



# Project idea

*This manuscript ([permalink](#)) was automatically generated from [lubianat/idea\\_hub@30c0fa8](#) on November 4, 2020.*

## Authors

---

- **Tiago Lubiana**

 [0000-0003-2473-2313](#) ·  [lubianat](#)

Computational Systems Biology Laboratory, University of São Paulo, Brazil

## Abstract

---

### Testing Manubot citation

[1]

## Reviews

### A review of cell-type annotation tools/methods

---

- Open collaborative review.
- Questions: what has been proposed so far? Which packages/modules are available?

### A review of single-cell analysis frameworks

---

- Question: What are the *de facto* standards for each step?
- Python / R / Julia / Other
- Seurat / Bioconductor SingleCellExperiment?
- Search on EuropePMC “single-cell RNA sequencing” AND (FIRST\_PDATE:2019) AND (((SRC:MED OR SRC:PMC OR SRC:AGR OR SRC:CBA) NOT (PUB\_TYPE:“Review”)))
- Review top 100.

Class:

- Method
- Primary analysis
- Reanalysis

Use the [Systematic Review Facility](#).

## Considerations based on information theory and human communication

(Written in Portuguese)

Modelagem de frases ambíguas

$P_a$  = probabilidades da percepção do ouvinte na perspectiva do falante  $P_b$  = probabilidades a percepção do ouvinte de fato

## Duas opções no modelo: comentário positivo e comentário negativo

$$P(X_{\text{intent}} = \text{positive\_comment}) + P(X_{\text{intent}} = \text{negative\_comment}) = 1$$

### 3 elementos: a intenção, a fala e a conclusão do ouvinte

$X_{\text{intent}}$  = intenção do falante  $X_{\text{spoken}}$  = as palavras usadas ("a fala")  $Y_{\text{guess\_about\_X\_intent}}$  = a conclusão do ouvinte sobre a intenção do falante

$Y_{\text{guess\_about\_X\_intent}} \mid X_{\text{spoken}}$  = a conclusão do ouvinte sobre a intenção do falante dada a fala

### A resolução de intenção por Bayes

O ouvinte conclui baseado em preconcepções a partir da fala. Por Bayes (exemplificando):

$$P(Y_{\text{guess\_about\_X\_intent}} = \text{"positive\_comment"} \mid X_{\text{spoken}} = \text{"example"}) =$$

$$P(X_{\text{spoken}} = \text{"example"}) * P(X_{\text{spoken}} = \text{"example"} \mid Y_{\text{guess\_about\_X\_intent}} = \text{"positive\_comment"})$$

### Exemplificando $X_{\text{spoken}}$ para esclarecer

$P(X_{\text{spoken}} = \text{"example"})$  = Probabilidade de você falar "example" exatamente daquele jeito, com aquela escolha de palavras. Dada todas as possibilidades de fala, qual é a probabilidade de você falar exatamente aquilo?

Bem, isso é, rigorosamente, não-modelável, mas podemos estimar com modelos linguísticos probabilísticos.

Para os fins aqui, vamos supor que há apenas 4 frases possíveis, todas igualmente prováveis *a priori*.

$$P(X_{\text{spoken}} == 1) = 0.25 \quad P(X_{\text{spoken}} == 2) = 0.25 \quad P(X_{\text{spoken}} == 3) = 0.25 \quad P(X_{\text{spoken}} == 4) = 0.25$$

### Exemplificando $P(X_{\text{spoken}} = \text{"example"} \mid X_{\text{intent}} = \text{"positive\_comment"})$

$P(X_{\text{spoken}} = \text{"example"} \mid X_{\text{intent}} = \text{"positive\_comment"})$  é a probabilidade de você falar "example" quando sua intenção é positiva.

Para as 10 frases possíveis, digamos que metade sejam consideradas por você mais positivas e metade mais negativas. Mas você nem sempre usa a mesma frase quando que passar uma mensagem positiva. Vamos supor que:

Dado a intenção positiva, , por exemplo:

$$P(X_{\text{spoken}} == 1 \mid X_{\text{intent}} = \text{"positive\_comment"}) = 0.5 \quad P(X_{\text{spoken}} == 2 \mid X_{\text{intent}} = \text{"positive\_comment"}) = 0.3 \quad P(X_{\text{spoken}} == 3 \mid X_{\text{intent}} = \text{"positive\_comment"}) = 0.2 \quad P(X_{\text{spoken}} == 4 \mid X_{\text{intent}} = \text{"positive\_comment"}) = 0.0$$

Dada uma intenção negativa, , por exemplo:

$P(X\_spoken == 1 \mid X\_intent = \text{"negative\_comment"}) = 0.0$   $P(X\_spoken == 2 \mid X\_intent = \text{"negative\_comment"}) = 0.2$   $P(X\_spoken == 3 \mid X\_intent = \text{"negative\_comment"}) = 0.3$   $P(X\_spoken == 4 \mid X\_intent = \text{"negative\_comment"}) = 0.5$

## Exemplificando $Y\_guess\_about\_X\_intent$ para esclarecer

Agora, o ouvinte não sabe o que  $X\_intent$  é, de fato. O que ele tem é  $Y\_guess\_about\_X\_intent$ , que é modificada pela fala. Há uma concepção, e essa concepção é atualizada. E é isso que o falante tem que pensar: no que a frase significa para o ouvinte, e não o que significa para ele.

Esse chute tem um valor a priori e um a posteriori ( $Y\_guess\_about\_X\_intent \mid X\_spoken$ ).

## Exemplificando $P(X\_spoken = \text{"example"} \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"})$

O ouvinte, então, tem seus modelos mentais de uso de linguagem.

