

ユーザー行動の平均値化

$$c_1 + c_1 + c_1 + \dots + c_1 = \sum_{k=1}^N c_k$$

$$\frac{1}{N} \sum_{k=1}^N c_k$$

click回数 1回



0回



2回



.....

1回



検索回数 1回目

2回目

3回目

N回目

ユーザー	1回目	2回目	3回目	4回目	検索回数合計 n_k	click数合計 c_k
A	1	0	検索なし	検索なし	2	1
B	3	1	2	1	4	7
C	2	4	2	3	4	11
D	1	2	1	検索なし	3	4

$$\text{平均検索回数} = \frac{1}{4} \sum_{k=1}^4 n_k = \frac{1}{4} (n_1 + n_2 + n_3 + n_4) = \frac{1}{4} (2 + 4 + 4 + 3) = 3.25$$

$$\text{平均click回数} = \frac{1}{4} \sum_{k=1}^4 c_k = \frac{1}{4} (c_1 + c_2 + c_3 + c_4) = \frac{1}{4} (1 + 7 + 11 + 4) = 5.75$$

北海道 旅行

検索語 北海道 旅行

ベクトル化

JTB
<https://www.jtb.co.jp> > 国内旅行 > 国内ツアー

北海道旅行・北海道ツアー | おすすめプラン・バック【JTB】

北海道旅行・ツアー・バックならJTBで。美味しいグルメが堪能できる北海道！初夏は富良野でラベンダー、冬は温泉宿で雪見温泉、北海道ならではの四季の楽しみ方が魅力です。

湯の川温泉(北海道)のホテル...・東京都発→札幌駅・大通 (飛行機...・札幌・函館

阪急交通社
<https://www.hankyu-travel.com> > 国内旅行

北海道旅行・ツアー おすすめプランやバックを紹介 - 阪急交通社

大自然に囲まれた北海道。白銀に覆われた季節はもちろん、新緑が輝く夏もおすすめ。夜を彩る函館の夜景や、季節によって色が変わる富良野の花畑、ゆったりと運河が流れる ...

手法の分類	具体的な処理の例																																										
クレイジング	ノイズ除去(クレイジング) JavaScriptのコード、HTMLのタグ																																										
単語分割	最小単位の単語に分割 動詞、名詞、形容詞など																																										
正規化	猫とネコ Webとweb など表記や大文字・小文字の統一、データの一貫性																																										
単語のベクトル化	<p>One-hot Encoding の他、様々な手法あり</p> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>名義尺度をダミー変数に変換した場合</p> <table border="1" style="border-collapse: collapse; text-align: center;"> <thead> <tr style="background-color: #4CAF50; color: white;"> <th>血液型</th> <th>A型かどうか?</th> <th>B型かどうか?</th> <th>O型かどうか?</th> <th>AB型かどうか?</th> </tr> </thead> <tbody> <tr><td>A型</td><td>1</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>B型</td><td>0</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>O型</td><td>0</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>AB型</td><td>0</td><td>0</td><td>0</td><td>1</td></tr> </tbody> </table> </div> <div style="text-align: center;"> <p>順序尺度をダミー変数に変換した場合</p> <table border="1" style="border-collapse: collapse; text-align: center;"> <thead> <tr style="background-color: #4CAF50; color: white;"> <th>震度</th> <th>震度1かどうか?</th> <th>震度2かどうか?</th> <th>震度3かどうか?</th> </tr> </thead> <tbody> <tr><td>震度 1</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>震度 2</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>震度 3</td><td>0</td><td>0</td><td>1</td></tr> </tbody> </table> </div> </div>		血液型	A型かどうか?	B型かどうか?	O型かどうか?	AB型かどうか?	A型	1	0	0	0	B型	0	1	0	0	O型	0	0	1	0	AB型	0	0	0	1	震度	震度1かどうか?	震度2かどうか?	震度3かどうか?	震度 1	1	0	0	震度 2	0	1	0	震度 3	0	0	1
血液型	A型かどうか?	B型かどうか?	O型かどうか?	AB型かどうか?																																							
A型	1	0	0	0																																							
B型	0	1	0	0																																							
O型	0	0	1	0																																							
AB型	0	0	0	1																																							
震度	震度1かどうか?	震度2かどうか?	震度3かどうか?																																								
震度 1	1	0	0																																								
震度 2	0	1	0																																								
震度 3	0	0	1																																								