからベクトル空間Yへの線型写像とすれば、f+gもベクトル空間Xからベクトル空間Yへの線型写像となり、kがスカラーなら、kf もベクトル空間Xからベクトル空間Yへの線型写像となります。

$$f, g \in \mathrm{hom}(X, Y) \text{ なら } f + g \in \mathrm{hom}(X, Y) \text{ で } kf \in \mathrm{hom}(X, Y)$$

このことは、hom(X,Y)自身がベクトル空間であることを意味します。



カテゴリーCのhom-setが
カテゴリーVである場合を考える

カテゴリーCのhom-setがカテゴリーVである時、
CはVによってenrich化されたカテゴリーであると言います。

先に見た例だと、ベクトル空間のカテゴリーVectは、Vect自身によってenrich化されたカテゴリーだということになります。

一般的なCategoryは、hom-setの構造についてそれがSetである以外特に条件を設けていないのですが、CategoryはSetによってenrich化されたカテゴリーだと考えることが出来ます。

もっとも、すべてのhom(X,Y)が集合となるとは限りません。こうした条件を満たすカテゴリーを”locally small”と言います。

言語のカテゴリーLに確率を導入するということ

ただ、先のenrich化の説明では、どのようなカテゴリーVがカテゴリーCをenrich化できるのかは、何も語られていません。

enriched category を一般的に定義するには、もう少し準備が必要です。それについては、おいおい説明することにしましょう。

ここでは、当初の目的である「言語のカテゴリーLに確率を導入すること」がどういうことなのかを整理してみようと思います。

言語のカテゴリーLはpreorderですので、LのオブジェクトXからオブジェクトYの射は一つしか存在しません。

$X \longrightarrow Y$

それは、hom(X,Y)が一つの要素しか持たない集合であると表現されます。

$\mathrm{hom}(X,Y)=\{*\}$

Lに確率を導入するという事は、hom(X,Y)に確率を表す数字のラベルをつけるという事です。

$X \xrightarrow{\text{確率}} Y$

このラベルをhom(X,Y)と同一視すると、それは、[0,1]の値を取ることになります。

$\mathrm{hom}(X,Y) \in [0,1]$

この値は、先の*を確率で置き換えたものです。

$\mathrm{hom}(X,Y)=0.03$

このことは、enrichするカテゴリーが[0,1]であることを示しています。

まだまだ足りないことがある

ただ、ここで説明したことは、preorder category のhom-setの*を確率を表す数字で置き換えればいいのではということを述べただけです。

だいたい、enrich化されるカテゴリーのhom-setを勝手にenrich化するカテゴリーに置き換えて上手くいくのでしょうか？
enrich化できるカテゴリーには、条件があるはずです。

そもそも、[0,1]がそうした条件を満たすカテゴリーであることの説明も行われていません。

こうしたことを、次回あらためて説明したいと思います。

enrich以前のカテゴリーと
enrich化されたカテゴリーを比較する

大規模言語モデルの数学的構造 II

カテゴリー VがカテゴリーCをenrichする時 Vの役割をみる

カテゴリー VがカテゴリーCをenrichする条件は、知られています。それは、Vがcommutative monoidal preorderであることです。

ここでは、commutative monoidal preorderの説明を後に回して、天下り的に、category Cと、C がcommutative monoidal preorderであるVによってenrich化されたcategory – 短く、V-categoryと言うことがあります – 両者の対応を示してみようと思います。

この比較は、Vの役割を見るためのものです。

元のcategory C ⇒ C enriched over V

カテゴリーを構成するもの

- オブジェクト： $obj(C)$
- 射： $x, y \in obj(C)$について
  $\hom(x, y): x \to y$

射の満たすべき性質

- 同一射の存在：
  すべての$x \in obj(C)$について
  $id_x: x \to x$ が存在する。
- 射の合成の結合性：
  $f: x \to y$, $g: y \to z$, $h: z \to w$の
  時、$(h \circ g) \circ f = h \circ (g \circ f)$が成り立つ

カテゴリーを構成するもの

- オブジェクト： $obj(C)$
- 射： $x, y \in obj(C)$について
  $C(x, y) \in V$

$C(x, y) \in V$の満たすべき性質

- すべての$x \in obj(C)$について
  $1 \leq C(x, x)$
- すべての$x, y, z \in obj(C)$
  について
  $C(y, z) \otimes C(x, y) \leq C(x, z)$

元のcategory C → C enriched over V

カテゴリーを構成するもの

- オブジェクト：　$obj(C)$
- 射：$x, y \in obj(C)$について
  $$\hom(x, y): x \to y$$

射の満たすべき性質

- 同一射の存在：
  すべての$x \in obj(C)$について
  $id_x: x \to x$ が存在する。
- 射の合成の結合性：
  $f: x \to y\,, g: y \to z\,, h: z \to w$の時、$(h \circ g) \circ f = h \circ (g \circ f)$が成り立つ

カテゴリーを構成するもの

- オブジェクト：　$obj(C)$
- 射：$x, y \in obj(C)$について
  $$\mathrm{C}(x, y) \in V$$

$C(x, y) \in V$の満たすべき性質

- すべての$x \in obj(C)$について
  $$1 \leq C(x, x)$$
- すべての$x, y, z \in obj(C)$について
  $$C(y, z) \otimes C(x, y) \leq C(x, z)$$

元のcategory C → C enriched over V

カテゴリーを構成するもの

- オブジェクト: $obj(C)$
- 射: $x, y \in obj(C)$について
  $\hom(x, y) : x \to y$

射の満たすべき性質

- 同一射の
  すべての$x \in obj(C)$について
  $id_x : x \to x$ が存在する。
- 射の合成の結合性:
  $f : x \to y$, $g : y \to z$, $h : z \to w$の時、$(h \circ g) \circ f = h \circ (g \circ f)$が成り立つ

カテゴリーを構成するもの

- オブジェクト: $obj(C)$
- 射: $x, y \in obj(C)$について
  $C(x, y) \in V$

$C(x, y) \in V$の満たすべき性質

$I \le C(x, x)$

- すべての$x, y, z \in obj(C)$について
  $C(y, z) \otimes C(x, y) \le C(x, z)$

二つのカテゴリーのオブジェクトは同一である

元のcategory C → C enriched over V

カテゴリーを構成するもの

- オブジェクト:　$obj(C)$
- 射:　$x, y \in obj(C)$について

  $\hom(x, y) : x \to y$

射の満たすべき性質

- 同一射の存在:

  すべての$x \in obj(C)$について

  $id_x : x$ -
- 射の合

  $f : x \to$

  時、$(h$

  り立つ

カテゴリーを構成するもの

- オブジェクト:　$obj(C)$
- 射:　$x, y \in obj(C)$について

  $C(x, y) \in V$

$C(x, y) \in V$の　たすべき性質

- す　$obj(C)$について

  $\leq C(x, x)$
- $j(C)$

  $\leq \ C(x, z)$

$C(x, y)$は、Vの要素である。
射はVによって置き換えられる
これを、V-hom オブジェクトという。

元のcategory C → C enriched over V

カテゴリーを構成するもの

- オブジェクト: $obj(C)$
- 射: x, y …
  hom…

射の満たすべき性質

- **同一射の存在**:
  すべての$x \in obj(C)$について
  $id_x : x \to x$ が存在する。
- **射の合成の結合性**:
  $f : x \to y$, $g : y \to z$, $h : z \to w$の時、$(h \circ g) \circ f = h \circ (g \circ f)$が成り立つ

