

ID:L-CH0043

データ収集日:2024年1月20日

聞いた素材:「データに隠れた規則性を探せ！～パターンが織りなす構造とAI～」(杉山麿人, 国立情報学研究所, 2021)

<https://youtu.be/i--NoqxSScA>

行番号	聞いた素材	協力者の発話内容		データ収集者の発話内容	備考
		発話	日本語訳		
1	みなさんこんにちは、国立情報学研究所の杉山麿人です。	大家好，我是国立情報学研究所的，呃，杉人，什么人，这个，这个字，我有点不知道该怎么念。	みなさんこんにちは、国立情報学研究所の、えー、杉人、なんとか人、これ、この字は、どう読むかちょっとわかりません。		協力者の発話の「杉人」は、漢字を中国語読みしたものと思われる。
2	今回は、データに隠れた規則性を探せというタイトルで、30分ほどお話しさせていただきます。	今天打算，呃，根据这个，题目，就是，数据中，被数据隐藏的规律性探索这个题目，进行一个三十分钟的，就是演讲吧。	今日は、えー、この、タイトルで、つまり、データの中に、データに隠れた規則性を探すというタイトルで、30分の、講演を行いたいと思います。		
3	この講演では質問とアンケートを募集しています。市民講座のウェブサイトかYouTubeの概要欄をご覧ください。	这个讲座，呃，正在募集一些问题和问卷调查。请在市民讲座的网页或者是，嗯，这个YouTube的概要栏，请看这个概，概要栏。	この講座は、えー、今質問とアンケートを募集しています。市民講座のウェブサイトまたは、うーん、このYouTubeの概要欄、この概，概要欄をご覧ください。		
4	お待ちしております。	呃，等您的参与吧。	えー、ご参加をお待ちしています。		
5	まずは、この購買データ解析の例を少しお話ししようと思います。	首先根据这个，购买的数据的解析的一些例子，嗯，想要，就是，简单的介绍一下。	まずこの、購買のデータの解析の例、うーん、簡単にご紹介したいと思います。		
6	スーパーマーケットがあったとして、まあ自分がたとえばスーパーマーケットのオーナーであるっていうことを想像してみてください。	请想象一下，比如说是超市，想象自己是超市的，嗯，什么。嗯。然后，我猜可能是顾客吧，但是我没有听清楚。	ちょっと想像してみてください。たとえばスーパーで、自分がスーパーの、うーん、なにかです。はい。そして、お客さんじゃないかと私は思いますが、はっきり聞けませんでした。		
7				前後の言葉から推測してくれたんですか。	
8		它上面有写是顾客，然后我就觉得是顾客，嗯。	上に書いてあるのが顧客なので、それで顧客だと思いました、はい。		
9	どういう商品が売れているのかっていうのをデータから解析しようということを考えます。	嗯，就是想要，嗯，就是通过买了什么东西，然后来，来进行解析吧。	うーん、なにを買ったかによって、解析をしようと思います。		
10				どうしてその解析をしようとするのかという目的はわかりますか。	
11		为什么要解析…通过顾客购买的东西，然后进行一个数据分析。	どうして解析をしようとするのか…顧客が購買した物によって、データ分析をします。		

12				どうしてそのデータが必要なのか、その理由はわかりますか。	
13		因为他想解析隐藏在数据后面的信息。	データの裏に隠れている情報を解析したいからです。		
14	たとえば、このスーパー、まあちよつと小さすぎますけど、商品が4つあって、	比如说这个超市里面，嗯，后面那个什么没有听清楚，然后但是有，呃，有四个商品。	たとえばこのスーパーの中に、うーん、後ろのほうのなにかは聞き取れなかったんですが、でも、えー、4つの商品があります。		
15				4つの商品があると聞いて、どうそのスーパーについてどう思いますか。	
16		可能超市规模比较小吧。	スーパーの規模は小さいほうかもしれません。		
17	パンと卵とリンゴとそれから牛乳、4つある。それで、	嗯，四个商品分别是，呃，面包，鸡蛋，苹果和牛奶。	うーん、4つの商品の分類は、えー、パン、卵、リンゴと牛乳です。		
18	たとえばお客さんが8人来て、どのお客さんがどの商品を買ったかってのが記録されています。	呃，比如说顾客有八个人，然后什么顾客购买了什么东西，然后这样记录。	えー、たとえば顧客が8人いて、そしてどんな顧客がどんな物を買ったのか、と記録します。		
19	たとえば、1人目のお客さんはパンと卵と牛乳を買って帰った。	比如说第一个顾客他购买了面包，鸡蛋和牛奶。	たとえば1人目の顧客はパンと卵と牛乳を買いました。		
20	2人目のお客さんはパンとリンゴですね、を買って帰った。	呃，第二个顾客购买了苹果，啊，面包和苹果。	えー、2人目の顧客はリンゴ、あ、パンとリンゴを買いました。		
21	そういうような情報がここにあります。	嗯，像这样的，呃，这里有像这样的信息。	うーん、このような、えー、ここにこのような情報があります。		
22	で、この中から、どういう商品売れてのかなってことを考えるわけですけど、	嗯，「ここから[日本語で発話する]」…像在这里…「ちよつとなんか[日本語で発話する]」，有，有点忘记刚刚他说的什么。	うーん，「ここから」…ここから…「ちよつとなんか」，ちよつと今言ったことを忘れました。		
23	最初は1つ1つの商品に注目して、それがどのぐらい売れたかっていうのを計算してみよう。	嗯，最初是一个一个商品，然后想要计算它们，是怎样被，怎样，被卖了多少。	うーん，最初は1つ1つの商品，それからそれらが，どのように，どのように，どのぐらい売れたのか計算します。		
24	これは、いわゆる商品の購入頻度を計算するってことに対応します。	这对应着商品的购入频度。	これは商品の購入頻度に対応します。		

25				最初、スーパーのお客さんと想像してくださいと聞いたと思いますが、ここまで聞いて、もし考えが変わったなら教えてください。	
26		嗯，确实，我觉得可能应该是店长。	うーん，確かに，店長じゃないかと思います。		
27				考えが変わった理由を教えてください。	
28		因为，他现在要分析的是商品的购入频度，如果是店长的话，呢，根据购入频度，就可以很好地看到哪一个商品最好卖。	彼は今商品の購入頻度を分析したいんですけど、もし店長だったら、えー、購入頻度によって、どの商品が一番よく売れるかよく見ることができるからです。		
29	たとえば、パンは8人中7人のお客さんが買って帰っているということで、	比如说面包，八个顾客中七个顾客都购买了。	たとえばパンは、8人の顧客の中の7人の顧客が買いました。		
30	8分の7で、0.875っていう頻度を計算することができます。	嗯，就计算出了八分之一，也就是零点八七五，的购入频率。	うーん，8分の1と計算されて、つまり0.875，の購入頻度です。		
31	同様にして、卵でしたら、3人のお客さんが買って帰っているので、8分の3で0.375っていう値を計算することができます。	呃，然后是鸡蛋，三个人购买了，所以是八分之三，嗯，所以购入频度是零点三七五。	えー，そして卵は、3人が買ったので、8分の3で、うーん，だから購入頻度は0.375です。		
32	っていう値を計算することができます。	嗯，然后这样，这样，可以这样计算。	はい，このように，このように，このように計算できます。		
33	まあこういうふうにしてどの商品がどのぐらいの頻度かっていうのは、簡単に計算できると思うんですけど、	嗯，像这样，就是，呃，什么商品，然后，然后购买了多少次，然后这样可以计算它的购入频率。	うーん，このように，えー，どんな商品を，何回買ったか，このようにその購入頻度を計算することができます。		
34	もう少し違うことを考えてみましょう。	嗯，接下来再想一些别的东西。	うーん，次はほかのことを考えます。		
35	商品の組み合わせを考えて、たとえば2つの商品が、同時に買われている頻度ってのが、どのくらいかってことを考えます。	嗯，比如说，嗯，商品的组合，嗯，两个商品一起购买的，呃，是怎样一起，是怎样一起购买的。	うーん，たとえば，うーん，商品の組み合わせ，うーん，2つの商品と一緒に買うのは，えー，どのように一緒に，どのように一緒に買うのか。		
36	たとえば、パンと牛乳っていうのを考えて、パンも牛乳も両方買って帰った人が何人いたかということを考えます。	嗯，比如说，呃，牛，呃，比如说面包和牛奶，嗯，就是一起购买面包和牛奶的人，大概有多少人。	うーん，たとえば，えー，牛，えー，たとえばパンと牛乳，うーん，パンと牛乳と一緒に買う人は，だいたい何人ぐらいいるのか。		
37	今の場合だと、1人目の人と、えー5番目の人、7番目の人、8番目の人	嗯，现在来说的话就是，嗯，第一个人，然后第五个人，第七个人和第八个人购买了 this 组合。	うーん，今だと，うーん，1人目の人，それから5人目の人，7人目の人と8人目の人がこの組み合わせを買いました。		

