

<https://brasileiraspln.ufscar.br/livro-pln-4ed-vol1/dados/cap-transdutivo/cap-transdutivo.html>



3º.Ciclo de
Encontros:
4 x PLN

Aprendizado Transdutivo em PLN
Capítulo 17 do Vol 1 do Livro de PLN do
BPLN (4a. Ed.).

Nataly Leopoldina Patti da Silva
Norton Trevisan Roman

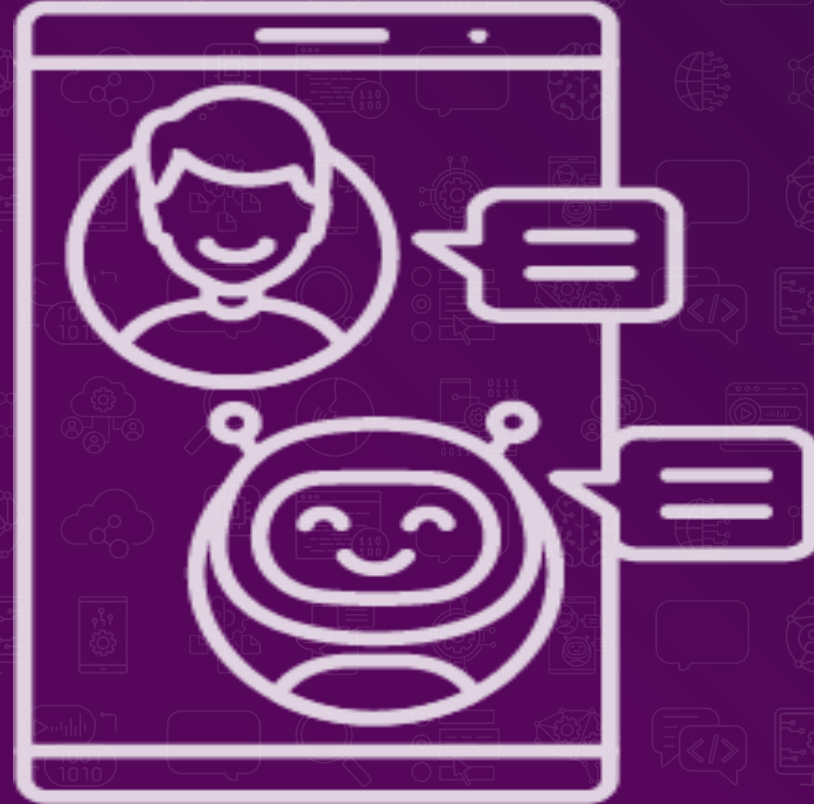
Sumário

- ❑ **Problema** | Poucos Rótulos em PLN
- ❑ **Conceito** | Da indução à transdução
- ❑ **Grafos** | Métodos baseados em propagação de rótulos
- ❑ **LLM** | Ajuste Fino Transdutivo
- ❑ **Conclusão** | Avaliação, benefícios e limitações



1. O problema

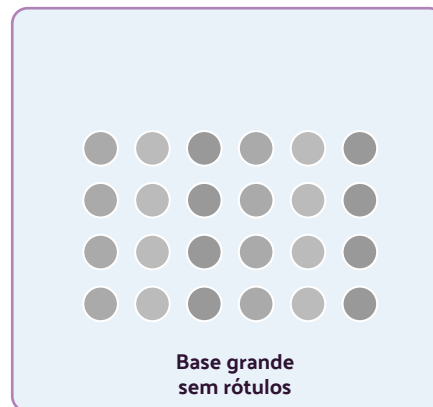
Por que aprendizado transdutivo aparece em PLN?



O ponto de partida: seu corpus raramente vem “perfeito”



Você tem rótulos, mas poucos
exemplos.



Você tem textos, mas não tem
anotação.