カテゴリーを構成するもの

- オブジェクト: $obj(C)$
- 射 … について
  … V

$C(x, y)$ …の満たすべき性質

- すべての$x \in obj(C)$について
  $1 \leq C(x, x)$
- すべての$x, y, z \in obj(C)$について
  $C(y, z) \otimes C(x, y) \leq C(x, z)$

これが同一射の存在を表す

元のcategory C → C enriched over V

カテゴリーを構成するもの

- オブジェクト: $obj(C)$
- 射: $x, y \in obj(C)$について
  $\hom(x, y) : x \to y$

射の満たすべき性質

- 同一射の存
  すべての$x \in obj(C)$について
  $id_x : x \to x$ が存在する。
- 射の合成の結合性:
  $f: x \to y\ , g: y \to z\ , h: z \to w$の
  時、$(h \circ g) \circ f = h \circ (g \circ f)$が成
  り立つ

カテゴリーを構成するもの

- オブジェクト: $obj(C)$
- 射: $x, y \in obj(C)$について
  $C(x, y) \in V$

$C(x, y) \in V$の満たすべき性質

- ついて
  (x, x)
- すべての$z \in obj(C)$
  について
  $C(y, z) \otimes C(x, y) \leq C(x, z)$

これが射の合成の結合性

C enriched over V

カテゴリーを構成するもの

- オブジェクト：　$obj(C)$
- 射：$x, y \in obj(C)$について
  $C(x,y) \in V$

$C(x,y) \in V$の満たすべき性質

- すべての$x \in obj(C)$について
  $1 \leq C(x,x)$
- すべての$x, y, z \in obj(C)$
  について
  $C(y,z) \otimes C(x,y) \leq C(x,z)$

commutative monoidal preorder

大規模言語モデルの数学的構造 II

symmetric monoidal preorder と commutative monoidal preorder

ここでは、カテゴリーVがenrichするカテゴリーである条件を述べたいと思います。

ただ、少し困ったことがあります。enriched category理論の論文を読むと、その条件は、Vがsymmetric monoidal preorder であることと述べられているのですが、Tai-Danaeらの論文では、Vがcommutative monoidal preorder であるとして議論が展開されています。

symmetric: $y \otimes x \leq x \otimes y$

commutative: $y \otimes x = x \otimes y$

の違いだと思います。

その意味では、symmetric monoidal preorder と commutative monoidal preorder の概念は異なるものです。symmetric monoidal preorder の方が広い概念です。

ただ、順序がpreorderではなく poset だとすると、

$y \otimes x \leq x \otimes y$ かつ $y \otimes x \leq x \otimes y$ から $y \otimes x = x \otimes y$ が導かれるので、両者の概念は一致します。

Tai-Danaeらの論文が扱っている言語の文字列表現の順序関係は、もちろん preorderでもありますが、posetです。

ですので、あの論文で、enrich化するカテゴリーとして commutative monoidal preorder で議論を進めているのは間違っている訳ではありません。

基本的に、論文の記述に沿って解説を行いたいと思います。

commutative monoidal preorder

定義1

commutative monoidal preorder $(V,\leq,\otimes,1)$ は、preorder $(V,\leq)$ 上の、次の条件を満たす commutative monoid $(V,\otimes,1)$ である。

$$x \leq x' \text{ かつ } y \leq y' \text{ならば、} x \otimes y \leq x' \otimes y'$$

この条件は、順序だけからなるpreorder $(V,\leq)$ の世界を、二項演算だけからなるmonoid $(V,\otimes,1)$の世界を結びつけて、monoidal preorder $(V,\leq,\otimes,1)$を構成する役割を持っている。

preorder $(V, \leq)$

順序≤が定義された集合Vで、Vに属する任意の要素 a, b, c が次の性質を持つとき、順序≤ はpreorder（前順序）と言われる。

1. $a \leq a$　（反射律）
2. $a \leq b$ で $b \leq c$ なら $a \leq c$　（推移律）

commutative monoidal preorderの定義

preorder $(V, \leq)$ 上のcommutative monoidal preorder $(V, \leq, \otimes, 1)$は、次の構成要素を持つ。

- monoidal unit: $1 \in V$
- monoidal product: $V \otimes V \to V$

これらは次の性質を持つ

1. $x \leq x'$ かつ $y \leq y'$ならば、$x \otimes y \leq x' \otimes y'$
2. $1 \otimes \mathrm{x} = \mathrm{x}$ かつ $\mathrm{x} \otimes 1 = x$
3. $(x \otimes y) \otimes z = x \otimes (y \otimes z)$
4. $x \otimes y = y \otimes x$

commutative monoidal preorder
ノテーション

$$(\boldsymbol{V}, \leq, \otimes, \mathbf{1})$$

preorder $(V, \leq)$ 上のcommutative monoidal preorderは、次の構成要素を持つ。

- monoidal unit: $1 \in V$
- monoidal product: $V \otimes V \to V$

V-enriched category

定義 2

$(V, \leq, \otimes, 1)$ を、commutative monoidal preorder とする。

この時、Vでenrich化されたカテゴリー（単純に、V-カテゴリーと呼んでもいい）を、次の様に定義する。

- V-カテゴリーは、Cと同じオブジェクトを持つ。
- また、V-カテゴリーは、その任意のオブジェクトのペア x, y について、V-homオブジェクトと言われる $C(x, y) \in V$を持つ。
- このV-homオブジェクトは、すべての$x, y, z \in V$について、次の条件をみたす。

$$1 \leq C(x, x)$$

$$C(y, z) \otimes C(x, y) \leq C(x, z)$$

単位区間 [0,1]は commutative monoidal preorder である

monoidal product $\otimes$ を普通の実数間の掛け算とし、monoidal unit を 1 とし、preorderの順序関係$\leq$ を 通常の実数間の大小関係とすればいい。

この単位区間 [0,1] で enrich化されたカテゴリー、[0,1]-カテゴリーは、Cと同じオブジェクトを持ち、すべての$x, y \in C$で定義された[0,1]に値を取る関数$(x, y) \longmapsto C(x, y)$を持つ。

この$C(x, y)$は、次の条件を満たす。

すべての$x \in C$ について、$C(x, x) = 1$

すべての$x, y, z \in C$ について、$C(y, z)C(x, y) \leq C(x, z)$

closed commutative monoidal preorder と internal hom

大規模言語モデルの数学的構造 II

closedなcommutative monoidal preorder

定義 3

commutative monoidal preorder Vは、すべての$x, y \in V$に対して、次の条件を満たす $[x, y] \in V$が存在する時、closed であると言う。

$$x \otimes y \leq z \ \ if\ and\ only\ if\ \ x \leq [y, z]$$

この $[x, y] \in V$ を、Vの internal hom という。

$(x, y) \in V$ なら $[x, y] \in V$ 、すなわち、このinternal homについて Vが閉じている時、Vはclosed commutative monoidal preorder と言われるということである。

closed commutative monoidal preorderの性質

closed commutative monoidal preorder は、次の様な性質を持つ。

Vのhom set V(x,y)を、Vのオブジェクトであるinternal hom $[x, y]$に置き換える。この時、この割り当て$(x, y) \longmapsto [x, y]$によって、Vは自分自身でenrich化されたカテゴリーであることを示すことが出来る。

V-enriched category の定義を振り返る

定義 2

$(V, \leq, \otimes, 1)$ を、commutative monoidal preorder とする。
この時、Vでenrich化されたカテゴリー（単純に、V-カテゴリーと呼んでもいい）を、次の様に定義する。