38	つていうふうに、4人が買って帰ってますので、8分の4で0.5つていうふうに計算することができます。	嗯，四个人购买了，所以是八分之四，嗯，购入频度就是零点五。	うーん、4人が買ったので、8分の4で、購入頻度は0.5です。		
39	同様にして、たとえば卵と牛乳、両方買った人、何人いますかっていうのを考えると、えー1人目の人と5番目の人と、8番目の人というふうにして、	嗯，接下来比如说，鸡蛋和牛奶一起购买的人有多少，那就是第一个顾客，第五个顾客和第八个顾客。	うーん、次はたとえば、卵と牛乳と一緒に買った人はどのぐらいいるのか、それは1人目の顧客、5人目の顧客と8人目の顧客です。		
40	ま、0.375という頻度を計算することができます。	然后就是，它的购入频度就是零点三七五。	つまり、その購入頻度は0.375です。		
41	で、おんなじように、商品の組み合わせを大きくしていくってことも可能で、3つ、たとえばパンと牛乳と卵の3つも買った人が何人いるかってことを	嗯，然后把这个组合可以扩大，比如说三个一起购买的，嗯，面包牛奶和鸡蛋一起购买的人有多少。	うーん、そしてこの組み合わせを拡大することもできて、たとえば3つ一緒に買った、うーん、パンと牛乳と卵と一緒に買った人はどのぐらいいるのか。		
42	計算することもできます。	也可以一起计算。	も一緒に計算できます。		
43	今の場合だと、1人目の人と5人目の人と8人目の人、というふうに3人いますので、ま、8分の3で0.375という頻度を計算することができます。	嗯，现在是顾客一，然后顾客五和顾客八都这样购买了，所以是八分之三，然后购入频度是零点三七五。	うーん、今は顧客1、それから顧客5と顧客8がこのように買ったので、8分の3で、購入頻度は0.375です。		
44	で、こういう商品の組み合わせのことを、パターンと呼んだりして、こういう組み合わせ的なパターンの中で、重要なパターンをデータから見つけてくる、こういった	嗯，这种组合叫做一种模式，然后后面模式，嗯，咋样，然后又给忘了。	うーん、こういう組み合わせを一種のパターンと呼んで、そしてそのあとパターン、うーん、どんなだったか、また忘れました。		
45	課題のことを、パターンのマイニングと呼んだりします。	模式的什么叫做课题，叫做什么课题。	パターンのなにかは課題と呼びます、なにかの課題と呼びます。		
46	今の場合だと、ま、たくさん売れている商品が知りたいので、たとえば頻度が0.5以上のものを見つけてきましよう、ってことを考えると、	嗯，想知，想知道，就是被，被购买最多的产品。	うーん、一番多く売れている商品を知りたい。		
47	パンと牛乳、それからパンと牛乳の組み合わせ、この3つ	嗯，就是最多购买的就是，嗯，那个，面包，牛奶和面包牛奶的组合。	うーん、一番多く売れているのは、うーん、あの、パン、牛乳とパンと牛乳の組み合わせです。		
48	が半分以上の人が買って帰っている、	嗯，就是超过一半的人购买了。	うーん、半分以上の人が買いました。		
49	頻出なパターンであるというふうに考えることができます。	嗯，就是一个什么，频度的模式，模式吗？	うーん、なにか、頻度のパターン、パターンでしょうか。		
50				今ここまで聞いてみて、たくさん売れてる商品のことを知りたいと言っていましたけど、どうして先生はたくさん売れてる商品の話をしたんでしょうか。	

51		因为，因为什么，因为他要总结嘛。	なにか、まとめたいからでしょうか。		
52	で、まあこれで、えっと頻出なパターンを、えっと全部見つけてくるってことができるんですけど、	嗯，像这样可以找到频度高的模式。	うーん、このように頻度の高いパターンを見つけることができます。		
53	一見すごい簡単なことをやっているように見えますと思います。	可以通过，很简单的，做很简单事情来，怎么样。	でも、とても簡単な、とても簡単なことをして、どうか。		
54	単に1つ1つの商品、もしくは商品の組みあわせっていうのを持ってきて、その頻度を計算して、でまあその頻度が高ければ出力すると、それをやっているだけなので、	什么找那些商品或者是商品的组合，然后然后把它，然后计算它们的，呃，频度。然后，然后把它们输出，然后后面有点长了，然后后面就记不太清了。	それらの商品か商品の組みあわせを見つけて、そしてそれを、そしてそれらの、えー、頻度を計算します。そして、そしてそれらを出して、それからそのあとはちょっと長くて、そのあとははっきり覚えていません。		
55				商品と商品の組みあわせの頻度で、先生はなにを説明したんでしょうか。	
56		他想说明整个这个计算是很简单的，就是说明一下这个过程是怎么样的。	この計算がとても簡単だということ、この過程がどんなかを説明したかったと思います。		
57	一見すごいシンプルなことをやっていると思います。だけど実は、	嗯。就是很简单的一个事情。	はい。とても簡単なことです。		
58	これは結構大変で、ま、多くの場合、こうやって単純に、1個1個のパターンを確認するだけでは、えっと全部頻出なものを見つけるってことは不可能である	嗯，然后又说是，还是很，很难的，很费劲的，因为要一个一个的去，去检查嘛。	うーん、そしてまた、やはりとても、とても難しい、とても大変です、1つ1つ検査しなければいけないので。		
59	ということが少し考えてみると見えてきます。たとえば商品の数20個ってことを考えましょう。	比如说，嗯，商品数一共是二十个。	たとえば、うーん、商品の数が全部で20個です。		
60	ま、20個しかないスーパー、まあ、まだまだ全然小さいと思うんですけど、	嗯，只有二十个商品的超市，嗯，虽然很小，	うーん、商品が20個だけのスーパーは、うーん、とても小さいですが、		
61	20個あるとしましょう。え、そうすると一人ひとりの、あ、1つ1つのパターンが、買うか買わないかの2通りがあるので、	嗯，一个一个商品，就是，呃，买还是不买，然后有一个单词没有听清楚，但是我记住它的意思，好像是「ミトリ」[日本語で発話する]]。	うーん、1つ1つの商品が、えー、買うか買わないか、それからある単語が聞き取れませんでした。が、その意味は覚えていますが、「ミトリ」だったと思います。		
62	ま、20個ってことはパターンの数としては、2を20回かけた値、ま、2の20乗ですね。	有二十个模式吗？然后就是人是二十个。	20個のパターンがありますか。そして人は20人です。		
63	だいたい100万ぐらいなんですけど、100万個ぐらいのパターンがあります。	「ひゃくご」[日本語で発話する]……一百与零，又有一百零五个模式了。	「ひゃくご」…100、また105個のパターンがあります。		