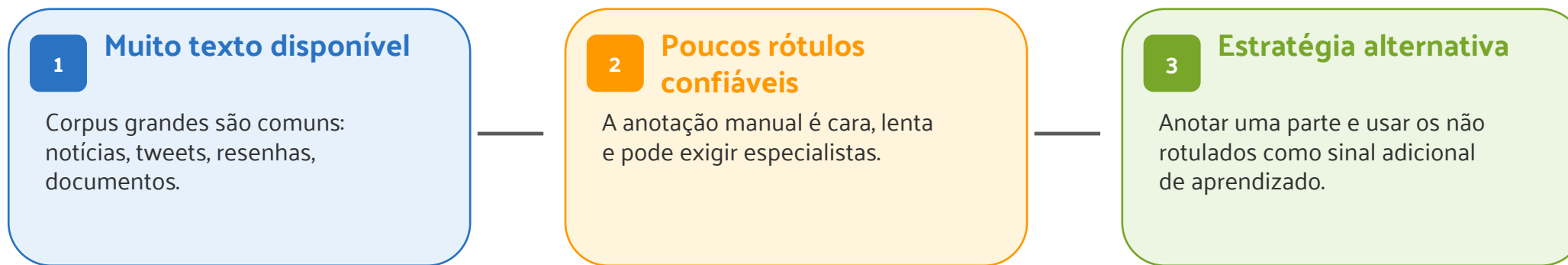
Você precisa coletar ou construir
tudo do zero.

Questão central

**Como aproveitar os textos não rotulados sem precisar
anotar tudo manualmente?**

O custo da anotação muda a estratégia

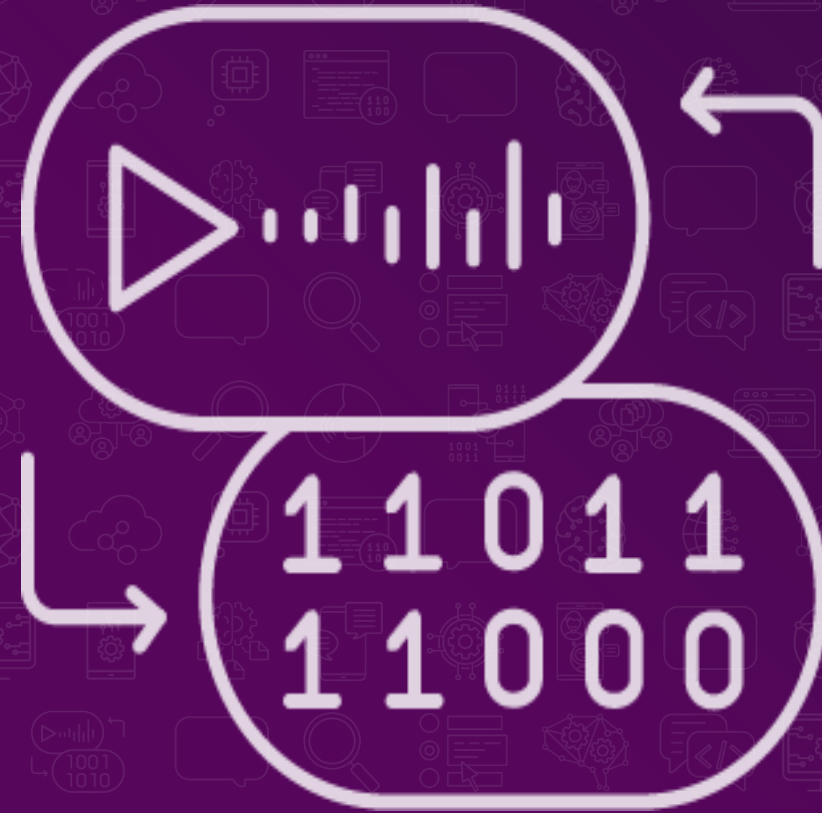
Em PLN, rótulos confiáveis costumam exigir critérios linguísticos, anotadores treinados e revisão de consistência.



O aprendizado transdutivo entra quando o **conjunto não rotulado já é conhecido** e queremos melhorar as previsões justamente nele.