- V-カテゴリーは、Cと同じオブジェクトを持つ。
- また、V-カテゴリーは、その任意のオブジェクトのペア x, y について、V-homオブジェクトと言われる $C(x, y) \in V$を持つ。
- このV-homオブジェクトは、すべての$x, y, z \in V$について、次の条件をみたす。

$$1 \leq C(x, x)$$
$$C(y, z) \otimes C(x, y) \leq C(x, z)$$

$$1 \leq C(x,x)$$
$$C(y,z) \otimes C(x,y) \leq C(x,z)$$

$$1 \leq [x,x]$$
$$[y,z] \otimes [x,y] \leq [x,z]$$

であることを示せばいい。

最初の、$1 \leq [x,x]$であることは、次のことからわかる。

まず、$1 \otimes x \leq x$ はいえる。
$x \otimes y \leq z$ は $x \leq [y,z]$ と同値なのだから、
$1 \otimes x \leq x$ は、$1 \leq [x,x]$ と同値である。

二つ目の、$[y, z] \otimes [x, y] \leq [x, z]$は、次の様にしてわかる。

$[y, z] \leq [y, z]$ は言える。左辺の$[y, z]$ をxとすると、$x \leq [y, z]$ 。
この式は$x \otimes y \leq z$ と同値である。よって、$[y, z] \otimes y \leq z$
同様に、$[x, y] \leq [x, y]$ から、$[x, y] \otimes x \leq y$ もいえる。

$[y, z] \leq [y, z]$ と $[x, y] \otimes x \leq y$ の二つの式から

$$[y, z] \otimes ([x, y] \otimes x) \leq [y, z] \otimes y$$

$(\because x \leq x'$ かつ $y \leq y'$ならば、$x \otimes y \leq x' \otimes y')$

先に見た様に、$[y, z] \otimes y \leq z$ であるので、

$[y, z] \otimes [x, y] \otimes x \leq [y, z] \otimes y \leq z$ よって、

$$[y, z] \otimes [x, y] \otimes x \leq z$$

これから、$[y, z] \otimes [x, y] \leq [x, z]$ がわかる。

$(\because x \otimes y \leq z$ は $x \leq [y, z]$ と同値)

単位区間[0,1]上のenrich化での internal hom

前回のセッションで見たように、単位区間[0,1]は、monoidal product $\otimes$ を普通の実数間の掛け算とし、monoidal unit を 1 とし、preorderの順序関係$\leq$ を 通常の実数間の大小関係とする、commutative monoidal preorder である、

補題 1

単位区間[0,1]で、すべての$a, b \in [0,1]$に対して、次のような切り捨て割算$[a, b]$を定義する。

$$[a, b] := \begin{cases} b/a & \textit{if } b < a, \\ 1 & \textit{otherwise}. \end{cases}$$

この時、単位区間[0,1]は、$[a, b]$をinternal homとするclosed commutative monoidal preorderになる。

先の切り捨て割り算の定義が、internal homの条件

$$x \otimes y \le z \ \ if\ and\ only\ if\ \ x \le [y, z]$$

を満たすことを示せばいい。

ここでは、任意の$a, b \in [0,1]$について、

$$ab \le c\ if\ and\ only\ if\ a \le [b, c]$$

を示せばいい。

二つの場合で考える。

$c < b$ の場合：$ab < c$ は、$a \le \frac{c}{b} = [b, c]$を導く。

$b \le c$ の場合：$[b, c] = 1$、よって$a \le [b, c]$

Part 4

enriched category論の言語理論への応用

大規模言語モデルの数学的構造 II

Part 4

enriched category論の言語理論への応用

- Syntax category
- enriched functor とenriched copresheaf
- copresheaf $[0,1]^L$とenriched Yoneda lemma
- Semantic category
- copresheaf-意味論の射程

enriched category論の言語理論への応用

Syntax category

大規模言語モデルの数学的構造 II

Syntax category L

定義 4

Syntax category L を [0,1]上でenrich化されたカテゴリーとして、次のように定義する。

- Lのオブジェクトは、言語の表現である。
- Lの[0,1]–オブジェクトは、

$$L(x, y) := \pi(y|x)$$

  である。

ここで、$\pi(y|x)$は、表現 yが表現 x の拡大である確率である。

$L(x, y)$ の例

$$L(red, red\ firetruck) = 0.02$$
$$L(red, red\ idea) = 10^{-5}$$
$$L(red, blue\ sky) = 0$$

短い表現については、こうした確率を、大規模言語モデルが学習できることは明らかだろう。

Transformer のSelf Attention は、そういう働きをする。

同時に、ここで得られた確率の情報は、言語の文法についての情報を与える。

Tai-Danaeらが、もはやpreorderのカテゴリーではないこのカテゴリーを、”Syntax category”と呼ぶのは、その為であろう。

長い表現 x,yについて、大規模言語モデルは、確率$L(x,y)$ を学習できるか？

表現x,yの関係がもはや「文法的」関係とは言えない、長い表現x,yの場合に、大規模言語モデルは、確率$L(x,y)$ を学習できるだろうか？

例をあげよう。表現 y=“I am going to the store to buy a can of” の後ろに”cat food” が連続して、x=“I am going to the store to buy a can of cat food” となる確率L(x,y) を、大規模言語モデルは評価出来る。

それは、「翻訳モデル」にはなかった「大規模言語モデル」の大きな特徴である。

この能力は、主要に”Next Sentence Prediction”を通じた学習によって支えられている。

enriched functor と enriched copresheaf

大規模言語モデルの数学的構造 II

このセッションの構成

このセッションの目的は、カテゴリー C からカテゴリーDへのfunctor $C \to D$ であるcopresheaf D^Cのenrich版を構成することです。

まず、functorの定義を振り返りながら、enrich化されたfunctorを定義します。

次に、functorカテゴリーの定義とcopresheafの定義を振り返りながら、enrich化されたcopresheafの定義を考えていきます。

functorの定義の振り返り

まず、functor の定義を振り返っておきましょう。

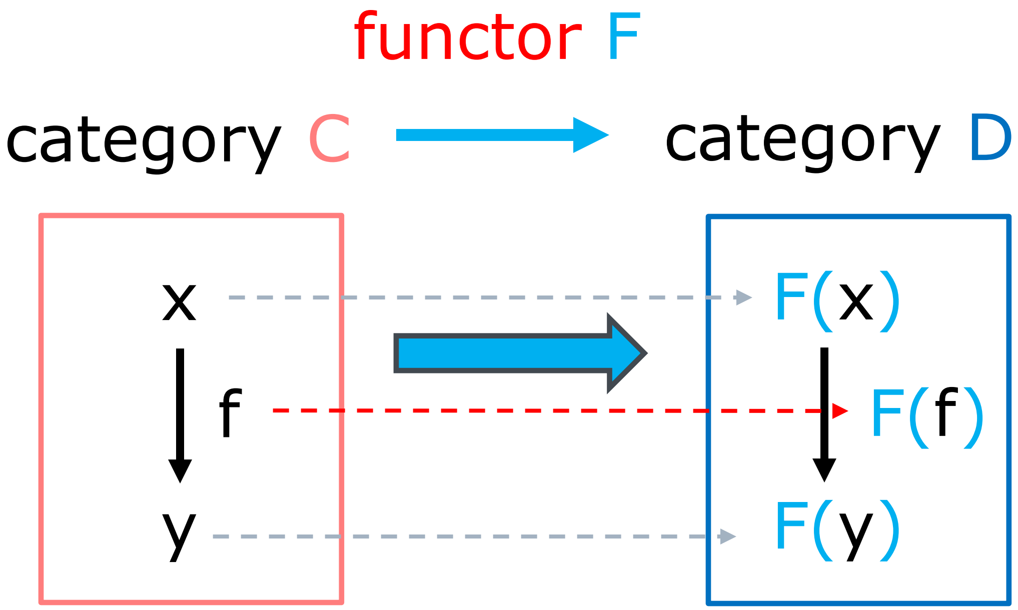
category C から category D へのfunctor $F\colon C \rightarrow D$ は、次のように定義されます。