### Um cenário

Se o ouvinte tem a mesma percepção de uso de vocabulário do falante, então as probabilidades de repetem:

$P(X\_spoken == 1 \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"}) = 0.5$   $P(X\_spoken == 2 \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"}) = 0.3$   $P(X\_spoken == 3 \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"}) = 0.2$   $P(X\_spoken == 4 \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"}) = 0.0$

$P(X\_spoken == 1 \mid Y\_guess\_about\_X\_intent = \text{"negative\_comment"}) = 0.0$   $P(X\_spoken == 2 \mid Y\_guess\_about\_X\_intent = \text{"negative\_comment"}) = 0.2$   $P(X\_spoken == 3 \mid Y\_guess\_about\_X\_intent = \text{"negative\_comment"}) = 0.3$   $P(X\_spoken == 4 \mid Y\_guess\_about\_X\_intent = \text{"negative\_comment"}) = 0.5$

### Outro cenário

Mas é possível que, por exemplo, na cultura do ouvinte, a frase 1 tenha uma palavra muito preconceituosa. O falante não sabe disso, mas isso pode afetar o ouvinte:

$P(X\_spoken == 1 \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"}) = 0.0$   $P(X\_spoken == 2 \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"}) = 0.6$   $P(X\_spoken == 3 \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"}) = 0.4$   $P(X\_spoken == 4 \mid Y\_guess\_about\_X\_intent = \text{"positive\_comment"}) = 0.0$

$P(X\_spoken == 1 \mid Y\_guess\_about\_X\_intent = \text{"negative\_comment"}) = 0.4$   $P(X\_spoken == 2 \mid Y\_guess\_about\_X\_intent = \text{"negative\_comment"}) = 0.1$   $P(X\_spoken == 3 \mid Y\_guess\_about\_X\_intent = \text{"negative\_comment"}) = 0.2$   $P(X\_spoken == 4 \mid Y\_guess\_about\_X\_intent = \text{"negative\_comment"}) = 0.3$

## O funcionamento e os ruídos do diálogo

---

Além das probabilidades de fato, a percepção de probabilidades pelo falante tem um papel importante.

A tarefa objetiva do falante é maximizar a  $P(Y\_guess\_about\_X\_intent = X\_intent)$ .

O falante sabe que a informação que o ouvinte tem é o a fala ("X\_spoken").

Logo,, na concepção do falante (Pfala) a probabilidade do outro entender aquilo que ele está falando:

$$P_{fala}(Y\_guess\_about\_X\_intent|X\_spoken) = P_{fala}(X\_spoken) \times$$

$$P_{fala}(X\_spoken|Y\_guess\_about\_X\_intent) / P_{fala}(Y\_guess\_about\_X\_intent)$$

Os diferentes P\_fala podem divergir dos P\_real, e isso causa conflito de diálogo.

## Implicações dos fatores latentes

### Divergência sobre percepção de intenção do falante afeta o diálogo

---

Se  $P_{fala}(Y\_guess\_about\_X\_intent)$  e  $P_{real}(Y\_guess\_about\_X\_intent)$  são muito diferentes, a comunicação fica complicada.

### Divergência sobre a predisposição do ouvinte entender que alguém com intenção "Y" falaria algo da forma "X"

---

$P_{fala}(X\_spoken/Y\_guess\_about\_X\_intent)$  e  $P_{real}(X\_spoken/Y\_guess\_about\_X\_intent)$  podem ser diferentes. Uma pessoa 'grossa' pode usar um conjunto de palavras e achar que está sendo positivo, mas para o interlocutor pode ser negativo.

### Divergência sobre o uso geral de frases

---

O último fator que entra no cálculo é o  $P(X\_spoken)$ . O  $P_{real}(X\_spoken)$  é a distribuição de frases para o ouvinte. Se o vocabulário é o mesmo, é provável que isso seja parecido para os dois.

O uso de gírias, por exemplo, pode complicar nesse caso. Se a outra pessoa não conhece uma gíria, a probabilidade de uso dessa frase no geral é muito baixa.

Para fins práticos, contudo, considerarei  $P_{fala}(X\_spoken) = P_{real}(X\_spoken)$ , mudando apenas as probabilidades condicionais.

## Typemaker

Given a Seurat object, a binarized dataframe and a maximum number of genes per type,

find the best combination of markers (using AND, OR, NOR, XOR) that better captures the clusters

## TechnotypeR

A simple annotation package for single-cell RNA-seq data based on the technotype framework.

Reference cell types based on markers.

## *sensu stricto* cell types

- ID: QXXX
- Markers:
  - Expresses X OR Y
    - A1
    - A2
  - Expresses X AND Y -B -C
  - NOT (Expresses X OR Y) -D1 -D2
  - NOT (Expresses X AND Y) -F -G

## Example

ID: Q001 label: "human B cell"

(at least 1) - Expresses X OR Y - MS4A1 - CD79A - CD79B (all) - Expresses X AND Y

(not express at least 1) - NOT (Expresses X OR Y)

(not express all)

- NOT (Expresses X AND Y) - CD3E - CD4 - CD8A - NKG7

## Steps:

- User decides what counts as expressed --> binarization algorithm
- Apply mask to dataset
- Cells are labeled with multiple labels.
  - For each mask, test all cells.
    - If a mask fits, label as technotype (e.g, using dataset\_ID\_Q001)
    - Relate technotype as a subclass of the uper class (e.g Q001)

## References

---

**1. Viral infections acquired indoors through airborne, droplet or contact transmission**

Giuseppina La Rosa, Marta Fratini, Simonetta Della Libera, Marcello Iaconelli, Michele Muscillo  
*Annali dell'Istituto Superiore di Sanità* (2013-01-01) <https://www.wikidata.org/wiki/Q38114797>

DOI: [10.4415/ann\\_13\\_02\\_03](https://doi.org/10.4415/ann_13_02_03)