

Motivation

- Multi-domain NMT, which translates multiple domains within a single model, should capture both **general** and **domain-specific** knowledge.
- Mutual Information (MI) between domain and translation represents the **dependency between the domain and the translated sentence**.
- A model with high MI tends to retain the domain-specific term in its translation.

Source	Beschreib ... Summenberechnung fur ein gegebenes Feld oder einen gegebenen Ausdruck.
Reference	Describe a way of computing totals for a given field or expression.
Model A with Low MI	Describe the kind of calculation for a given field or expression.
Model B with High MI	Describe the way of computing totals for a given field or expression

Table 1: Examples from different MI distributions in IT

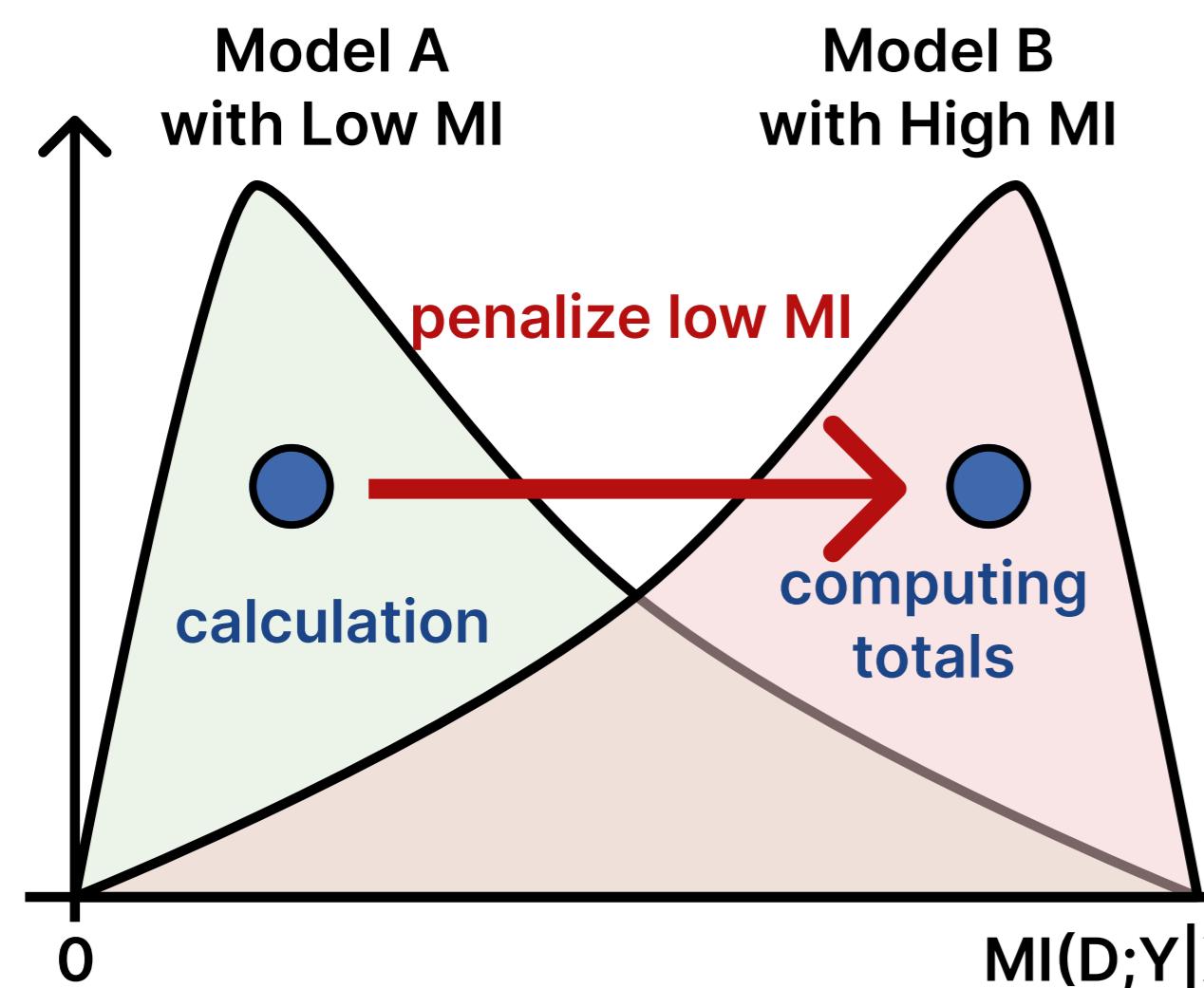


Fig 1. Overview of two models with different MI distributions

- In this study, we specialize multi-domain NMT by penalizing low MI to have higher value.

MI in Multi-domain NMT

Given X (source sentence), Y (target sentence), D (domain), MI is calculated as follows:

$$MI(D; Y|X) = \mathbb{E}_{D,X,Y} \left[\log \frac{p(Y|X, D)}{p(Y|X)} \right]$$

We approximate it with a parameterize model, namely cross-MI (XMI). (DA = domain adapted, G = general)

$$XMI(D; Y|X) = \mathbb{E}_{D,X,Y} \left[\log \frac{p_{DA}(Y|X, D)}{p_G(Y|X)} \right]$$