Cのオブジェクトと射について、

- Cのすべてのオブジェクトx について、
  $F(x)$はDのオブジェクトである。
- Cのすべての射 $f : x \rightarrow y$ について、
  $F(f)$はDの射 $F(f) : F(x) \rightarrow F(y)$ である。

要するに、functor Fによって、CのオブジェクトはDのオブジェクト F(x)にうつされ、Cの射 fはDの射 F(f)にうつるということです。

図で表すと、次のようになります。



enriched functor

定義 5

CとDを、あるcommutative monoidal preorder (V,⊗,≤, 1)上でenrich化されたカテゴリーとする。

この時、Cのすべてのオブジェクト x, yについて

$$C(x,y) \leq D(fx, fy)$$

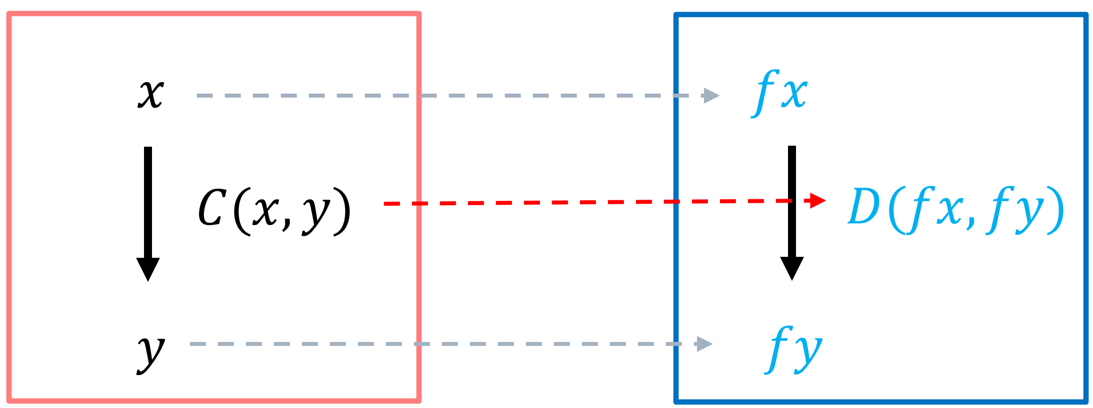
を満たす関数 $f: C \to D$ を、enriched functorと呼ぶ。

enriched functor

$$C(x,y) \leq D(fx,fy)$$

enriched functor f

enriched category C → enriched category D



functor category D^Cと copresheaf の振り返り

重要なことは、functor は、それ自身category とみなすことができるということです。

一般に category C から category D へのfunctor から構成されるcategory を、functor category といい、D^Cで表します。（ [C, D] という表記を使うこともあります。）

- category D^C のオブジェクトは、CからDへのfunctor
- category D^C の射は、CからDへのfunctor間のnatural transformation

で、functor category を定義できます。

集合に値を持つ表現可能なfunctorである意味のcategory Set^Lを、L上のcopresheafといいます。

意味の$category$ Set^L

copresheaf

Set^L

$L(x, -)$

$L(z, -)$

$L(a, -)$

$L(y, -)$

すべてのLのオブジェクトxについて
L(x, −)なるfunctorの集まり

enriched copresheaf の定義

Vがclosed commutative monoidal preorder の時、Vは自分自身でenrich化されたカテゴリーになります。

先の、enrich化されたfunctor $C \to D$ で$D = V$とした enrich化したfunctor $C \to V$を考えます。

定義 6

Cを、closed commutative monoidal preorder V上でenrich化されたカテゴリーとする。

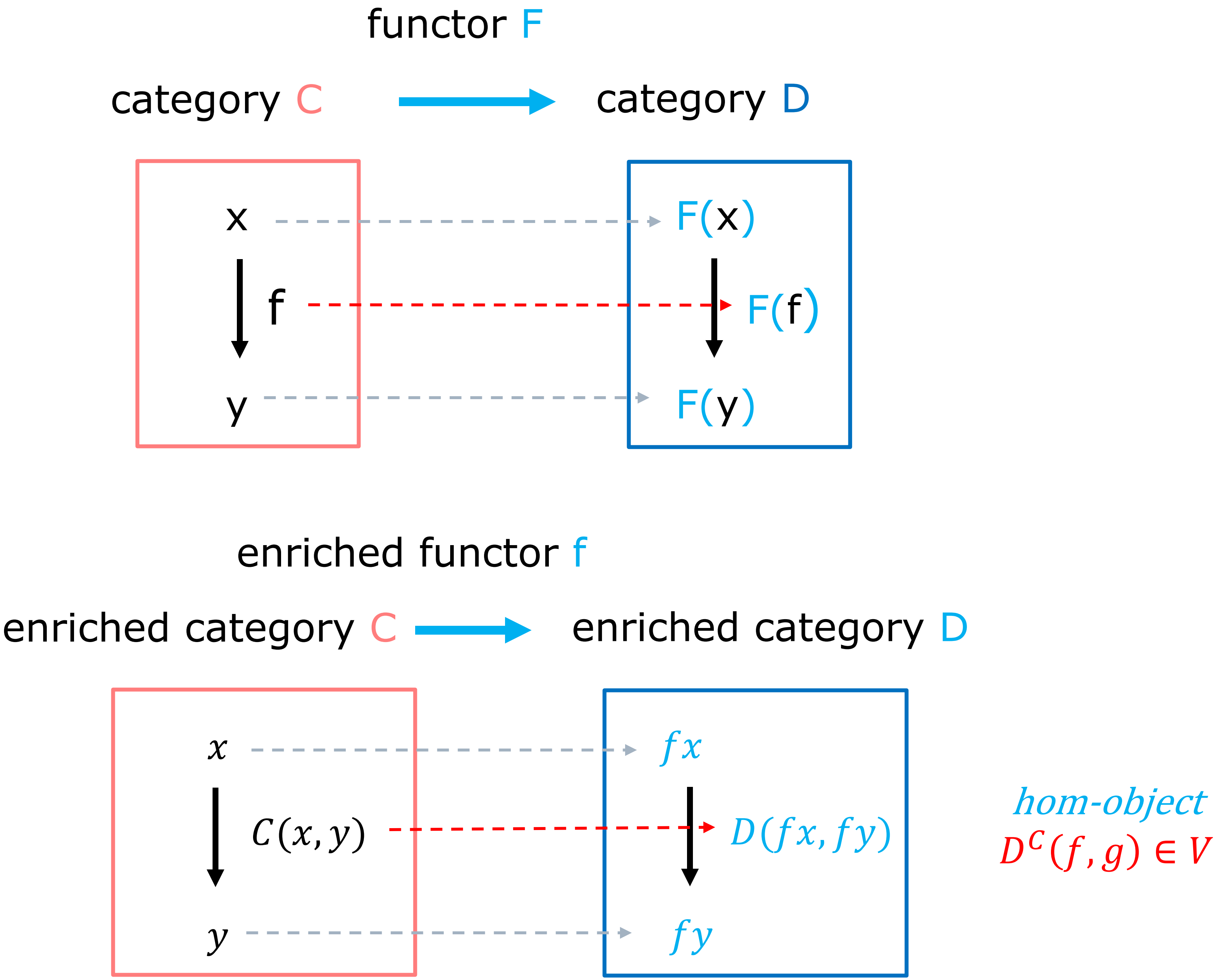
enrich化されたcopresheaf V^Cは、Cのすべてのオブジェクトx, yについて$C(x, y) \le V(fx, fy)$を満たす関数で定義される。