2. Da indução à transdução



Três formas de inferência

Indução

Aprender uma relação geral a partir dos dados.

$$Y = f(X)$$

Dedução

Aplicar uma relação já aprendida a novos pontos.

$$X_{\text{novo}} \rightarrow \hat{Y}$$

Transdução

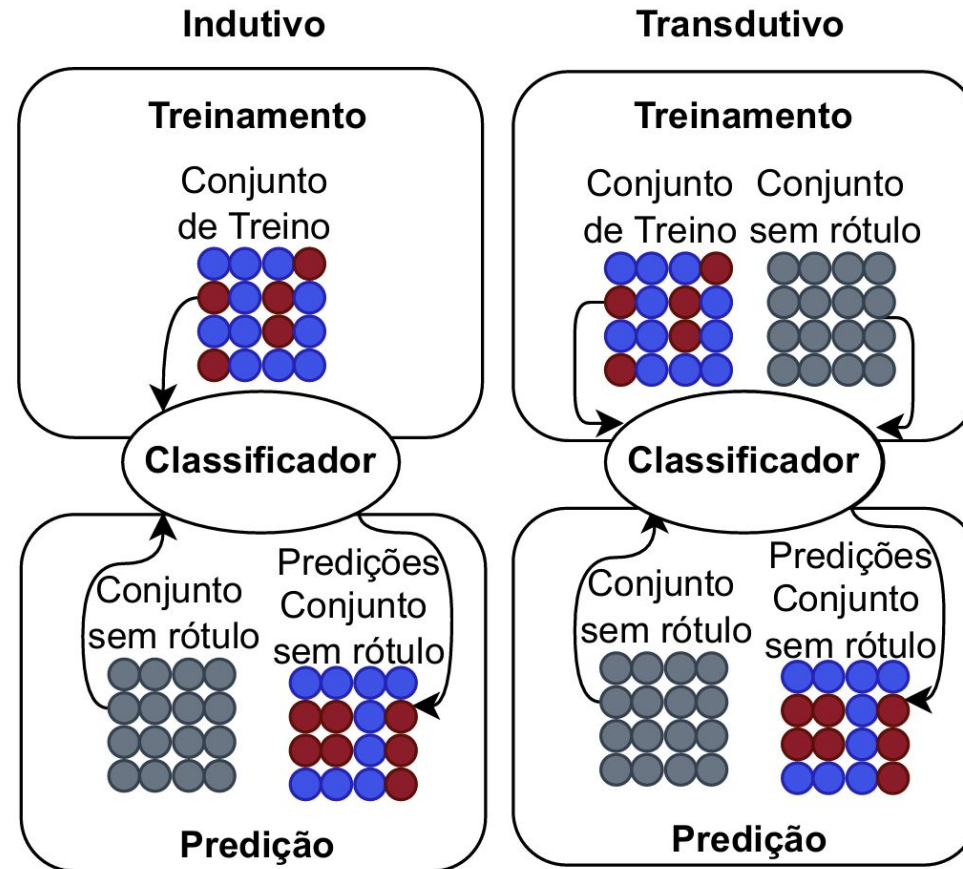
Inferir diretamente os rótulos dos pontos de interesse.

$$(X_{\text{treino}}, Y_{\text{treino}}, X_{\text{teste}}) \rightarrow \hat{Y}_{\text{teste}}$$

Resumo intuitivo

No aprendizado transdutivo, não buscamos necessariamente uma função universal para qualquer dado futuro. Buscamos **boas respostas para um conjunto específico de exemplos** não rotulados já disponível.

Indutivo vs. transdutivo



No caso transdutivo, os dados não rotulados participam do treinamento – não apenas da predição final.

Formalização mínima

A transdução recebe poucos rótulos + muitos exemplos sem rótulo e devolve rótulos estimados apenas para esse conjunto.