64				さきほどミトリ、という言葉がわからないと言ってたんですが、推測できたら教えてください。	
65		啊，是不是「ニトリー[日本語で発話する]」啊？那就是两个一组啊。	あ、「ニトリー」ではありませんか。2つ1組です。		
66				音から推測したんですか、それとも前後の文脈から推測したんですか。	
67		就是也有文章也有声音。	文章も音もです。		
68				2つの組みあわせというのは具体的にどういものかわかりますか。	
69		就好比说二十个里面有，有两个，其中两个商品是牛奶和鸡蛋，那么就是牛奶和鸡蛋的组合。	たとえば20個の中で、2つ、その中で2つの商品が牛乳と卵なら、牛乳と卵の組みあわせです。		
70	で、その100万個くらいのパターンに対して、それぞれの頻度を計算して、頻度が大きかったら出力するってことをやればいいんですけど、	有一百零五万的组合，然后那些组合就是计算它们的频度，也是可以，但是，	105万の組みあわせがあつて、それらの組みあわせの頻度を計算することもできるんですけど、		
71	だいたいどのぐらい時間かかるかってのをざっくり計算してみると、まあ1秒もかからないですね。このぐらいだったら、一瞬でコンピューター	呃，如果就是粗略的计算一下，需要的计算时间的话，大概就是这零点零零零五九秒，然后还是很快的，就是一瞬间的事情。	えー、もし大まかに計算したら、必要な計算時間は、だいたい0.00059秒で、やはりとても速くて、一瞬のことです。		
72				今言った計算はなにを計算するものですか。	
73		在计算，就是，计算这个模式的，频，频度数吗？频度，频度吗？	計算、このパターンの、頻、頻度数の計算ですか。頻度、頻度ですか。		
74				なんの頻度かわかりますか。	
75		就是刚刚说的组合出现的频度吧。	今言った組みあわせが出現する頻度ですよ。		
76				もう少し具体的に聞きたいんですけど、なぜこの計算方法になるのかとかはわかりますか。	

77		现在好像还不太知道。	今まだよくわかりません。		
78	使ったら計算できます。なので、まあ、まあまあ、大丈夫そうって感じなんですけど、	嗯，嗯，好像觉得没问题。	うん、うん、大丈夫そうです。		
79	今度、商品の数40個だったとしましょう。	接下来商品数变为四十。	次に商品の数を40に変えます。		
80	よん、ま、20個しか増えてないんですけど、40個のパターン全部チェックして、	嗯，仅仅增加了二十个，然后，四十个组合，嗯，这样看下来的话，	うーん、20個増やしただけで、そして、40個の組みあわせを、うーん、こうやって見ていくと、		
81	堅実なものを見つけてくるっていうのをやろうとすると、	頻度，頻率，頻，頻率，发现頻率…	頻度，頻度，頻，頻度，頻度を発見して…		
82	10分かかります。	大概要花费十分钟。	だいたい10分かかります。		
83				なにが10分かかるんですか。	
84		嗯，那个模式的，頻率吗？	うーん、そのパターンの、頻度ですか。		
85				さっきは1秒かからないと言ってましたが、それと違うのはなぜですか。	
86		因为之前二十个的话，它那个模式的数量是，嗯，二的二十次方，但是现在有四十个了，它的那个模式数就变成二的四十次方了，就变得非常的多，所以需要很多时间。	前の20個の場合、そのパターンの数は、うーん、2の20乗でしたが、今は40個になって、そのパターン数は2の40乗になりました。非常に多くなったので、多くの時間が必要です。		
87				その計算が時間がかかるということですか。	
88		嗯，对，我觉得是这个计算花费时间。	はい、そうです。この計算が時間がかかるのだと思います。		
89	さっきよりはだいぶ時間かかるんですけど、まあ10分くらいだったらまだ、10分待てば答えが出てくるということであれば、ま、そんな大したことないかなっていうふうに思います。	就是等待十分钟，嗯，觉得没什么大不了的。	10分待つのは、うーん、大したことないと思います。		

90	じゃあ今度50個ですね。また10個だけ商品の数を増やしてみました。	嗯，接下来是商品数变为五十个，然后相比之前增加了十个。	うーん、次は商品の数を50個にして、前と比べて10個増やしました。		
91	50個だと、1週間かかります。全部チェックするのに。	嗯，如果变成五十个的话，嗯，计算时间就变成了七天。	うーん、もし50個にしたら、うーん、計算時間は7日になります。		
92	1週間、まあ結構時間かかりますけど、まあ1週間待てば、自分の欲しかった答えが全部手に入るってことであれば、まあまあ1週間くらいコンピューターまわしっぱなしにして、まあできる	嗯，一周也还好啦，然后等待一周，可以得出自己的，想要的结果，也是可以的。然后通过电脑计算，一周，时间。	うーん、1週間もまあいいですね、1週間待てば、自分の、欲しい結果が得られるので、まあいいです。そしてコンピューターを通じて、1週間、計算します。		
93	って人は結構いるんじゃないかなっていうふうに思います。	然后有什么这样的人，前面。	そしてこのような人がいます、前の。		
94	だけど今度じゃあ70個にしてみましょう。パターンの数は2の70乗なんですけど、これ全部確認するのに、ま、だいたいざっくりですけど、1万9000年かかります。	然后接下来变成商品数为七十，然后，这个所需要的计算时间就变成了，一万九千年，就大概的计算一下。	そして次は商品の数を70にして、そして、これに必要な計算時間は、だいたいの計算で、1万9000年になります。		
95	1万9000年はまあもちろん全然無理で、	一万九千年就是完全不行的，时间太长了。	1万9000年は全然無理です、時間が長すぎます。		
96	あの生きてる間にはこの計算終わらないということで、	嗯，就是，就是，好像就是，意思就是死了以后，嗯，计算都还没有结束。	うーん、つまり、つまり、意味は、死んでも、うーん、まだ計算は終わっていません。		
97	もう70個になっただけで、全然計算できない。	如果是变成九，七十个的话就完全无法计算。	もし9、70個になったら全然計算できません。		
98				1万9000年というのはなにを計算した数ですか。	
99		应该是计算那个频度高的商品组合。	たぶん頻度の高い商品の組み合わせを計算したときです。		
100				7日ではなく1万9000年になったのはなぜですか。	
101		嗯，因为它的就是组合的数量很多。	うーん、その組み合わせの数が多いからです。		
102	だけどまあ最近の、もしかしたらスーパーコンピューターとか、コンピューター発達したら、今でも1000個の並列とかありますし、	最近超市的，嗯，电脑很比较发达，然后，后面那一句没有听懂。	最近スーパーマーケットの、うーん、コンピューターも発達して、後ろの一文は聞き取れませんでした。		