そのオブジェクトが、CからDへのV-functorである Vでenrich化されたカテゴリーD^Cを構成するには、$f, g : C \to D$であるすべてのfunctor f, g について、hom object $D^C(f, g) \in V$が必要です。

通常のカテゴリーでは、カテゴリーの間の写像が functor category を定義するように、

enrich化されたカテゴリーでは、enrich化されたカテゴリー間の写像が、enrich化されたfunctor category を定義します。

（base categoryが、closedでcompleteな場合）



functor F
category C
category D
x
f
y
F(x)
F(f)
F(y)
enriched functor f
enriched category C
enriched category D
x
$C(x,y)$
y
fx
$D(fx,fy)$
fy
hom-object
$D^C(f,g) \in V$

Lが[0,1]上のenriched category であることの形式的説明

L(x,y)が、enrichするカテゴリーの V-hom オブジェクトの条件

$$1 \leq C(x,x)$$

$$C(y,z) \otimes C(x,y) \leq C(x,z)$$

を満たしていることを示す。

$$1 \leq L(x,x)$$

$$\mathrm{L}(y,z) \otimes L(x,y) \leq L(x,z)$$

を示せばいい。

$$L(x,y) \coloneqq \pi(y|x)$$

である。

$\pi(y|x)$は、xの下でのyの条件付き確率であるので、

$$\pi(x,x) = 1$$
$$\pi\ (z|y)\pi(y|x) = \pi(z|x)$$

L(x,y)は、enrich化するカテゴリーの [0,1]-オブジェクトの性質を満たしている。

copresheaf $[0,1]^L$ とYoneda Lemma

大規模言語モデルの数学的構造 II

end

そのオブジェクトが、CからDへのV-functorである Vでenrich化されたカテゴリーD^Cを構成するには、$f, g : C \to D$であるすべてのfunctor f, g について、hom object $D^C(f, g) \in V$が必要です。

形式的には、それは、Vのlimitの一つであるend として、次のように定義されます。

$$D^C(f, g) = \int^{c \in C} D(fc, gc)$$

complete

通常のカテゴリーでは、カテゴリーの間の写像が functor category を定義するように、

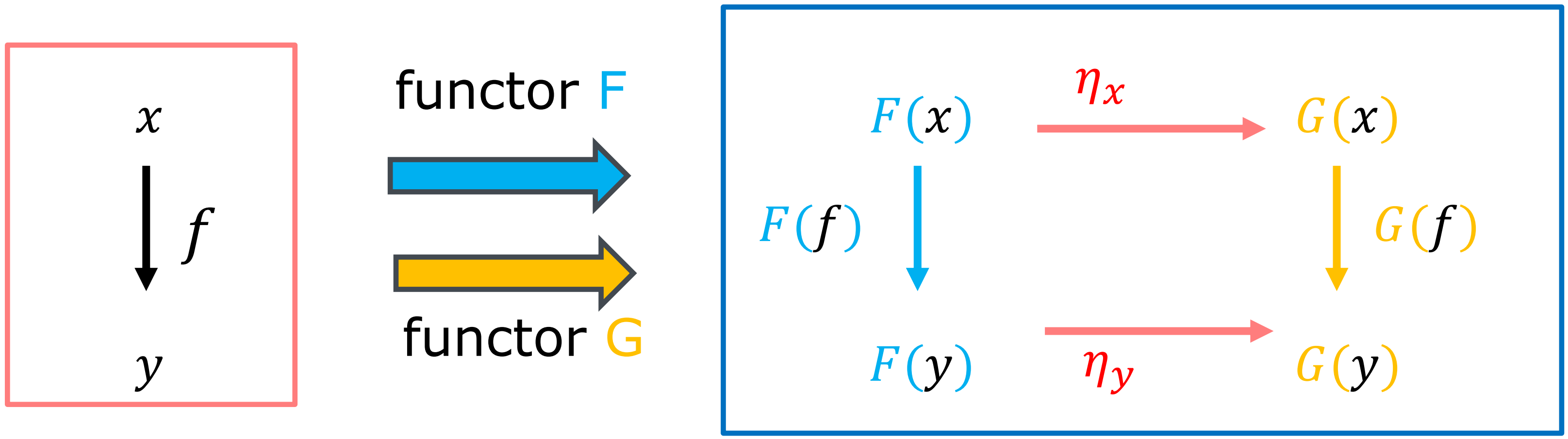
enrich化されたカテゴリーでは、enrich化されたカテゴリー間の写像が、enrich化されたfunctor category を定義します。

（base categoryが、closedでcompleteな場合）

あるカテゴリーCがcompleteだというのは、Cがlimitを持つことである。

functorの間の natural transformation

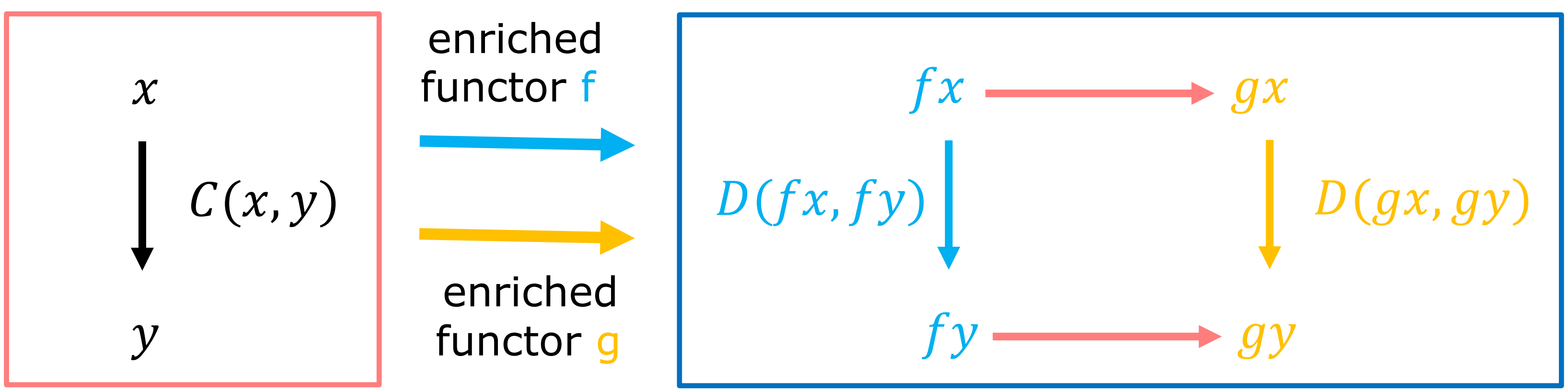
category C　　　　　　　　　　category D



$$G(f) \circ \eta_x = \eta_y \circ F(f)$$

enriched functorの間の natural transformation

enriched category C $\longrightarrow$ enriched category D



copresheaf $[0,1]^C$は [0,1]上でenrich化されたカテゴリーか？

補題 2

Cを[0,1]上でenrich化されたカテゴリーとする。この時、copresheaf $\hat{C} := [0,1]^C$も [0,1]上でenrich化されたカテゴリーとなる。

二つの copresheaf 間の [0,1]-object である $f, g: C \to [0,1]$ は、cがCのオブジェクトすべてを走る時、次の下限で与えられる。

$$\hat{C}(f, g) = \inf_{c \in C}\{[fc, gc]\}$$

単位区間[0,1]のinternal hom は、切り捨て割り算になります。

$$[a,b] := \begin{cases} b/a & \mathit{if}\, b < a, \\ 1 & \mathit{otherwise}. \end{cases}$$

