

ゆずこ 電車、久しぶりに乗るね～
唯 あー、そうかもなー
ゆずこ いつぶりくらいかな
縁 去年の海？
唯 いやいや、さすがにもっと乗ってるだろ
ゆずこ 電車、あんまり乗らんよね～
縁 あんまりね～
...
ゆずこ いろんな人がいるね
縁 うん
ゆずこ なんか、さ
縁 うん？
ゆずこ いろんな人がいるんだなーって。思うよね？
縁 あー、わかるー
唯 ...。
唯 あんまじろじろ見るなよ
縁 えー？
唯 ゆずこ、他の人のことじろじろ見てただろ？
ゆずこ なにそれ？
縁 ？
ゆずこ わたしだけを見てろってこと？
唯 いやそうじゃなくて...
縁 あ、席空いたよー
ゆずこ 座る？
唯 ...。
...
縁 かわいかったねー
ゆずこ かわいかった！
唯 そうだな
縁 でもすごいぐずってたねー
ゆずこ ぐずってた！
唯 (なんか、ゆずこみたいだった...)
ゆずこ ...？
唯 なんだよ？
縁 なんかねー

ゆずこ ん？
 縁 今の赤ちゃん、ゆずちゃんみたいだった
 ゆずこ ほへ？
 唯 あー、うん...
 ゆずこ えー？ なにそれー？ それ喜んでいいの？
 縁 なんかー、しゃべり方とか？
 唯 喋ってたか？
 縁 だーだーって
 ゆずこ それ、バカにされてる気がするんですが...

...

ゆずこ 喃語。乳児の発する言葉。言語を獲得する前段階
 唯 さっきの赤ちゃんだ
 縁 わんわんとか？
 ゆずこ んーと。幼児語。乳幼児の会話に用いられる言葉
 縁 ふーん？
 ゆずこ わんわんとかは幼児語っていうんだって
 ゆずこ 言語の獲得...
 唯 あ、これテレビで見たことある
 唯 ブーバ・キキ効果
 唯 2つの図形を見せて、
 唯 どっちがブーバで、どっちがキキかを答えさせるという心理実験
 縁 ブーバー？
 唯 うーんと、言葉に意味はなくて、音の響きとかから判断するんだって
 ゆずこ ぶ～ば～
 縁 なんかー、キキの方が痛そうな感じがする
 ゆずこ 黒板みたいな？
 縁 あー、確かに
 唯 ん？ どっちが黒板？

...

ゆずこ E. Mark Gold さん^[1]
 縁 誰？
 ゆずこ 幼児の言語獲得の定式化を試みた人、だってー
 ゆずこ 幼児は親が話す言葉だけから言語を学習する...
 縁 うーん、それでー？
 ゆずこ 正しい言葉だけから正しい言語を学習できる
 ゆずこ これを Learning from positive data、と言うそうです
 唯 それっ当たり前じゃない？ 何が難しいんだ？

縁 うーん、よくわからない
 ゆずこ 子供は学んだ言葉を使って新しく作文するかもしれない
 ゆずこ それが正しい言葉であるかどうかを、親の反応から学習する
 ゆずこ この場合は、「正しくない言葉」も学習できるチャンスがある
 縁 ふーん？
 ゆずこ あ、こんな例があるよ¹

Q. 次の A は自然数 $(1, 2, 3, \dots)$ の部分集合です. A はどんな集合ですか?

$$A = \{2, 4, 6, 8, \dots\}$$

唯 偶数、だろ？
 縁 うんうん

A. A は偶数全体.

$$A = \mathbb{N}_2 = \{2, 4, 6, 8, 10, 12, \dots\}$$

ゆずこ さんね～ん
 ゆずこ とは限りません

A. A は偶数または 13 の倍数からなる集合.

$$A = \mathbb{N}_2 \cup \mathbb{N}_{13} = \{2, 4, 6, 8, 10, 12, 13, 14, \dots\}$$

縁 えー、インチキだよー
 ゆずこ 最後まで聞かないほうが悪いのです
 唯 お？ 急になんだ？
 ゆずこ そうじゃなくってー
 縁 あー、「…」ね

¹参考文献 [3] の例を改変

ゆずこ そう！「...」なのです

唯 有限の情報だけから決めるのは無理、ってことか？

ゆずこ 正しくは positive data (または informant) ね

ゆずこ これは正しい言葉ですよ、っていう

縁 言葉？

ゆずこ あ、そう。これは言葉なのです

- 一般に (有限とは限らない) 集合で、学習したい対象のことを Concept という. 例えば一つの言語 (日本語とか英語とか) は Concept の一例. 例えば自然数の部分集合は Concept の一例.
- ある要素が学習したい Concept に属するという情報を positive data (informant) と呼ぶ. 例えば親が話す言葉は positive data. 例えば自然数の部分集合 A について $x \in A$ は一つの positive data.