103	まあ1万並列ぐらいまでもしいけば、もしかしたら1年くらいで、この計算できるようになるかもしれません。	嗯，「ヘーレツ、ヘーレツ」[日本語で発話する]，不知道，啊，是什么，并列吗？然后如果需要一年的话，计算也是可以的。	うーん，「ヘーレツ、ヘーレツ」，わかりません。あ、なんか、並列ですか？それでもし1年あれば、計算もできます。		
104				なぜ1年で計算できると先生は言ったんですか。	
105		这里没有太听懂。	ここは聞き取れませんでした。		
106	まだ望みはあると。	嗯，还是有什么，希望的。	うーん，まだ，希望はあります。		
107	ところが、もうちょっとだけ商品の数増やして、100個っていう状況を考えますと、	嗯，然后再，嗯，再增加到一百，然后想一下这种情况吧。	うーん，そしてさらに，うーん，さらに100個まで増やして，こういう状況を考えてみましょう。		
108	これ2の100乗なので、全部確認するのに200億年かかっちゃいます。	然后它的那个组合数，就变成了二的一百的次方，然后全部计算的话需要的时间，是二百亿年。	そしてその組み合わせの数が，2の100乗になって，そして全部計算すると必要な時間は，200億年です。		
109	でまあ，だいたい宇宙の年齢137億年ぐらいとかって言われてるので、	然后宇宙的年齢大概是一百三十七亿年。	そして宇宙の年齢はだいたい137億年です。		
110	200億年もかかる計算というのはもう全然できるはずがなくて、	然后要花费200亿年的这种计算是完全不行的。	そして200億年かかるという計算は完全に無理です。		
111	コンピューターがちょっと高速になったらできるとかそういうレベルではもう全然ない。絶対できないっていうぐらい、	嗯，电脑的这种什么，反正是完全不可以的，这种计算是完全就是行不通的。	うーん，コンピューターのこのようななにか，とにかく完全にできない，このような計算は完全に無理なものです。		
112				コンピューターがなにをして全然できないと言っているか，もし推測できたら教えてください。	
113		即使电脑计算得很快，但是也不可能把这个计算完了，反正，嗯。	たとえコンピューターがとても速く計算しても，この計算を終わらせるのは不可能です，いずれにしても，はい。		
114				そう推測したのはなぜですか。音から推測したんですか，それとも前後の文脈からですか。	
115		文章吧。没有，声音，没听出来。	文章ですね。音は聞き取れませんでした。		

116	むちゃくちゃ大量にパターンがある。それでもまだ商品の数は100個しかない、という状況なので、	嗯，就是它这个，呃，组合数特别多，但是它的商品数其实只有一百。	うーん、この、えー、組みあわせの数はとても多いですが、商品の数は実は100だけです。		
117	とにかくやっている1つ1つのことは単純なんですけど、パターンの数が多すぎるので、それを計算するってことが、あの全然できないと、	嗯，它那个组合一个一个的都比较的简单，但是，嗯，它的组合数特别多，所以就是没有办法计算。	うーん、その組みあわせ1つ1つは簡単なんですけど、うーん、組みあわせがとても多いので、計算できません。		
118	まあそういう状況になってます。	嗯，就是这种情况。	うーん、こういう状況です。		
119	じゃあどうすればいいかっていうことなんですけど、	那么怎么做才好呢。	ではどうしたらいいのでしょうか。		
120	この今のパターンマイニングの場合ですと、頻度が持つ性質ってのをうまく使うことで、計算を効率化すると。	这句话，没有，没有怎么听懂呢，嗯，嗯，通过什么东西计算频率，然后，嗯…	この文は、聞き取れませんでした。うーん、うーん、なにかを通して頻度を計算して、そして、うーん…		
121	それによって、その200億年かかるような計算をすっ飛ばして、	嗯，跳过那个计算需要两百亿年的。	うーん、200年必要なその計算を飛ばします。		
122	あの、まあ同じ答えを見つけてくるってことが可能になります。	自己怎么样还是可能的。	自分がどうとすると可能です。		
123	これは先ほどの、まあ商品の組みあわせを、少し違う書きかたをしているんですけど、	这是刚刚的那个，嗯，商品的组合，以，以和刚才稍微不同的形式，展示。	これはさっきのあの、うーん、商品の組みあわせを、さっきと少し違う形式で、示しています。		
124	下から上に行くほどパターンのサイズが大きくなっていっているのがわかると思います。	嗯，从下往上，它那个，嗯，组合就是越来越大，这个是可以看出来。	うーん、下から上に、あの、うーん、組みあわせがだんだん大きくなっているのが、見て取れます。		
125	で、それに対応して頻度も書いてます。	嗯，和组合对应的这个頻度，也就是写出来。	うーん、組みあわせに対応する頻度も、書いてあります。		
126	で、この矢印とこの頻度の関係をじっと見てみると、頻度っていうのは、上に行けば行くほど、同じか、もしくは小さくなっているっていうことがわかっていただけたと思います。	像这样记载的，然后越，越往上的话，它那个頻度就越来越小。	このように記載されているように、上に行くほど、頻度はだんだん小さくなります。		
127	たとえば牛乳の頻度は0.625なんですけど、牛乳とパンになると0.5になって、牛乳とパンと卵になると0.375になっている。	比如说牛奶吧，然后单个牛奶是零点六二五，但是牛奶和面包的组合是零点五，然后牛奶面包鸡蛋的组合就是零点三七五。	たとえば牛乳ですね、牛乳だけだと0.625ですが、牛乳とパンの組みあわせは0.5、そして牛乳とパンと卵の組みあわせは0.375です。		
128	で、これは必ず成り立つ性質で、矢印をどこでもいので上にたどっていけばいくほど、必ず同じく小さくなっていくっていう性質が成りたってます。	嗯，这都是是一样的，反正就是从下往上的话，就会越来越小，那个頻度。	うーん、これは全部同じで、とにかく下から上は、だんだん小さくなっています、頻度が。		

129	これは一見不思議なんですけど、よくよく考えてみると当然で、	虽然看起来有点不可思议，但这是必然的。	少し不思議に見えますが、必然的です。		
130	この頻度っていうのは何人がその商品買ったかってことなので、たとえば牛乳をそもそも8人中5人しか買ってないってことなので、	嗯，就是几个人购买了这个产品，比如说，面，比如说面包，刚刚八个人中，有七个人都购买了这个产品。	うーん、何人の人がこの商品を買ったか、たとえば、パン、たとえばパンは、さっき8人中、7人がこの商品を買いました。		
131	牛乳もパンも買った人が何人いるかって考えたら、5人以上いるはずはないですね。そもそも牛乳5人しか買ってないから。	然后面包和牛乳一起购买的人，然后有多少呢？然后大概就是，呃，呃，原？他说一个什么，原来猜想有五个人，但是其实只有四个人。	それからパンと牛乳と一緒に買った人は、何人いるでしょうか。だいたい、えー、えー、もともと？なにか、もともと5人だと思いましたが、実は4人でした。		
132				もともと5人というのはどうやってわかったんですか。	
133		是不是因为我听错了。它那个牛奶，单个买牛奶的人是有五个人，但是买牛奶和面包的人的这个组合，是四个人，这样啊。	私が聞きまちがえたからではないでしょうか。牛乳、牛乳だけを買った人は5人ですが、牛乳とパンを買った人という組み合わせは、4人、ということです。		
134	だから、この頻度っていうのは、絶対8分の5以上にはならない、あ、8分の5より大きくはならないというふうになっています。	嗯，所以它，它那个牛奶和面包一起购买的人，不会超过八分之五。	うーん、だから、その牛乳とパンと一緒に買った人は、8分の5を超えることはありません。		
135	ということは、今、頻度が高いパターンが欲しいので、この下のほうから見れば、全部を見なくても、頻度の高いパターンが手に入るんじゃないかってのが基本的なアイデアです。	现在是想要就是频度高组合嘛，所以就是看下面的数据，其实就好，因为它是从下往上逐渐变小，所以看下面的数据就可以。	今は頻度が高い組み合わせが欲しいので、下のデータを見れば、実はいいんです。下から上に向かってだんだん小さくなるので、下のデータを見ればいいんです。		
136	ここで、まあ少し抽象的に、このすべてのパターンがなる空間というのを書いています。	这里就是写地稍微抽象一点的，组合和它的频度。	ここは少し抽象的に書いた、組み合わせとその頻度です。		
137	それぞれの商品を、まあABCDって書いてまして、	嗯，各个商品写作ABCD。	うーん、それぞれの商品がABCDと書いてあります。		
138	商品を買ったか買わなかったっていうのは、まあ1と0で、デジタルな感じで表現してます。	嗯，就是买还是不买商品由一和零表示。	うーん、商品を買ったか買わなかったかを1と0で表示しています。		
139	1っていうのは、その商品を買ったってことで、0っていうのは、その商品を買わなかったということです。	一是买商品，然后零是没有买。	1は商品を買った、そして0は買わなかった。		
140	たとえば1人目の人はパンと牛乳と卵を買ってたと思うんですけど、1、1、0、1という文字列に対応している。	比如说第一个人，他买了面包，鸡蛋和牛奶，所以他就由一一零一来表示。	たとえば1人目は、パンと卵と牛乳を買ったので、1、1、0、1と表示しています。		
141	で、この右側の図っていうのは、そのありうるパターンを全部書き並べていて、パターンが大きくなることに、こう線を、こう引つ張っていってると、まあなんかそんな感じですよ。	嗯，然后这个右边的图呢，就是表示它频率的图，然后下面也写了那些一和零，然后中间由线连着。但是我不知道它为什么用线连着。	うーん、そしてこの右側の図は、その頻度を表している図で、下には1と0が書いてあって、真ん中は線で繋がっています。でもなぜ線で繋がっているのか私はわかりません。		