だから、

$$\hat{C}(f,g) = \inf_{c\in C}\{[fc, gc]\} = \inf_{c\in C}\{1. gc/fc\}$$

です。

$C(x,-)$について

補題 3

Cを[0,1]-category とする。この時、Cのすべてのオブジェクトx について、関数

$$h^x := C(x,-)$$

は、[0,1]-functor となる。

V=[0,1]とし、Dも[0,1]とする。定義 5 で見たように、

関数 $C \to [0,1]$が、enrich化されたfunctorである条件は、

$$C(x,y) \le D(fx,fy)$$

である。

$$C(c,d) \le [h^x(c), h^x(d)]$$

が成り立っているか、チェックすればいい。

$C(c,d) \leq [h^x(c), h^x(d)]$は、$C(c,d) \ \leq \ [C(x,c), C(x,d)]$ である。
$a < [b,c]$は$ab < c$ と同値なので、

$$C(c,d)C(x,c) \ \leq \ C(x,d)$$

をチェックすればいい。

Cは、[0,1]上でenrich化されたカテゴリーなので、この条件は成り立っていることがわかる。

copresheafの表現可能性

定義 7

[0,1]-category のすべてのオブジェクトx について、functor $h^x \coloneqq C(x, -)$ をxで表現されたcopresheaf と呼ぶ。

もし、ある$x \in C$について $f = h^x$ なるfが、存在するなら

copresheaf $f: \mathrm{C} \to [0,1]$は表現可能であるという。

[0, 1]-enriched Yoneda lemma.

定理 1

(Enriched Yoneda Lemma)

[0,1]-category Cのすべてのオブジェクトxと、すべての[0,1]-copresheaf $f: C \to$[0,1]について、次が成り立つ。

$$\hat{C}(h^x, f) = f(x)$$

Cのオブジェクト xと、copresheaf f を固定する。

$\hat{C}(h^x, f) = \inf_{c \in C}\{[h^x(c), fc]\}$なので、どんな$c \in C$を選んでも、

$\hat{C}(h^x, f) \leq [h^x(c), fc]$ は成り立つ。c＝xとする。

$$\hat{C}(h^x, f) \leq \ [h^x(x), fx] \ = \ [1, fx] \ = \ fx$$

一方、fは、$f : C \to [0,1]$なる[0,1]-functor なので、

$$C(x, c) \leq [fx, fc]$$

これは、$C(x, c)fx \leq fc$と同値である。これはまた

$$fx \leq [C(x, c), fc]$$

と等しい。

$fx \leq [C(x, c), fc]$が与えられていると、すべての$c \in C$について

$$fx \leq \inf_{c \in C}\{[h^x(c), fc]\} = \hat{C}(h^x, f)$$

よって、

$$\hat{C}(h^x, f) = f(x)$$

系 1

[0,1]-category Cで、すべてのオブジェクトx,yについて、

$$C(y,x)=\hat{C}(h^x,h^y)$$

定理 1で、$f=h^y$ とおく。

$$\hat{C}(h^x,h^y)=h^y(x)=C(y,x)$$

系 2

すべての[0,1]-category Cで、割り当て$x \longmapsto h^x$は、enrich化されたfunctor $C^{op} \to \hat{C}$を定義する。

このfunctorは、C^{op}を$\hat{C}$のenrich化されたsubcategoryとして埋め込む。

C^{op}のopは、oppositeを表す。C^{op}のオブジェクトはCと同一だが、C^{op}の射の向きはCと逆向きである。

系 1で見たように、$x \longmapsto h^x$は、射を逆向きにするからである。

Semantic category

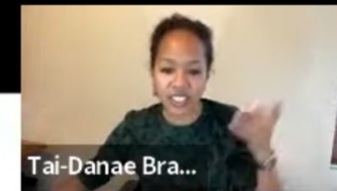
大規模言語モデルの数学的構造 II

copresheaf
XからYへの関数の集合 {X→Y} を考えることの意味

数学的対象Xの性質がよく分からないけど、数学的対象Yの性質・構造はよくわかっているとしましょう。

そうした時、Xはそうでなくても、XからYへの関数の集合{X→Y}を考えると、それがはっきりした構造を持つことがよくあります。

例えば、任意の集合上の実数値関数はベクトル空間を形成します。



THEME

If "blah" has nice structure, then functions from a set into "blah" have nice structure, too.

$$\mathbb{R}^X := \{f\colon X \to \mathbb{R}\} = \left\{ \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix} \right\} \cong \mathbb{R}^3$$

Vector spaces are an example. A similar story holds in category theory....

So if I look at all functions, from X into the reals.

VECTOR SPACE	(CO)PRESHEAF CATEGORY
$\mathbb{R}^X$	$\mathsf{Set}^{\mathsf{C}}$

Algebra and category

単位区間[0,1]は、カテゴリー理論の観点から見ると豊かな構造を持っています。それは closed commutative monoidal であり, complete で cocompleteです。

単位区間[0, 1]でenrich化された任意のカテゴリーCに対して、Cから単位区間へのfunctor $C \to [0,1]$のカテゴリーとして、C上のcopresheaf $\hat{C} = [0,1]^C$ を考えます。

このcopresheaf $\hat{C}$ は、単位区間[0,1]の豊かな構造を継承します。それは、単位区間[0,1] と同様に、product, coproduct, internal hom のenrich化されたバージョンを持っています。

補題 2 は、任意の [0,1]-category Cは、豊かな構造を持つ[0,1]-category copresheaf $\hat{C} = [0,1]^C$ に埋め込まれることを示しています。