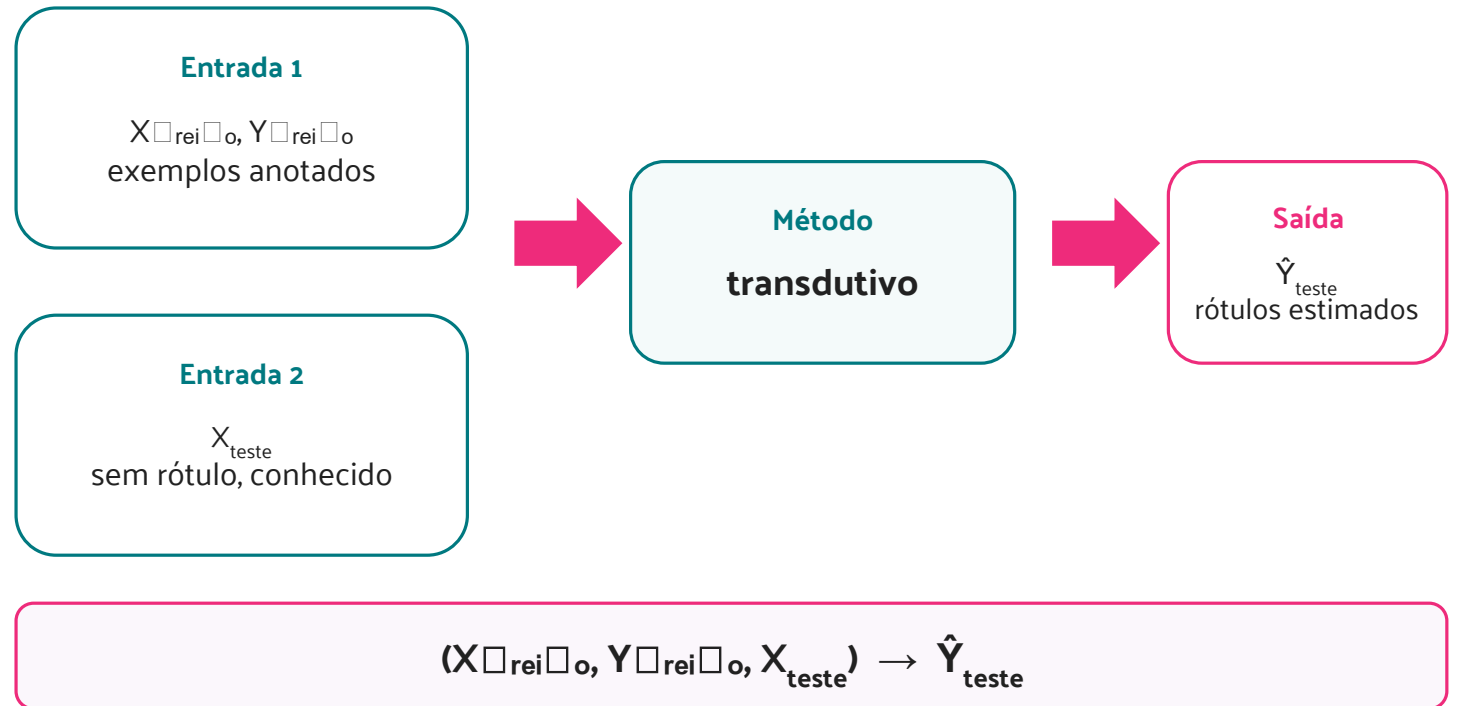
Exemplo: corpus fechado

$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$

x_1	positivo
x_2	negativo
x_3	positivo
x_4	?
x_5	?
x_6	?

Rotulados: $X_{\text{rei}}, Y_{\text{rei}}$

Alvo sem rótulo: X_{teste}



Quando faz sentido usar transdução?

1

O teste é conhecido

Ex.: corpus completo a ser anotado offline.

2

Não preciso generalizar agora

O objetivo é rotular aquele conjunto.

3

Anotar tudo é inviável

Uso rótulos manuais apenas como semente.



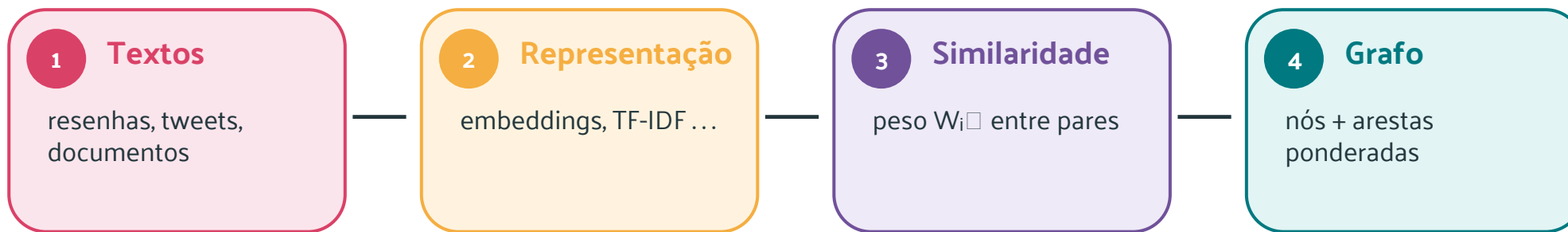
3. Propagação de rótulos

A ideia de espalhar informação em um grafo



Como transformar textos em um grafo?