唯 ってことは集合 A が言語ってことか...？

縁 自然数が言葉でー？

...

唯 でも、そんなの無理だろ？

ゆずこ ？

唯 有限個の情報、あ、informant か？ しかくれなかったんだから

唯 そこから、そんな、「または 13 の倍数」なんて分かるわけないだろ？

縁 うんうん

ゆずこ そこで、「極限同定」という発想が生まれるわけです

唯 ほーん？

次のような学習の仕方を「極限同定」と呼びます。

- ある Concept をこれから学ぼうとする学習者がいます。
- 学習者は一つの positive data (informant) を得ます。^a
- 学習者はそれまでに得た positive data (informant) から、一つ、考えつく Concept を推論します。
- 上の、positive data を受け取って一つの Concept を推論することを一つのステップとして、これを繰り返します

更に次のようなとき、極限同定が「成功した」と言います。

- あるステップで真の Concept を推論し、かつ、
- 以降のステップでは、常にその Concept を推論する。

^aこの情報の得方を正提示 (positive presentation) と言いますが、厳密にはもっと強い制約があります。それは Concept の任意の要素はいつかは必ず提示される、というものです。

縁 じゃあさっきの例で言うとなー

唯 うん

縁 2, 4, 6, 8, ... っるのが正提示でー

唯 うんうん

縁 数字一つを受け取るごとに、

唯 positive data な

縁 集合を答えるのが、Concept の推論、ってことー？

ゆずこ そうそう

ゆずこ ちなみに、推論した Concept の列を「推論列 (guessing sequence)」っていうよ

正提示 2, 4, 6, 8, ... に対して、推論列 $A_1, A_2, A_3, A_4, \dots$

2,	4,	6,	8,	...
↓	↓	↓	↓	
$A_1,$	$A_2,$	$A_3,$	$A_4,$...

ゆずこ あ、これは小さい順に数字を並べてるけど、別にそれに意味はなくて
 ゆずこ 順序はどうでもいいし、それに重複してもいいんだって
 縁 2, 10, 2, 10, ... みたいに？
 ゆずこ うん。でもこの注釈があって
 ゆずこ いつかは 13 の倍数も出現しないといけないわけ
 唯 時刻 t の推論を A_t って書いたの？
 ゆずこ そうっす
 唯 じゃあさっきの「極限同定の成功」ってのは

$$\lim_{t \rightarrow \infty} A_t = A$$

唯 って書いていいのか？
 ゆずこ うーん、いいんじゃないかな
 ゆずこ 文字通り形式化すると $\exists T, \forall t > T, A_t = A$ だけだね
 ゆずこ 「極限同定が成功」したとき、Concept を学習したって見做すらしい
 のさ

...

縁 それで？
 ゆずこ うん？
 縁 これで、人間の言葉が学習できるの？
 ゆずこ うーん、、、どうなんだろう？
 縁 えー
 唯 どうなんすか？ ゆずこさーん
 ゆずこ むーん。。。

...

ゆずこ あ！
 唯 なになに
 ゆずこ ”文脈自由文法の学習は不可能である” [1] ²
 縁 えーなんでー？
 唯 推論の具体的な方法も決めてないのに
 ゆずこ TABLE 1 に書いてるのだとー...
 唯 なになに？
 ゆずこ 文脈自由文法も文脈依存文法も正規言語も、正提示からの極限同定は
 不可能
 縁 だからなんでー？
 ゆずこ うーん、詳しい記述はどこにあるんだろう

²英語は文脈自由文法であると言われる。いや、少なくとも反証はされていない。

...

ゆずこ あ、あったあった

ゆずこ "super-finite class of languages"

縁 クラス？

ゆずこ うん、言語族、のことかな？

唯 そんなの今まで話に出てこなかったぞ

ゆずこ ごめんごめん。説明がめんどそうだったから...

ゆずこ 「自然数の部分集合」とか「自然言語」みたいに、推論する Concept を中から選べるための枠組みみたい

縁 え？ じゃあ、推論って、選択のこと？

唯 なんか急に問題が簡単に見えてきたなあ

列挙による推論.

Concept のクラスが添字つき集合^aの場合、列挙による推論がありえる.
この方法ではまず Concept を列に並べる.

$$\mathcal{C} = \{C_1, C_2, C_3, \dots\}$$

次のような方法で推論を行う.

- 添字 $i = 1$ を持っておく.
- positive data (informant) を受け取る.
- それまでに受け取った positive data と C_i とが矛盾するか調べる.
- 矛盾するなら、添字 i を 1 増やす. 矛盾しないなら何もしない.
- 推論として C_i を推論する.