Syntax category Lの定義の振り返り

ここでは、定義 4 で定義した、[0,1]上でenrich化されたSyntax category L上のpresheaf を考えます。

定義 4

Syntax category L を [0,1]上でenrich化されたカテゴリーとして、次のように定義する。

- Lのオブジェクトは、言語の表現である。
- Lの[0,1]−オブジェクトは、

$$L(x, y) \coloneqq \pi(y|x)$$

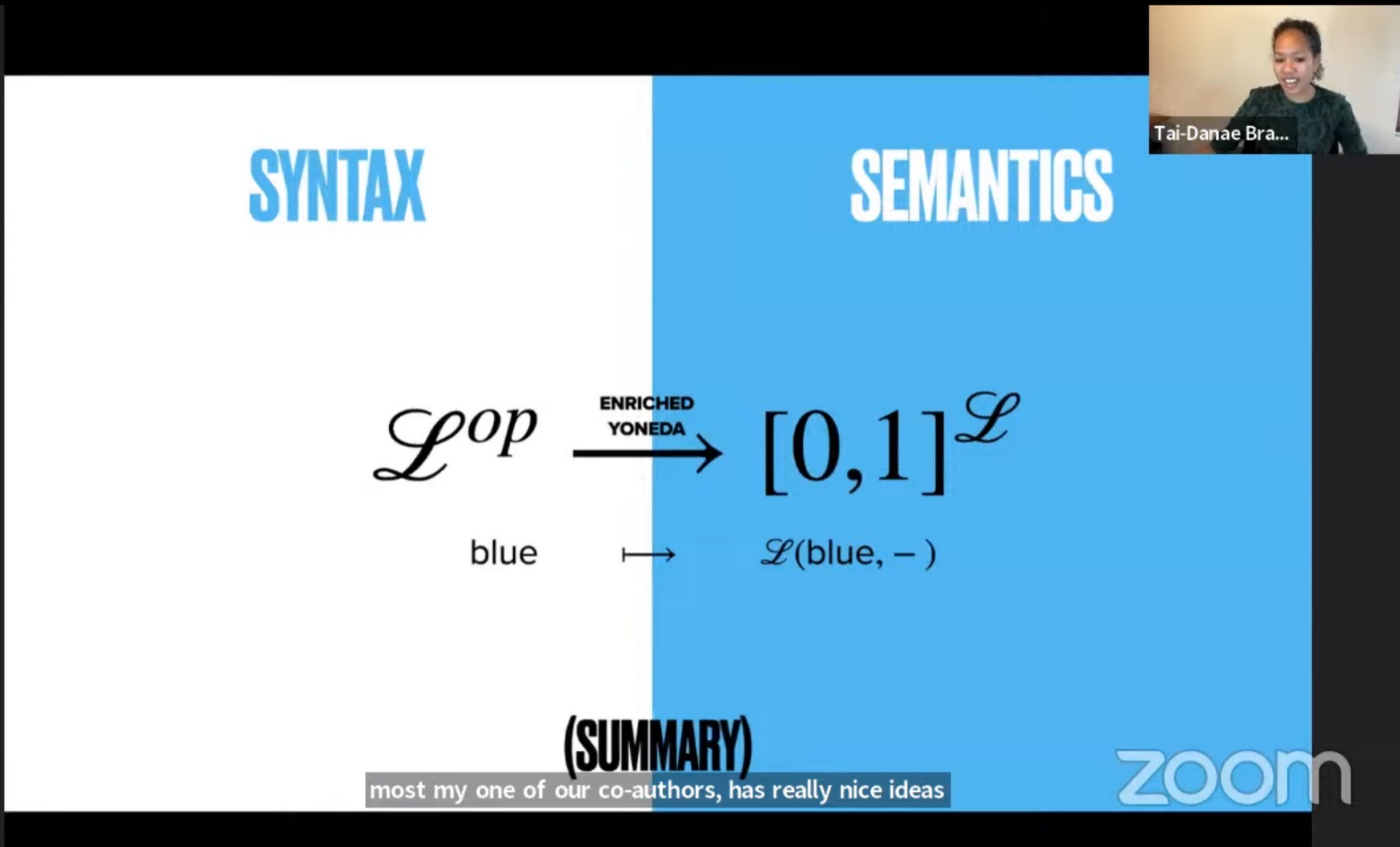
である。
ここで、$\pi(y|x)$は、表現 yが表現 x の拡大である確率である。

Semantic category

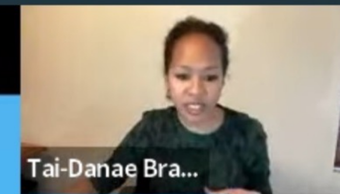
定義 8

LをSyntax category とする。この時、Semantic category を
$\hat{L} = [0,1]^L$ と定義する。

Semantic category $\hat{L}$は、[0,1]-categoryである L 上で
[0,1]でenrich化されたcopresheaf の[0,1]-categoryである。


SYNTAX
SEMANTICS
$\mathscr{L}^{op} \xrightarrow[\text{YONEDA}]{\text{ENRICHED}} [0,1]^{\mathscr{L}}$
blue $\longmapsto$ $\mathscr{L}(\text{blue}, -)$
(SUMMARY)
Tai-Danae Bra...
most my one of our co-authors, has really nice ideas

以前のバージョン



SYNTAX **SEMANTICS**

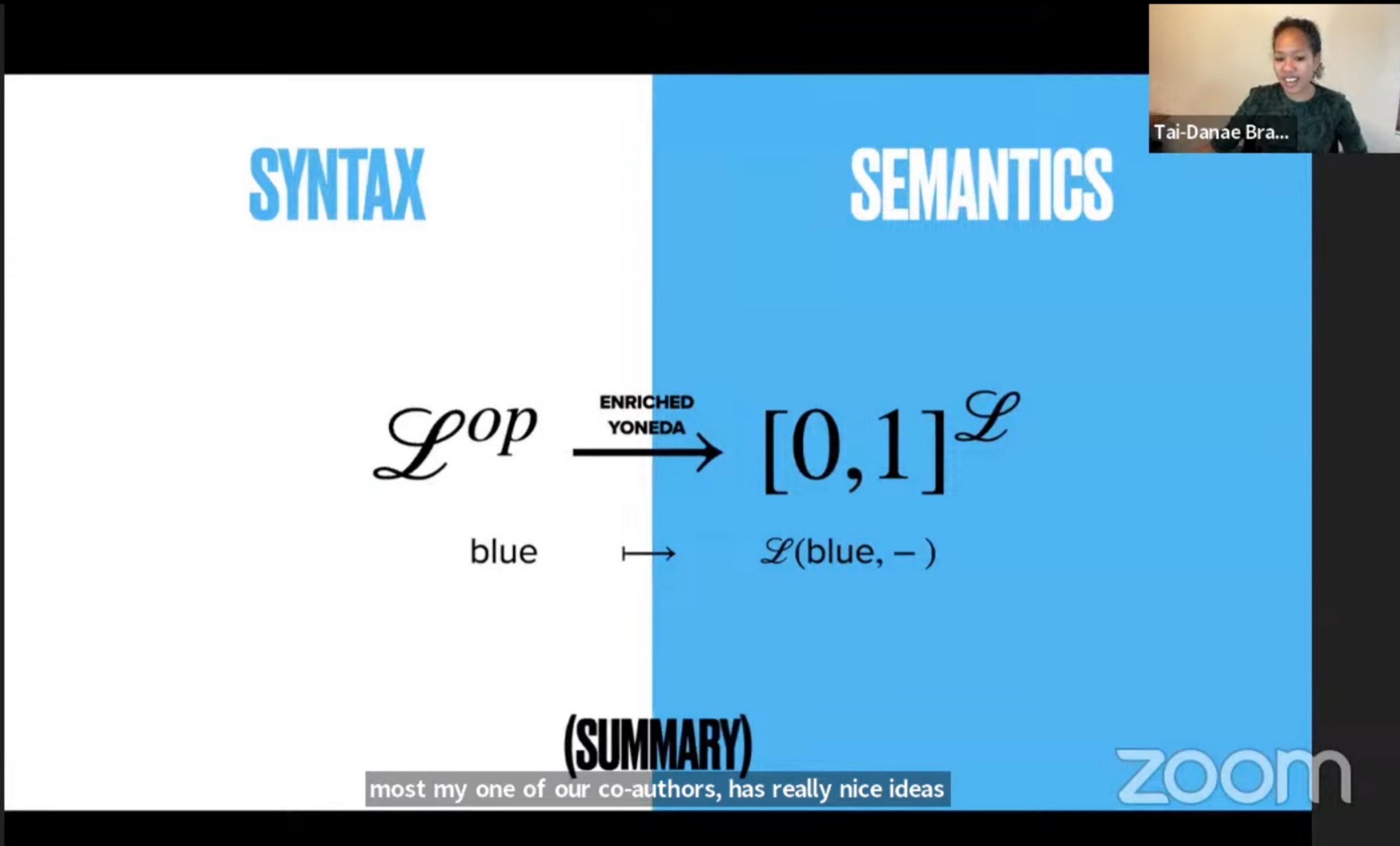
$$\mathsf{L}^{op} \xrightarrow{\text{YONEDA}} \mathsf{Set}^{\mathsf{L}}$$

$$\text{blue} \longmapsto \mathsf{L}(\text{blue}, -)$$

(SUMMARY)