A ideia: representar textos como nós e conectar exemplos semelhantes.



Matriz de afinidade

W_{ij} guarda o quanto x_i e x_j são parecidos.
Quanto maior W_{ij} , maior a influência entre os nós.

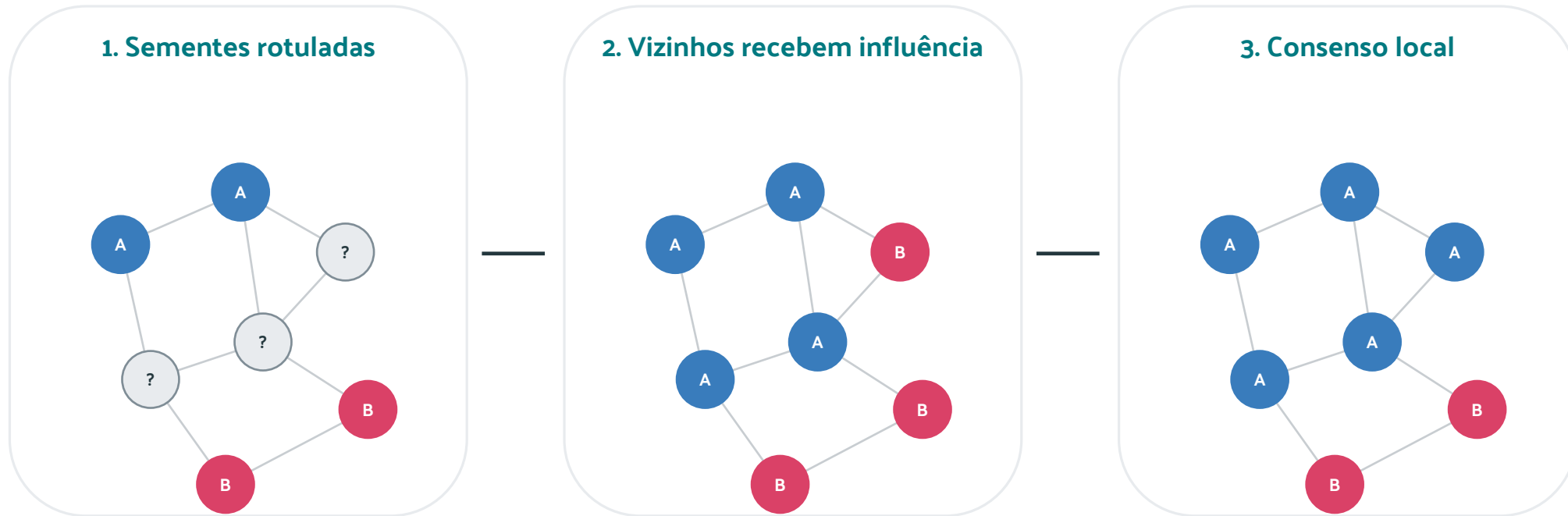
Hipótese de suavidade

Exemplos próximos no grafo tendem a compartilhar rótulos ou distribuições de rótulos semelhantes.

Em PLN, a qualidade do grafo depende diretamente da representação textual e da medida de similaridade.

Intuição: textos parecidos formam vizinhanças

Poucos nós começam rotulados. A cada iteração, eles influenciam seus vizinhos até formar regiões consistentes.



o rótulo não é “adivinhado” isoladamente; ele emerge da vizinhança

Propagation vs. Spreading

Os dois espalham informação no grafo, mas diferem no controle da propagação.

Label Propagation

Atualiza rótulos diretamente a partir dos vizinhos.

Regra típica:

$$Y^{(t+1)} = D^{-1} W Y^{(t)}$$

Mais simples, mas pode ser mais sensível a ruído nos rótulos ou nas arestas.

Label Spreading

Introduz suavização/regularização durante a propagação.

