

Tradução entre Sistemas de Classes em NLI

[Inferência em Linguagem Natural](#) [Tradução entre Sistemas de Classes em NLI](#) [SICK-BR](#) [ASSIN](#) [ASSIN2](#) [MNLI](#) [SNLI](#)

Classes

- E: *Entailment*/Acarretamento
- P: Paráfrase
- \emptyset : *None*
- C: Contradição
- N: Neutralidade

Tabela de Tradução

Origem		Destino		
Sistema	Classe	ASSIN	ASSIN2	SICK
ASSIN	E(a,b)		E(a,b)	E(a,b)
	P(a,b)		$E(a,b) \wedge E(b,a)$	$E(a,b) \wedge E(b,a)$
	$\emptyset(a,b)$		$\emptyset(a,b)$	$C(a,b) \oplus N(a,b)$
ASSIN2	E(a,b)	$E(a,b) \oplus P(a,b)$		E(a,b)
	$\emptyset(a,b)$	$\emptyset(a,b)$		$C(a,b) \oplus N(a,b)$
SICK	E(a,b)	$E(a,b) \oplus P(a,b)$	E(a,b)	
	C(a,b)	$\emptyset(a,b)$	$\emptyset(a,b)$	
	N(a,b)	$\emptyset(a,b)$	$\emptyset(a,b)$	

⚠ MNLI e SNLI usam sistemas de classes análogos ao SICK e a mesma regra de conversão pode ser aplicada dos demais para eles.

Interpretação

- A classe de destino c_D é a classe que o modelo de destino corretamente classificaria um exemplo da classe de origem c_O num conjunto de dados do sistema de origem.
- Nos casos em que $c_D = A \wedge B$ é esperado que o modelo dê ambas as classificações A and B . Em todos os casos na tabela isso implica uma avaliação dupla pelo modelo sendo testado, garantindo que a classe composta seja distinta de outras classes, de forma que o modelo de destino pode ser avaliado no modelo de origem.
- Nos casos em que $c_D = A \oplus B$ o modelo de destino retornará ou uma classe ou A ou B para classificar o exemplo fornecido, o que implica que em tempo de avaliação, ambas as classes A e B devem ser traduzidas para a c_O correspondente.
- Nos casos em que temos a mesma c_D para mais de uma c_O do sistema de origem não conseguimos fazer a avaliação do modelo de destino no dataset de origem, precisamos então traduzir as c_O s correspondentes no dataset de origem para a c_D do dataset de destino e, então, fazer a avaliação.
- O SICK possui anotação de todos os pares em ambos os sentidos (a, b) e (b, a) . Durante o treinamento, normalmente o que fazemos é considerar duas entradas distintas, uma (a, b) e outra (b, a) . Quando estou testando um modelo que diferencia paráfrase em um corpus como esse, seguindo a tradução acima, poderia acontecer de dar acerto a uma previsão do modelo que é inconsistente. Por exemplo, se $E(a, b)$ e $N(b, a)$ e meu modelo classifica o par (a, b) com P , eu consideraria isso um acerto (porque a tabela acima me manda traduzir $P(a, b) \rightarrow E(a, b)$), mesmo que globalmente a resposta estivesse errada. Isso poderia inflar a métrica de desempenho do ASSIN no SICK, por exemplo, mas fazer qualquer verificação nesse sentido durante a avaliação do modelo atrapalharia a comparabilidade com outros resultados de outros modelos, além de ser inconsistente com a maneira com a qual os modelos são treinados, na qual mede-se a performance apenas localmente, exemplo à exemplo. Dessa

forma, é melhor utilizar a tradução da tabela acima, mesmo neste caso, mas mantendo essa ressalva no que tange à análise desses resultados.