^a集合の要素をもれなく列として並べられるということ

縁 さっきの自然数のは？ これで解ける？

唯 いやいや、無理だろ

縁 どうして？

唯 自然数の部分集合全体は可算じゃないから

縁 そっかー

ゆずこ もっと単純なやつじゃないと、これはダメみたいだね

縁 なんならいいんだろう

唯 「ある数の倍数」とかじゃないか？
 縁 それなら可算だー

自然数 n の倍数全体を \mathbb{Z}_n と書く. Concept class, $\mathcal{C} = \{\mathbb{Z}_n : n \in \mathbb{N}\}$ は列挙による方法で正提示から極限同定可能?

ゆずこ うーん...
 ゆずこ いや、おかしいよ！
 縁 えー、なんでー？
 ゆずこ だってほら、4 の倍数だとするじゃん？
 ゆずこ 4, 8, 16, ... ってきて、
 ゆずこ ほら、2 の倍数の可能性を捨てきれないわけじゃん
 唯 あっ
 ゆずこ でしょ？
 唯 添字の順序を工夫すれば... うーん... わからんなあ...
 縁 どゆこと？
 唯 ほら、 $\mathbb{Z}_1, \mathbb{Z}_2$ って順に見てくとするだろ？
 唯 そしたら、正提示はずっと2の倍数だから、添字が2で停まるわけよ
 縁 あーそっかー
 唯 \mathbb{Z}_4 を \mathbb{Z}_2 の前に持ってくればいいんだけど、そしたらキリがないし
 縁 あーでも、ちょっとズルをしたら上手くいくかも
 唯 どんな？
 縁 最初の positive data の数だけ先に見ちゃってー、
 縁 例えばそれが m だったら、

$$\mathbb{Z}_m, \mathbb{Z}_{m-1}, \mathbb{Z}_{m-2}, \dots, \mathbb{Z}_2, \mathbb{Z}_1$$

縁 って並べるの
 唯 あー、それなら上手くいくなー
 唯 上限を決めて、降りるように並べるわけだな
 ゆずこ ていうか
 ゆずこ これって、提示される数の最小値を取ればいいだけだね、これ³
 縁 あー、確かに
 唯 整数だから、絶対値は取らないとだけだな

³極小言語 (minimal language; MINL) 戦略の自然数バージョンです

自然数 n の倍数全体を \mathbb{Z}_n と書く. Concept class, $\mathcal{C} = \{\mathbb{Z}_n : n \in \mathbb{N}\}$ は次のような方法によって正提示から極限同定可能.

- 数 $m = \infty$ を持つ
- positive data (informant) m_t を受け取る
- m を m と $|m_t|$ の最小値とする ($m \leftarrow \min\{m, |m_t|\}$)^a
- Z_m を推論

^a $|\cdot|$ は絶対値

縁 でもなんか、

縁 言葉っぽくないね？

唯 あ、ていうか、文脈自由文法は？ ”super-finite” は？

ゆずこ あ、そうだったそうだった。忘れてた

...

唯 正提示から学習できる言語。パターン言語 [2]

ゆずこ お？

唯 ンンン?? $\Sigma A \dots$?

ゆずこ どれどれ

唯 んー。パターン言語の定義らしいんだけど...

ゆずこ ああ、たぶんこれ、Kleene 閉包だよ

唯 こんなヘンな記法が...

非形式的にパターン言語を説明します。パターン言語とはあるパターンによって「説明される」言語です。そしてここで言うパターンとは、要は、「空欄のある文」です。

- 例. “世界が でありますように”

また空欄には「名前」をつけることができます。

- 例. “ (x) が (x) を (y) てました”

そして、このパターンが「説明する」言語とは、空欄を自由に埋めて出来る文からなる集合のことです。ただし、同じ名前の空欄には同じものを埋めます。

- “私が私を見つめてました”
- “深淵が深淵を覗いてました”
- 等々

縁 あー、なんか、言語って感じ

唯 うーん、そうかー？

唯 同じ名前の空欄って、同じものが何度も出現するってことだろ？

唯 あんまり自然言語にそういうのって出てこないような

ゆずこ 同じ名前の空欄が出てこないものは、正則パターン言語、と言うそうです

縁 あ、これってもしかして、深さが 1 しかない文脈自由文法？

唯 あー、なるほど

唯 ん？

参考文献

本物語は次の参考文献をヒントに創作しました。

- [1] E. Mark Gold: “Language Identification in the Limit”, in *Information and Control* 10 (1967)

- [2] Angluin: “Positive Inference of Formal Languages from Positive Data”, in *Information and Control* 45 (1980)
- [3] Hiroki Arimura, Takeshi Shinohara and Setsuko Otsuki: “Finding Minimal Generalizations for Unions of Pattern Languages and Its Application to Inductive Inference from Positive Data”, in *In Proc. the 11th STACS, LNCS* 775 (1994)
- [4] @cymph: “言語の極限同定みたいな話”,
<http://cymph.cc/study/language-identification/history/it.pdf>