Regra típica:

$$Y^{(t+1)} = \alpha L Y^{(t)} + (1-\alpha)Y^{(0)}$$

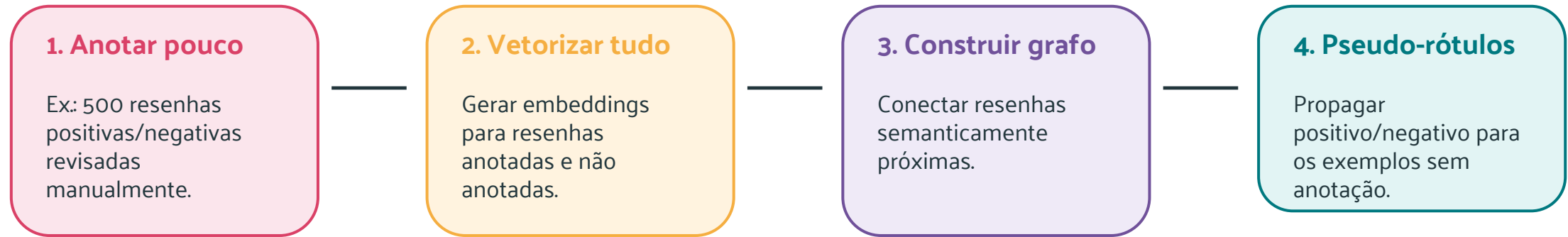
Mais controlado; tende a ser mais robusto quando as sementes são incertas.

Como ler o parâmetro α no Label Spreading

α alto \rightarrow confia mais na estrutura do grafo; α baixo \rightarrow preserva mais os rótulos iniciais. O termo $(1-\alpha)Y^{(0)}$ funciona como uma “âncora” contra propagação excessiva de ruído.

Passo a Passo

Cenário: poucas resenhas anotadas e muitas resenhas sem rótulo.



Resultado esperado

Base parcialmente anotada → base ampliada

Os pseudo-rótulos podem ser usados para análise, revisão humana ou novo treinamento supervisionado.

Cuidados práticos

Não confiar cegamente

Erros podem se propagar se as sementes ou a similaridade estiverem ruins.

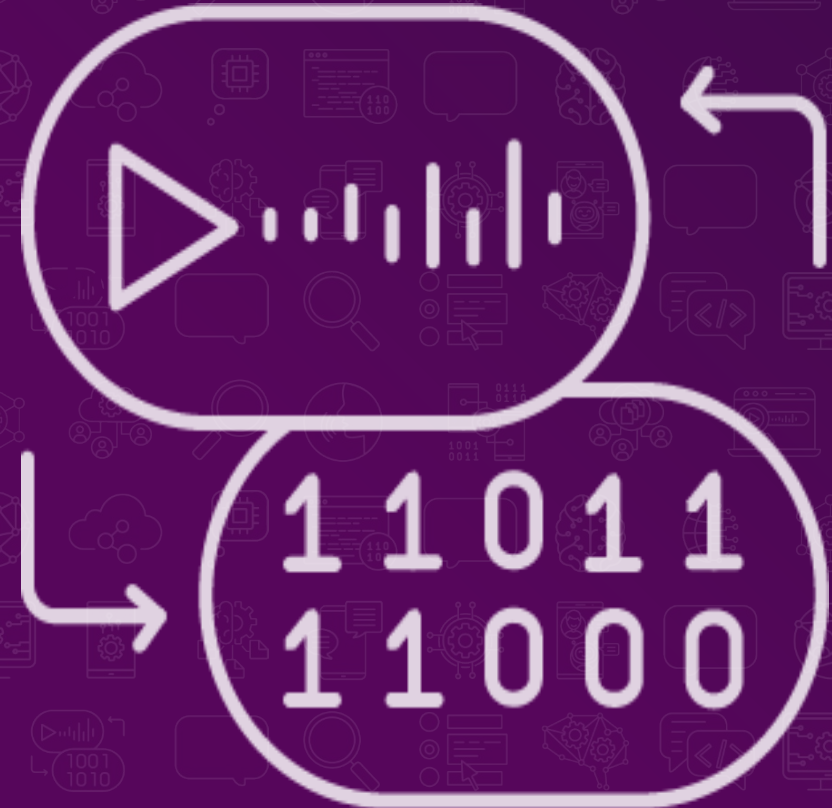
Avaliar depois

Conferir confiança, vizinhos, amostra humana e impacto em validação.