142	で、あとは頻度を全部のパターンにそれぞれ書いてます。	然后接下来是它的频度，都每一个组合上面都写了它的频度。	そして次はその頻度で、すべての組みあわせの上にその頻度が書いてあります。		
143	で、たとえば、この状態で、頻度が、じゃあ8分の3以上のパターンを全部見つけるということをやりますよ。	比如说，我们要找频度大于八分之三的组合。	たとえば、頻度がだいたい8分の3の組みあわせを探します。		
144	そうすると、この今赤で色を付けたところが実はちょうど対応しているんですけど、	比如说现在，红色的，红色的部分就是对应着频度大于八分之三的。	たとえば今、赤の、赤の部分が頻度がだいたい8分の3のものに対応しています。		
145	これはまあざっと見ていただくとわかるように、この全体の空間の、この下のほうに固まっているというのがわかっていただけたと思います。	嗯，就是粗略的看一下，大概它就是占据这个表的下面的部分。	うーん、ざっと見て、だいたいそれはこの表の下の部分を占拠しています。		
146	頻度っていうのは上に行けば行くほどだんだん小さくなるので、頻度が高いものを見つけるってことは、この下のほうの空間を見つけてくるといことに対応している。	然后频度比较小的就在表的上方部分，所以想要找频度比较高的话，就在表的下方寻找就可以。	それから頻度が小さいものは表の上の部分にあるので、頻度が高いものを探したければ、表の下のほうを探せばいいのです。		
147	じゃあこれを探すっていうのはどうすればいいかっていうと、下から順番にチェックしていつて、あの、欲しいものを見つけないことをやればいい。	嗯，那么，如何找它这个频度比较高的部分呢。那就是从下面往上，然后依次来寻找即可。	うーん、では、この頻度が高い部分をどうやって探すのでしょうか。それは下から上に、順番に探していけばいい。		
148	たとえば、まず0, 0, 0, 0から出発して、最初1, 0, 0, 0パンを買った人ってところをチェックする。で、頻度は8分の7なので8分の3より大きいので、たとえば次こち行て、パンも卵も買った人って何人いるかというのをチェックする。	嗯，首先从出发点零零零开始出发，然后往上，然后买面包的人有多少呢，然后是八分之七，然后再接着再往，这个左边往上，然后买面包和鸡蛋的又是八分之三。	うーん、まず出発点の000から出発して、上に向かって、パンを買った人は何人いるか、8分の7、続けて、左の上に向かって、パンと卵を買ったのはまた8分の3です。		
149	そうすると8分の3なので、まだOKってことで、もう1こ上に行く。で、そうするとパンと卵とリンゴを買った人なんですけど、これは1人しかいないので、頻度で8分の3を超えてない。	嗯，然后接下来就是买面包，鸡蛋和苹果的人。然后能看出来它是只有一个人。然后是它，而且它是可能超过之前的八分之三的。	うーん、そして次はパンと卵とリンゴを買った人です。そしてそれは1人しかいないことが見てとれます。そして、しかもそれは前ので8分の3を超えることはありません。		
150	ということでこれはいらなくてことなんですけど、これなにを教えてください。	所以这是就是不需要的，然后我是想告诉大家什么呢，	だからこれは必要なくて、それから私がみなさんに何を教えたいかというと、		
151	もう1こ上ですね、この全部の商品を買った人っていうのは、この1こ手前の頻度よりも絶対に大きくはならないので、もう、この頻度っていうのは見ることなく8分の1以上、より大きくはないということが保証されるということになります。	因为，嗯，再往上看的话，它也就只有一个人购买了所有产品，然后这个频度，也是不可能超过上一个频度的，所以就没有看的必要。	うーん、さらに上に向かって見ると、1人だけ全部の商品を買った人がいて、この頻度は、やはり1つ前の頻度をを超えることはないので、見る必要がありません。		
152	なので、この1, 1, 1, 1というすべての商品を買った人がどのくらいいたかっていうのは計算をする必要すらないということがわかります。	所以要看的，就是所有产品都购买的人有多少呢，就没有计算的必要。	だから、全部の商品を買った人が何人いるか見るのは、計算の必要がありません。		
153	1こ手前がもう全然頻度が足りないんで、上はもう見る必要がない。	以下，就是频度也是完全的不够的，所以就是上面就是没有继承的必要。	以下，頻度も完全に足りないんで、上は続ける必要がありません。		
154	今の場合そういうふうにして順番にチェックしていくと、この3つのパターンに対して、あの、実際には計算をする必要すなくて、	嗯，就是如果这样一个一个地往上看的话，嗯，这三个，嗯，就是没有计，没有计算的必要，没有看的必要。	うーん、もしこのように1つ1つ上に向かって見ていった場合、うーん、この3つ、うーん、計、計算する必要がありません、見る必要がありません。		

