Controlling MT for Multiple Attributes with Additive Interventions

Andrea Schioppa Artem Sokolov David Vilar Katja Filippova

Google Research

Motivation

- Increase user trust by enabling agency on the attributes of interest.
- Length constraints in video subtitling.
- Monotonic translations for online education and speech MT.
- Generation in the desired politeness register.

Baseline: Tagging

Disadvantages:

- Continuous attributes need to be bucketized.
- Effect of tags order is poorly understood.
- Modifications needed if controlling only a *subset* of attributes.

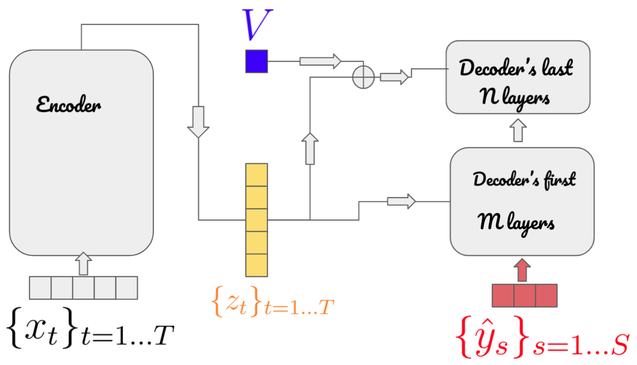

Figure 1:Our approach

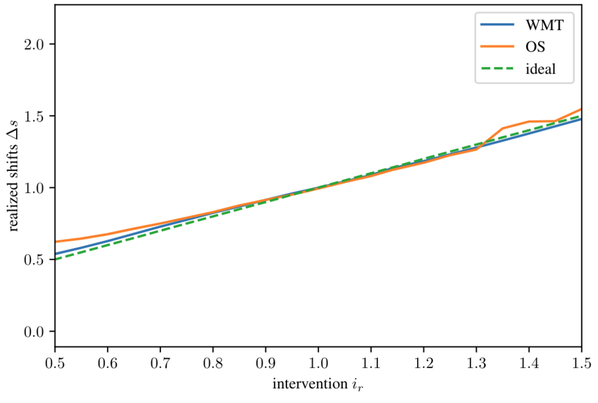

Figure 2:Fine-grained length control.

Additive Control

- Apply a control intervention V by shifting *each* of the final encoder representations $z'_t = z_t + V$.
- Decompose V as a sum across attributes: $V = \sum_a w_a V_a$.
- Set $w_a = 0$ when no control is requested (`Neutral` mode).

Advantages

- **Simultaneous control of multiple attributes**.
- **No bucketing of continuous attributes**,
  $\Rightarrow$ more fine-grained control.
- Works when z'_t are fed to only the last 2 decoder layers,
  $\Rightarrow$ **faster fine-tuning**.
- **Order-invariance** and control of **any attribute subset**.

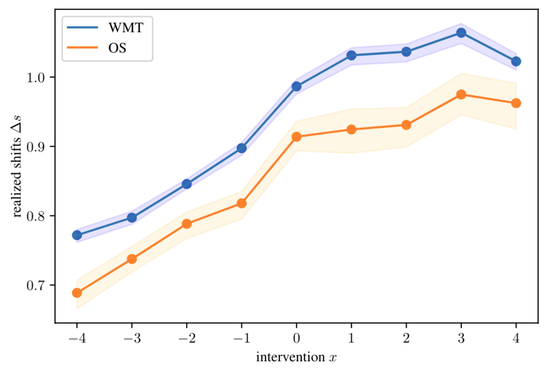

Figure 3:Tagging gives coarse length control with different OOD behavior.

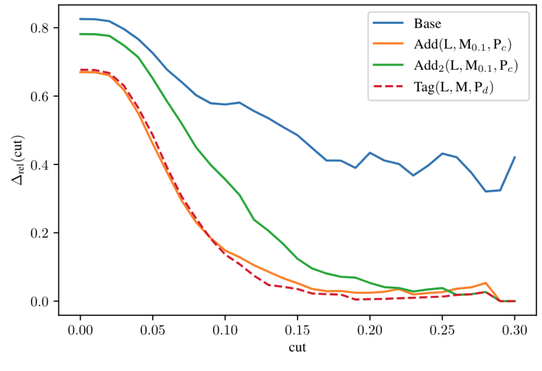

Figure 4:Increasing monotonicity: the lower the curve the better.

Model	Mode	BLEU
Base	-	27.11
Tag(L, P_d, M)	Oracle	26.58
Tag(L, M, P_d)	Oracle	27.32
Add	Neutral	26.92
Add	Oracle	26.99
Add$_2$	Neutral	27.43
Add$_2$	Oracle	27.76

Table 1:BLEU on WMT EN-DE.

input	why don't you come sit down with me?
reference	こっちに来て一緒に座らない?
	over here come together not sitting down?
$i_r = 0.30$	座って
	sit down!
$i_r = 0.50$	一緒に 座って
	together sit down!
$i_r = 0.70$	一緒に座ろう
	together let's sit down
$i_r = 0.90$	一緒に 座ったら?
	together why don't sit down?
$i_r = 1.00$	一緒に座らないか?
	together not sitting down?
$i_r = 1.10$	俺と一緒に座ったらどうだ?
	with me together when sitting down how's it?
$i_r = 1.20$	なぜ私と一緒に座らない?
	why with me together not sitting down?
$i_r = 1.30$	なぜ私と一緒に座らないの?
	why with me together not sitting down?
$i_r = 1.50$	なぜあなたは私と一緒に 座らないか?
	why you with me together not sitting down?

Table 2:Controlling length of EN-JA translations. The system adjusts optional words, e.g. pronouns.

Model	Method	*unknown*	*informal*	*polite*	*formal*
Base	-	14.60	14.92	15.61	21.95
Tag	Neutral	14.51	15.66	15.06	27.87
Tag	Oracle	14.42	19.63	20.03	52.85
Add	Neutral	15.16	14.25	17.24	38.31
Add	Oracle	15.60	19.32	20.28	53.28
Add$_2$	Neutral	16.11	15.42	15.73	20.44
Add$_2$	Oracle	16.31	17.79	18.97	45.15

Table 3:Politeness control improves quality of EN-JA translations.