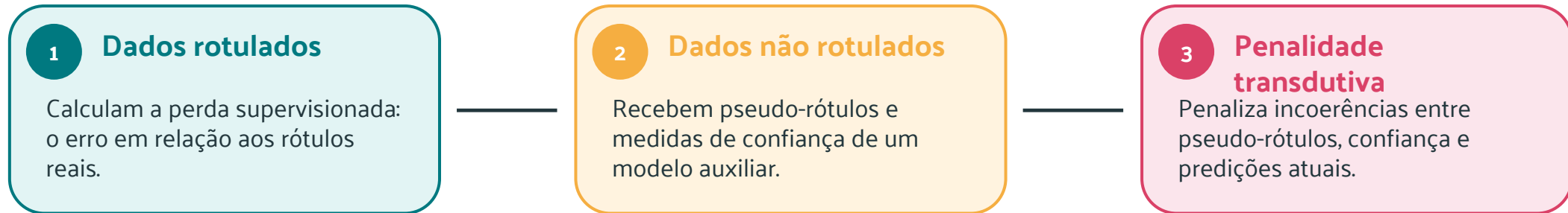


4. Ajuste fino transdutivo

Quando a transdução entra na função de custo



Como dados não rotulados entram no ajuste fino?



$$\text{Perda final} = \text{perda supervisionada} + \lambda \cdot \text{perda transdutiva}$$

λ controla quanto os dados não rotulados influenciam o treinamento.

Primeiro: onde a transdução entra na perda total?

Antes de detalhar os símbolos, localizamos a perda transdutiva dentro do objetivo de treinamento.

Perda supervisionada

Vem dos exemplos rotulados.

Perda Supervisionada = média de $\ell(f\theta(x_i), y_i)$

Pergunta: “o modelo acertou os rótulos reais?”

Perda transdutiva

Vem dos exemplos não rotulados.

Perda transdutiva = média de $k_i \cdot k_{\square} \cdot \delta_i \cdot S_i$

Pergunta: “o modelo está coerente com os dados sem rótulo?”

Perda total = Perda supervisionada + λ · Perda transdutiva

λ pequeno: quase supervisionado | λ grande: maior influência dos pseudo-rótulos



De onde vem cada termo da fórmula?

Depois de ver a fórmula, podemos ler cada símbolo como uma etapa do processo.

1. Modelo auxiliar f_{θ_0}

Recebe X_{teste} sem rótulo e gera:

- pseudo-rótulo: \hat{y}_i^0
- confiança: $k(x_i)$

Ex.: $f_{\theta_0}(x_3) = \text{classe 0 com confiança } 0,6$

2. Comparação δ

Compara pseudo-rótulos de dois exemplos:

$$\begin{aligned}\delta(x_i, x_j) &= 1 \text{ se } \hat{y}_i^0 \neq \hat{y}_j^0 \\ \delta(x_i, x_j) &= 0 \text{ caso contrário}\end{aligned}$$

Só pares de classes diferentes geram penalidade.

3. Modelo em ajuste f_{θ}

Produz as previsões atuais:

$$f_{\theta}(x_i), f_{\theta}(x_j)$$

A função S mede se essas previsões estão parecidas, por exemplo por cosseno.

4. Penalidade

Aumenta quando:

- pseudo-rótulos são diferentes;
- f_{θ_0} está confiante;
- f_{θ} trata os exemplos como parecidos.

$$\text{Termo: } k_i \cdot k_j \cdot \delta_{ij} \cdot S_{ij}$$

Leitura final: a penalidade é alta quando f_{θ_0} separa dois exemplos com confiança, mas f_{θ} ainda produz previsões muito parecidas.

A função de custo fica com dois termos

Agora abrimos o termo transdutivo: ele é calculado sobre pares de exemplos não rotulados.

$$\text{Perda Transdutiva } (X_{\text{teste}} | f_{\theta}, S, \delta, k) = (1/N) \sum_{(i,j)} k(x_i) \cdot k(x_j) \cdot \delta(x_i, x_j) \cdot S(x_i, x_j)$$

$$k(x_i) \cdot k(x_j)$$

CONFIANÇA

quanto confiar nos pseudo-rótulos gerados por f_{θ}

$$\delta(x_i, x_j)$$

SELEÇÃO DO PAR

vale 1 se os pseudo-rótulos forem diferentes; caso contrário, 0

$$S(x_i, x_j)$$

SIMILARIDADE

quão parecidas são as previsões atuais geradas por f_{θ}

Leitura intuitiva: a penalidade cresce quando o par deveria ser diferente, os pseudo-rótulos são confiáveis e o modelo atual ainda produz previsões muito parecidas.