155	見なくていいというのが出てきます。で、今の場合は、あの、商品の数が4つしかなかったので、	現在の話只有四个商品。	今の場合は商品が4つしかありません。		
156	まあ見なくていいパターンというのは3つしかない、あんまり得してないんですけど、	嗯，所以现在只有三个，就是不需要看的，不需要计算的数据。	うーん、だから今3つだけが、見る必要がない、計算する必要がないデータです。		
157	実際にたとえば100個とか、あの、なってくると、200億年かかるぐらいのこのむちゃくちゃ大量の空間ってのがあるわけで、	那，如果商品数变成了一百个的话，就是需要两百亿年的，计算时间的数据的话，数据就非常的多。	では、もし商品の数が100個になったら、つまり200億年の、計算時間が必要なデータだったら、データは非常に多いです。		
158	その大量の空間のほとんどすべてっていうのは見なくていいと。この下のほうだけをチェックして行って、自分が欲しい頻度の高いパターンだけを取りだしてくる。まあそういうことをすればいいというふうになります。	嗯，如果是一百个的话，数据就是非常的大量，然后在大量的空间中，只需要看下面的部分，嗯，就可以。	うーん、もし100個だったら、データは非常に大量で、大量の空間で、下の部分だけ見れば、うーん、いいんです。		
159	こういうふうになら順にチェックして行って、自分の好きなところで止めるっていうような方法のことをアプリアリっていうふうに呼んで、	嗯，像这种从下面然后往上看，然后进行，嗯，检查的，然后在自己想停的地方停下来的种方法叫做「アプロリ」[日本語で発話する]]。	うーん、このように下から上に向かって見て行って、検査を、うーん、進めて、自分が止めたいところで止めるという方法を「アプロリ」と呼びます。		
160	もともと1994年に提案された手法なんですけど、	嗯，是一九九四年，嗯，提案的，被提案的方法。	うーん、1994年に、うーん、提案の、提案された方法です。		
161	今でもこの同じ原理が使われています。	嗯，现在也用着同样的原理。	うーん、今も同じ原理を使っています。		
162	で、これを使うことによって、たとえば商品数が100のとき、もしなにも考えずに全部チェック、確認していたら、200億年かかっていたんですけど、	呃，用这种方法，虽然商品数变成一百了，需要计算的时间是两百亿年，但是用这种方法计算的话，	えー、この方法を使うと、商品の数が100になって、必要な計算時間が200億年になっても、この方法を使って計算した場合、		
163	アプリアリを使えば数秒、ときには1秒以下で同じものを計算することができます。	嗯，就是用几秒，或者是一秒就能吧，就能计算出来。	うーん、数秒、あるいは1秒でできます、計算ができます。		
164				どうして200億年かかっていた計算が1秒で済んでしまったんでしょうか。	
165		因为以前，那个计算是需要把所有的数据都计算出来，但是现在我们只需要看，呃，那个表，表中最底下的数据就可以了，计算量可能没有那么多，所以很快。	以前、計算はすべてのデータを計算する必要がありましたが、今は、えー、その表の、表の中の一番下のデータだけ見ればよくて、計算量がそれほど多くないので、速いです。		
166	で、得られる結果はまったく同じです。なので200億年かけて計算するか、ま、1秒でこのアプリアリで計算するかっていうのはもう全然違うわけですけど、出てくる結果はまったく一緒です。	就是计算出来的结果是完全一样的。虽然那个需要计算需要两百亿年，但这个只需要一秒。	計算した結果は完全に同じです、必要な計算が200億年必要だったんですけど、これは1秒しかありません。		
167	だからなにも損することなく、ただただ、ま、200億倍ぐらい高速にできると、そういうことになっています。	什么什么不需要，嗯，这句话没有太听懂。	なにになにが必要ない、うーん、この文は聞き取れませんでした。		

168	これはパターンが持つ性質っていうのをうまく使ったことで、あの、上に行けば	嗯，这个是什么，什么什么…什么什么模式什么，他可能是想说总结，但是我…他说太快了。呃，这种模式拥有性质吗？然后，好好地利用一下这种性质	うーん，これはなに，なにに…なにに…なにのパターンのなにか，まとめを言いたいのかもしれませんが，私は…話しかたが速すぎました。えー，このようなパターンが持つ性質ですか。そして，このような性質をうまく使う		
169				なぜまとめだとわかったんですか。	
170		因为他上面的例子已经说完了，就应该总结了吧。	上の例はもう話し終わったので，まとめるでしょう。		
171	行くほど，頻度が小さくなる，そういう性質によって，うまく解決できたということになります。	就是越往表的上面走，然后它的频度越小，然后好好利用这个性质就可以，解析出我们想要的的数据。	表の上に行けば行くほど，頻度が小さくなる，この性質をうまく使えば，私たちが欲しいデータを解析することができます。		
172	この，商品の数が少し増えただけで，確認しないといけないパターンが，むちゃくちゃ	就是稍微增加一些，呃，商品的数量，然后它的组合就增加得特别多，然后需要一个一个，一个一个检查，嗯。	商品の数が，えー，少し増えと，組み合わせが非常に多くなって，1つ1つ，1つ1つ検査しなければいけない，はい。		
173	増えてしまうっていうのは，爆発的に数が増えるってことで，組み合わせ爆発と呼んだりするんですけど，この組み合わせ爆発の，まあ面白さであったり，恐ろしさってものについては，	嗯，就是这种组合的，就是数量增长了，就，就被称为组合的爆发，然后，呃，接下来根据这种组合性的爆发的，呃，有趣性，或者说就是说是恐怖性	うーん，このような組み合わせの，数が増えと，すると，組み合わせの爆発と呼ばれて，それから，えー，次はこのような組み合わせの爆発による，えー，おもしろさ，あるいは怖さ		
174	ま，こちらの動画，すごい有名なもので，ご存じのかたも多いかもしれないですけど，あの，ぜひ，もしご興味があったらご覧いただけたいというふうに思っています。	嗯，接下来这个，嗯，这个是一个它根据，呃，有关那个有趣性，和恐惧性的一个讲解的视频。然后肯定知道这个视频的，嗯，人也是有的。然后如果大家有兴趣的话，一定要去看一下。	うーん，次はこの，うーん，これは，えー，そのおもしろさと，怖さに関する解説の動画です。そしてもちろんこの動画を知っている人，うーん，人もいるでしょう。そしてもしみなさんが興味があったら，是非見てください。		
175				どうして組み合わせ爆発と呼ばれているか，もう一度説明してもらえませんか。	
176		嗯，因为，嗯，只是稍微增加一下商品的数量，然后它的组合就会变得非常的多，所以非常的多，就是一种爆发吧，就是很多。	はい，うーん，商品の数が少し増えただけで，組み合わせは非常に多くなるので，非常に多いと，一種の爆発ですね，つまりとても多いということです。		
177	ここまでは，えーと商品の組み合わせってものについて，少し具体例を使って説明してきたんですけど，	嗯，到此为止，就是稍微介绍了一下这些组合，嗯，的这种模式吧。嗯。	うーん，ここまでは，これらの組み合わせの，うーん，このようなパターンを少し紹介しました。はい。		
178	パターンにはもちろん，いろんなものがあって，商品だけではなく，	嗯，这种组合当然不只有商品，还有别，可以有别的东西的组合。	うーん，このような組み合わせはもちろん商品だけではなくて，ほか，ほかのものの組み合わせもあり得ます。		
179	たとえばソーシャルネットワークの部分グラフっていうのは，あのまあ，友だちかどうかで線を引いていくと，その友人関係みたいのが，いわゆるグラフと呼ばれる構造で表現できるわけですけど，	嗯，比如说在网上，嗯，你自己和朋友，然后是这种朋友关系，然后我们中间，连了线，这样。	うん，たとえばインターネット上で，うーん，あなた自身と友だち，このような友人関係の，間を，線で繋ぎました。このように。		
180	その部分的な構造っていうのを見てくるっていうのもパターンを調べることに対応しますし，	嗯，就是上面圈起来的部分，然后作为一个组合也是可以进行搜索的吗？	うーん，上の丸で囲った部分は，1つの組み合わせとして検索することもできるんじゃないか。		