But there are other limits, and co limits. So the point

現在のバージョン



SYNTAX
SEMANTICS
Tai-Danae Bra...
$\mathscr{L}^{op} \xrightarrow{\text{ENRICHED YONEDA}} [0,1]^{\mathscr{L}}$
blue $\longmapsto$ $\mathscr{L}$(blue, –)
(SUMMARY)
most my one of our co-authors, has really nice ideas

xで表現可能なcopresheaf $h^x := L(x,-)$

Lのオブジェクトであるすべての表現xについて、xで表現可能なcopresheaf $h^x := L(x,-)$を 表現xを拡張する条件付き確率として次のように定義する。

$$c \mapsto h^x(c) := \begin{cases} \pi(c|x) & \text{if } x \leq c \\ 0 & \text{otherwise.} \end{cases}$$

ここで、"$x \leq c$"は。「xは、cの部分文字列に含まれている」ことを表している。

表現xの意味

表現xの意味を、表現可能なenrich化されたcopresheaf

$$h^x := L(x, -)$$

であると考える。

h^xは、xを含むすべての可能な表現上で定義される。

このことは、あるテキストの意味は、そのテキストを含みうる可能なすべてのコンテキストによって、可能的には変化しうることを意味する。

補題 2の埋め込みは、テキストxをその意味h^xに割り当てる。

[0, 1]-enriched Yoneda Lemmaの逆向きの射の解釈

[0, 1]-enriched Yoneda Lemmaの逆向きの射、それがcontravariant functorであることの意味を考えてみよう。

テキストyをテキストxを拡大したもので、$L(x, y) = a \neq 0$ だとしよう。図で書くとYoneda Lemma は、次のことを主張している。

$$L$$

$$x \xrightarrow{a} y$$

$$\hat{L}$$

$$h^y \xrightarrow{a} h^x$$

L

$$x \xrightarrow{a} y$$

$\hat{L}$

$$h^y \xrightarrow{a} h^x$$

左は、Syntax category Lで、テキストyはテキストxの拡大であることを表している。

右は、Semantic category $\hat{L}$ で、h^xはxの意味を表している。ただしそれは、xを含みうる潜在的に可能なすべてのコンテキスト上でのxの意味である。

右で、h^yは、同様に、テキストyが現れる可能なコンテキスト上でのyの意味を表している。

$$L$$

$$x \xrightarrow{a} y$$

$$\hat{L}$$

$$h^y \xrightarrow{a} h^x$$

xが出現可能なコンテキストは、yが出現可能なコンテキストの拡大である。

ある表現の具体的な継続は、その表現を利用することのできる、潜在的に可能なコンテキストを制限するのである。

copresheaf-意味論の射程

大規模言語モデルの数学的構造 II

ここまでで見てきたこと

ここまで見てきたこと、すなわち、Syntaxカテゴリーから Semantic カテゴリーの構成は、大規模言語モデルの次のような数学的モデルを構築できたということです。

「何の構造もないように見える具合的なテキスト・データから、その表現の継続の確率分布を大量に学習することによって、大規模言語モデルは内部に意味の世界を構築している。」

重要なのは、copresheaf-意味論

この数学的言語モデルには、重要なポイントがあります。

それは、言語の意味が住む世界を、[0,1]-copresheaf として、明確に定義したことです。

我々は、言語の文法構造については、いろいろ知ることができます。しかし、意味の世界がどのような構造をしているかについては、必ずしも明確な概念を持っているわけではありません。

しかし、大規模言語モデルが内部に構築する意味の世界を、[0,1]-copresheaf としてとらえることは、意味へのアプローチを大きく変えるものです。

copresheaf-意味論の含意

もし、意味が [0,1]-copresheaf で表現されるのなら、さまざまな意味の構成には、カテゴリー論的操作が対応しているということができます。

Part 2 で見た、論理的概念「red and blue」「red or blue」にカテゴリー論的な product と coproduct に対応づけたのは、その一例です。

もっともPart 2での議論は、productもcoproductも、未だenrich化されていませんでした。論理的概念「AND」と「OR」を[0,1]-copresheaf で表現するためには、productもcoproductも、enrich化されなければなりません。

今回のセミナーでは論文を最後まで紹介できませんでした。

実は、論文はその後半部分で､「OR]という概念の構成を可能とするcoproduct のenrich化､「AND]という概念の構成を可能とするproduct enrich化を行っています。

この方向でのこの論文の重要な成果は､「implies」という論理的含意の概念が､$\hat{L}$のinternal-hom $[h^x, h^y]$であることを示したことです。

含意 $x \Longrightarrow y$ (xならばy)は、

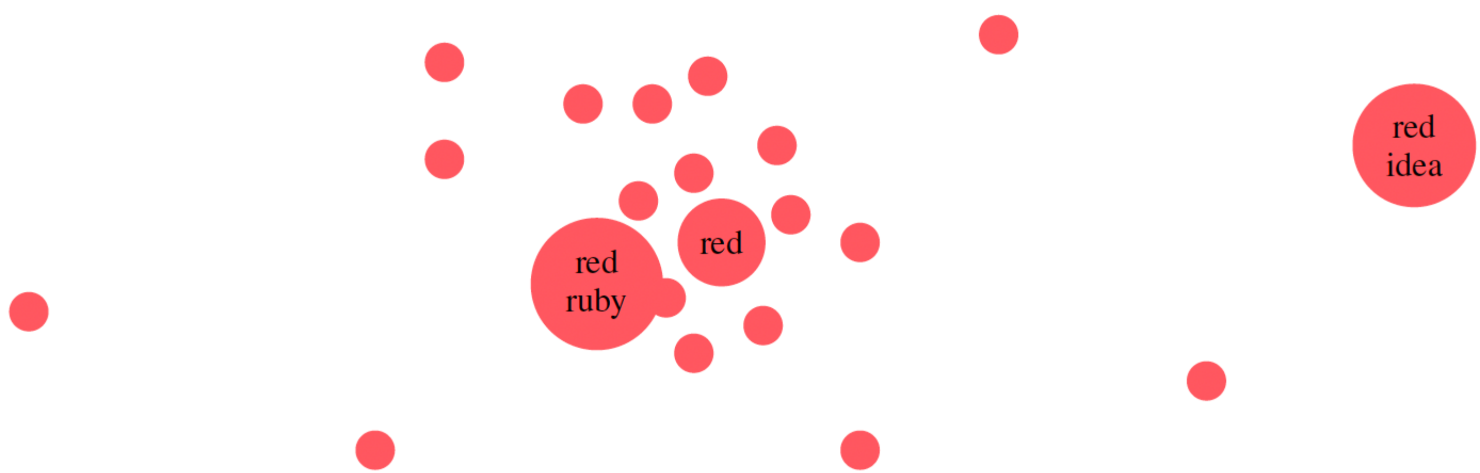
copresheaf $[h^x, h^y]: L \to [0,1]$と定義できる。

copresheaf-意味論のvariant
$[0,\infty]$-$copresheaf$

この論文は、単位区間[0,1]でenrich化するcopresheaf-意味論のvariantとして、$[0,\infty]$区間でenrich化するcopresheaf-意味論も紹介されています。

それは言語を一般的な計量空間と考えるものです。表現 redは、$[0,\infty]$-$copresheaf\ d_M(red,-)$ とみなされます。

redと継続しそうもない表現はredから離れた場所に、継続しそうな表現は、redの近くに置かれます。



copresheaf-意味論は「出発点」

意味の世界の構造を探る糸口は掴めたのですが、我々は、大規模言語モデルが学習した意味の世界でさえも、実は、よく理解できているわけではありません。

ただ、意味の情報は [0,1]-copreshef カテゴリーに蓄えられていることを理解し、この情報に カテゴリー論的操作がどのように作用するかを理解することは、近未来のさまざまの具体的で魅力的な応用に道を開きます。

論文の” 6.1. Applications and future directions”には、興味深い方向が示されています。

copresheaf-意味論はAI研究の新しい「出発点」なのです。