Exemplo numérico: juntando todos os termos

Agora aplicamos a sequência: perda supervisionada + confiança + δ + similaridade + perda total.

1. Dados rotulados

$x_1: y=1, f\theta(x_1)=[0,2; 0,8]$
 $x_2: y=0, f\theta(x_2)=[0,7; 0,3]$

Perda supervisionada = 0,29(entropia cruzada)

2. Pseudo-rótulos de $f\theta_0$

x_3 : classe 0, $k=0,6$
 x_4 : classe 1, $k=0,7$

Como são classes diferentes:
 $\delta(x_3, x_4)=1$

3. Predições atuais de $f\theta$

$f\theta(x_3)=[0,6; 0,4]$
 $f\theta(x_4)=[0,7; 0,3]$

São muito parecidas:
 $S(x_3, x_4) \approx 0,9827$ (similaridade de cosseno)

Perda total = Perda supervisionada + λ · Perda transdutiva

Perda transdutiva = $k_3 \cdot k_4 \cdot \delta_{34} \cdot S_{34} = 0,6 \cdot 0,7 \cdot 1 \cdot 0,9827 = 0,4127$

Com $\lambda=1$: Perda total = 0,29 + 0,4127 = 0,7027

Interpretação: o modelo recebe uma penalidade porque está aproximando exemplos que o pseudo-rotulador separou com alguma confiança.



Como avaliar pseudo-rótulos gerados?

Confiança

probabilidade máxima ou margem entre classes

Consistência local

vizinhos próximos têm o mesmo pseudo-rótulo?

Amostra humana

revisão manual representativa

Impacto downstream

modelo melhora com pseudo-rótulos?

Regra prática: não trate pseudo-rótulo como verdade absoluta. Trate como hipótese controlada por confiança, consistência e validação.

Benefícios e limitações

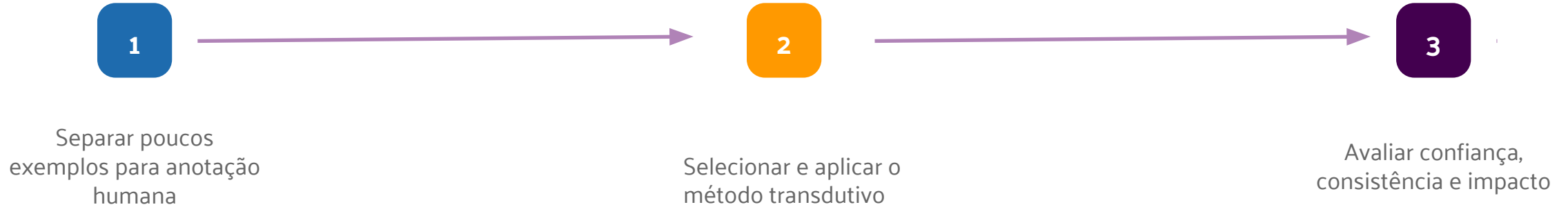
Benefícios

- ✓ menos dependência de bases rotuladas
- ✓ melhor uso de recursos humanos
- ✓ predição especializada para o conjunto conhecido

Limitações

- ! maior complexidade computacional
- ! baixa generalização fora do conjunto
- ! dependência de boas predições iniciais

Roteiro prático para aplicar em um projeto de PLN



O resultado ideal não é “anotar automaticamente tudo”, mas reduzir custo mantendo controle de qualidade.

Mensagens finais

- 1 Transdução é útil quando o conjunto não rotulado já está disponível.
- 2 Ela transforma dados sem rótulo em sinal de aprendizado.
- 3 Em PLN, é especialmente interessante para anotação e construção de corpora.
- 4 O ganho depende de avaliação cuidadosa dos pseudo-rótulos.



Obrigada!