181				どうして例の1つとしてこのネットワークの話をしたんでしょうか。	
182		嗯，因为他也想介绍吧，举例子说，除了商品的组合还有别的组合。	うーん，例を挙げて，商品の組み合わせの他にも別の組み合わせがあることを紹介したかったからです。		
183	分子の部分構造，分子っていうのもグラフ的な構造を持っているので，その分子の部分的な構造を探してくる。	就分子の部分構造也是一样的。可以，嗯，就是画出，比如说画出的这一分子的部分构造，它也是可以，嗯，做一个组合进行检索。	分子の部分構造も同じです。うーん，グラフにした，たとえばグラフにしたこの1分子の部分構造も，うーん，1つの組み合わせとして検索することができます。		
184	まあそういったものもパターンの解析というふうに捉えることができます。	也可以作为一个组合，然后进行搜索。	1つの組み合わせとして，検索をすることができます。		
185	それからDNAみたいな，こういう文字列ですね，部分列，で，そういうところから部分列をとってくるっていうのも一種のパターンマイニングです。	嗯，DNA也是一样的。然后可以把DNA的这部分文字列，或者是叫部分列，然后把它做一个组合，然后检索。	うーん，DNAも同じです。DNAのこの部分の文字列，あるいは部分列と呼びますが，これを1つの組み合わせとして，検索することができます。		
186	まあ画像に対しても実はおんなじような考えかたをすることができます。	嗯，画，画像，画面也是可以，也是同样的。就是可以挑出一部分，然后进行检索。	うん，画，画像，画像もできます，同じです。一部分選んで，検索することができます。		
187	まあ，画像についてももう少し詳しく見てみますと，	嗯，嗯，画像，画，画像，这个东西我想更详细地说明一下。	うん，うーん，画像，画，画像，これはもう少し詳しく説明したいと思います。		
188	たとえばこれは4掛ける4の白黒画像，まあ今どきこんな小さい画像っていうのはないと思うんですけど，	比如说这是一个四乘四的黑白画像，然后他说什么比这更用小的画像，怎么，怎么样，没听懂。	たとえばこれは4掛ける4の白黒画像で，それからなにかこれより小さい画像をもっと使って，どう，どうとか言いましてけど，聞き取れませんでした。		
189				なにが小さいと思いますか。なにか推測できたら教えてください。	
190		嗯，没有想到。	うーん，想像できませんでした。		
191	まああの，こういう簡単な例を使ってみましょう。	想，想用这个比较简单的一个例子。	この簡単な例を使ってみたいと思います。		
192	で，白か黒かしかないの，こんな感じ，	嗯，只有黑或者是白。	うん，黒か白かしかありません。		
193	1個ずつの四角っていうのは，それぞれピクセルに対応してます。	嗯，他说一个一个，要什么什么，这个没有听清楚。	うーん，1つ1つ，なにになにする，これは聞き取れませんでした。		

194	で、この画像を、こういう文字列としても表現できますよっていう話なんですけど、	嗯，它这种画，画像可以通过，就是文字，文字来表示。	うーん、このような画，画像は，文字，文字で，表示することができます。		
195	結局このピクセルに、まあそれぞれたとえば番号を振ってあげて、今の場合4掛ける4なので1から16まで。	嗯，然后这种也可以通过数字，然后，一块一块来表示。比如说这个是四乘四的，那么就通过一到十六来表示。	うーん、それからこういうのは数字でも、ひとかたまりずつ表示することができます。たとえばこれは4掛ける4なので、1から16までで表示します。		
196	で、この1から16までを単に横に、白だったら0、黒だったら1っていうふうに書き並べたのがこの右側の表現です。	然后右边表示出来的是一到十六，是每一块，如果是白色的话就用零来表示，如果是黑色的话就用一来表示。	そして右側に表示されているのは1から16までで、それぞれの四角が、もし白だったら0で表示し、もし黒だったら1で表示します。		
197	ま、こういうふうにしてあげると、まあ画像って言っても結局、0と1の列、16個0と1が繋がった列として扱うことができます、	然后这种图像也就变成了一行文字。	そしてこのような画像も1列の文字になります。		
198	まあやっていることは、さっきの商品の組みあわせを調べてたときと、まあまったく同じっていうふうに考えることができます。	嗯，怎么样做呢？其实就和刚刚的商品是一样的。	うーん、どのようにするのでしょうか。実はさっきの商品と同じです。		
199	たとえば1のピクセルっていうのはバンだったとか、2のピクセルは卵みたいな、さっきは、さっきだったらスーパーマーケットの商品ですし、今だったらこのピクセルと、そういうふうに対応していて、	嗯，然后现，如果是刚才的话一就是面包嘛，但是现在的话，一就是它那个白色的格子。嗯，这样，对。但是我还是不知道它那个ピクセル[日本語で発話する]是什么东西。	うーん、それから、もしさっきのだったら1はバンでしたが、今は、1はその白の四角です。うーん、そういうことです、はい。でもやはり「ピクセル」がなにかわかりません。		
200				ピクセルは名詞とか動詞とか、どういうものかはなんとなくわかりますか。	
201		嗯，那个是不是，可能是那个格子叫「ピクセル[日本語で発話する]」啊。我没有好，嗯，没有学过这个词。	はい、もしかしたら、その四角が「ピクセル」というのではないでしようか。よく、うーん、この言葉は勉強したことはありません。		
202				じゃあ今のところだと16個あるということなんでしょうか。	
203		我现在猜可能是吧。	そうではないかと思います。		
204	たとえばこの領域が3つ黒ですとかっていうのは、結局この10、11、12の部分が、あの、まあ全部が黒、というふうにそういうパターンを考えると、いうことに対応してます。	比如说，嗯，十，十一，十二这个部分，然后就有对应的右边那个，展示的那样，嗯。	たとえば、うーん、10、11、12という部分は、右側のそれが、そのように示されているのが対応しています。はい。		
205	で、こういうふうに、一見その商品の組みあわせ、スーパーマーケットの商品を考えるってことと、まあ、こういう画像みたいなものを見ると、いうことは、一見全然違うことをやっているんですけど、表現の仕方を工夫して少し抽象化すると、	嗯，就是刚刚那个商品的组合和现在画像组合其实是完全不同的，完全不一样，做的事完全不一样的事情。	うーん、さっきのあの商品の組みあわせと今の画像の組みあわせは実は全然違うものです、全然違います、していることは全然違うことです。		
206	まったく同じアルゴリズムであったりとか、手法で、あの、そういうデータを解析できるようになる。	嗯，但是可以通过相同的手法来进行，来对它们进行数据解析。	うん、でも同じ手法で、それらのデータ分析を行うことができます。		

207				違うものに見えるのに、同じ手法で解析できるのはどうしてでしょうか。	
208		嗯，因为它们转化成就是数字，然后进行排列，然后进行分析。	うーん、それらを数字に変えて、並べて、分析を行うことができます。		
209	ま、そういうこともあります。で、こういうパターン解析にはもちろん、あの普通、商品の数とかすごい多い、たとえばスーパーだったら5000種類とか商品が	就是像这种数据解析的话，就是肯定是有很多组合的。	このようなデータ分析の場合、当然とても多くの組み合わせがあります。		
210	あったりと思うので、とにかくパターンが数えきれないほど無数にあるという、そこがデータ解析の難しいところですよ。	嗯，就是数据解析的最难的部分就在于，它的组合就是有很多，可能就是数不清，接近无数吧。	うん、データ分析の一番難しい部分は、組み合わせが多い、数えきれない、無数に近いことです。		
211	なので、有限個であるということはわかってる、だけど、全部を見るってことはできないという状況です。	嗯，虽然是有限的，但是数量特别多，所以不可能所有都看。	うーん、有限ではありますが、数がとても多くので、全部を見ることは不可能です。		
212	なので、パターンを相手にするデータ解析では、いかに全体を見ることなく、一部分だけで自分たちの欲しい答えを出せるかっていうところが、	嗯，所以，嗯，不需要全部看，只需要看自己想要得到的部分，想要调查的部分就可以了。	うーん、だから、うーん、全部見る必要はなくて、自分が欲しい部分、調査したい部分だけ見ればいいんです。		
213	このパターン解析の肝になってきます。	嗯，这里就是什么，组合数据分析的疑问吗？「疑問[日本語で発話する]」？	うーん、ここはなんでしょう、組み合わせデータ分析の疑問ですか。「疑問」？		
214				最後の言葉は疑問と推測してくれたと思うんですが、どうして疑問と推測してくれたんでしょうか。	
215		嗯，因为我没有听清楚，嗯，感觉他那个发音很像疑问，「疑問[日本語で発話する]」。	うーん、聞き取れなかったので、うーん、発音が疑問，「疑問」に似ている感じがしました。		
216	こういう例をもう1つご紹介したいと思うんですけど、	嗯，想再介绍一个像这样的例子。	うーん、またこのような例を紹介したいと思います。		
217	今度はパターンが出現する傾向を分析しようということを考えます。	嗯，这次想分析这个组合的出现的倾，倾向。	うーん、今度はこの組み合わせの出現の傾、傾向を分析したいと思います。		
218	今後どういうパターンがどのぐらい出てくるのか、たとえばこういう、今の画像でパターンを表してますけど、	嗯，就是，嗯，那些组合是，大概出现的频度是多少呢？就是像。	うーん、つまり、うーん、それらの組み合わせは、だいたい出現の頻度はどのぐらいなのでしょう、かな、のような。		
219	こういう画像が出てきやすいとか、こういう画像が出てにくいとか、	比如说这种图像比较难，比较容易出现呀，那种图像比较难出现呀，这种。	たとえばこのような画像は比較的、比較的出現しやすいとか、あのような画像は比較的出現しにくいとか、こういう。		

220	商品の組みあわせだったらこういう、えっと購入のされかたが今後出てくとか、ま、そういうことを考えたい、	商品的，商品的購入，然后是这样，这样比较多，还是怎么样，嗯。	商品の、商品の購入，はこういう、こういうのが多くて、またどうか、はい。		
221	すなわちパターンの出現確率というのを推定したいという問題になります。	嗯，然后可以推测出这些组合的出现的，出现的，频，频率。	うーん、そしてこれらの組みあわせが出現する、出現する、頻、頻度を推測することができます。		
222	たとえばこの画像っていうのは30%ぐらいの確率で出てきますよとか、この画像は、たぶん15%ぐらいの確率で出てきますよとか、	嗯，比如说这种图像它出现的频度是百分之三十，然后，中间那个是百分之十五，这样。	うん、たとえばこのような画像が出現する頻度は30%で、そして、真ん中は15%、ということです。		
223	もしそういうことがわかれば、まあ、パターンの傾向がわかるので、今後の解析に役立つ、	嗯，如果像这样的话，就可以知道它的，嗯，就可以知道它的出现频度了，嗯。	うん、このようにすれば、その、うん、その出現頻度を知ることができます、はい。		
224	ということなんですけど、残念ながら、こういう確率を計算するということは、基本的にはできません。	但是很可惜，像这样的计算，它的出现频度是不行的。	でも残念ながら、このような計算は、出現頻度はだめなんです。		
225	なんでかっていうとやっぱり、パターンの数がひたすら多すぎるからです。	嗯，嗯，原因还是因为它的组合数太多了。全部把它计算完是不可能的。	うん、うーん、理由はやはり組みあわせの数が多すぎるからです。全部計算しおえることは不可能です。		
226	確率の値っていうのは0から1で、全部を足したら1になるから意味があるんですね。	嗯，这种出现的频度的话是从零到一开始，然后全部加起来就是一嘛。	うーん、このような出現の頻度の場合は0から1で始まって、全部足すと1になります。		
227	だけどパターンの数が多すぎるので、全部足して1になるように値をそろえるっていうことができなくて、	嗯，它的组合很多，然后把它们全部加起来等于是有有点不可能的吗？	うーん、組みあわせが多くて、全部足すのはちょっと不可能に等しいのでしょうか。		
228	全部足すのは正規化と呼ばれているんですけど、ほんとだったら全部足してその数で割ればいいんですけど、それができない。	他好像说的意思是，如果全部加起来等于一的话是叫正规化嘛。但是全部加起来等于一又是是不可能的。	言っているのは、もし全部足して1になるなら正規化と呼ぶという意味のようです。でも全部足して1になるのはまた不可能です。		
229	なので残念ながら、確率の値を明示的に計算するということが不可能です。	嗯，所以就很可惜，嗯，像这种把它全部计算出来，然后正规化是不可能的。	うーん、だから残念ながら、うーん、このようなものを全部計算して、正規化するのは不可能です。		
230				正規化とはなにか、もう一度教えてもらえませんか。	
231		正规化我听他的意思好像是说，就是把它们全部计算出来，然后加起来等于一这么一个过程嘛。	正规化というのは聞いたところ、それらを全部計算して、足して1になるという過程のことのようです。		
232	じゃあどうするかっていうことなんですけど、その正規化されていない、確率のような値であれば計算できる。	嗯，那怎么解决呢？就是不要，不正规化，然后就是，普通地解析一下。	うーん、ではどうやって解決するのでしょうか。正規化しないで、そして、普通に解析します。		

233	で、この確率のような値っていうのを使って、うまいことパターンの解析をしていきましょう、ということを考えます。	嗯，然后是把它们组合起来，然后计算它们的，计算它们的出现频率。	うーん、それらを組みあわせにして、そしてそれらを計算する、それらの出現頻度を計算します。		
234	で、この確率のような値にもいろいろあるんですけど、今回はこのエネルギーというものを使います。	嗯，有很多频率，但是这次用「エネルギー[日本語で発話する]」这种东西，嗯。	うん、多くの頻度がありますが、今回は「エネルギー」というものを使います、はい。		
235	エネルギーの定義は、まあここに書いてるんですけど、とにかく確率と違うのは、マイナス無限大から無限大の値をとる。	嗯，虽然在这里有写「エネルギー[日本語で発話する]」的定义，但是「エネルギー[日本語で発話する]」是什么东西呢？它是，是负无穷到正无穷的值。	うーん、ここに「エネルギー」の定義が書いてありますが、「エネルギー」とはなんでしょう。それは、マイナス無限大から無限大の値です。		
236	ま、どんな値でもいい。たとえばこのパターンのエネルギーはマイナス0.21とか、このパターンのエネルギーはなんか1.2とか、ちょっとわからないですけど、	比如说这种，就是上面那个组合，它的「エネルギー[日本語で発話する]」是负零点二，然后中间的那个，它的组，组合的「エネルギー[日本語で発話する]」是一点二。	たとえばこういう、上の組みあわせは、「エネルギー」がマイナス0.21で、真ん中の、組み、組みあわせの「エネルギー」は1.2です。		
237	ま、そういうふうにエネルギーつてもものが計算される。	嗯，就像这样，然后，可以这种エ，「エネルギー[日本語で発話する]」可以被计算出来。	うーん、このように、このような「エ、エネルギー」が計算されます。		
238	で、確率と違うのは、まあちょっとややこしいんですけど向きが反対で、	嗯，就是和，和那个，确率，就是出现频率不同。嗯，「エネルギー[日本語で発話する]」有一点，稍微有点复杂。然后后面那句话没有听懂。	うーん、あの、確率，出現頻度とは違います。うーん，「エネルギー」はちょっと，少しちょっと複雑です。それからそのあとの文は聞き取れませんでした。		
239	小さいエネルギーのほうが、高い確率を意味しています。	大概可能是它底下那句话，就是小的「エネルギー[日本語で発話する]」意味着更高的出现频率吧。	たぶん一番下のその文で、小さい「エネルギー」はより高い出現頻度を意味しているですね。		
240				いきなりエネルギーの話になりましたが、どうして先生はエネルギーの話をしたのかわかりますか。	
241		嗯，他说，之前不是那个，没有办法用正规化来计算嘛。所以他这次提出了一个解决方法，如何就是不用正规化计算，然后利用「エネルギー[日本語で発話する]」来进行计算。	うーん、その前に、正規化を使って計算することができないと言いました。だから今度は、どうやって正規化を使わないで計算するか、「エネルギー」を使って計算するか、解決方法を示しました。		
242				なにを解決する方法でしたか。	
243		解决那个计算组合的出现频率吧。	組みあわせの出現頻度を計算することの解決ですよ。		
244				先生はエネルギーはどんなものだと説明していましたか。	
245		他说是，嗯，就是和那个出现频率相，好像是相反的一个东西吧。然后它，越小的「エネルギー[日本語で発話する]」可以代表更高的出现频率这样。	先生が言っていたのは、うーん、出現頻度と反対、反対のもののようなですね。そしてそれは、より小さい「エネルギー」ほどより高い出現頻度を表すというものです。		