

Investigating the Effectiveness of Multiple Expert Models Collaboration

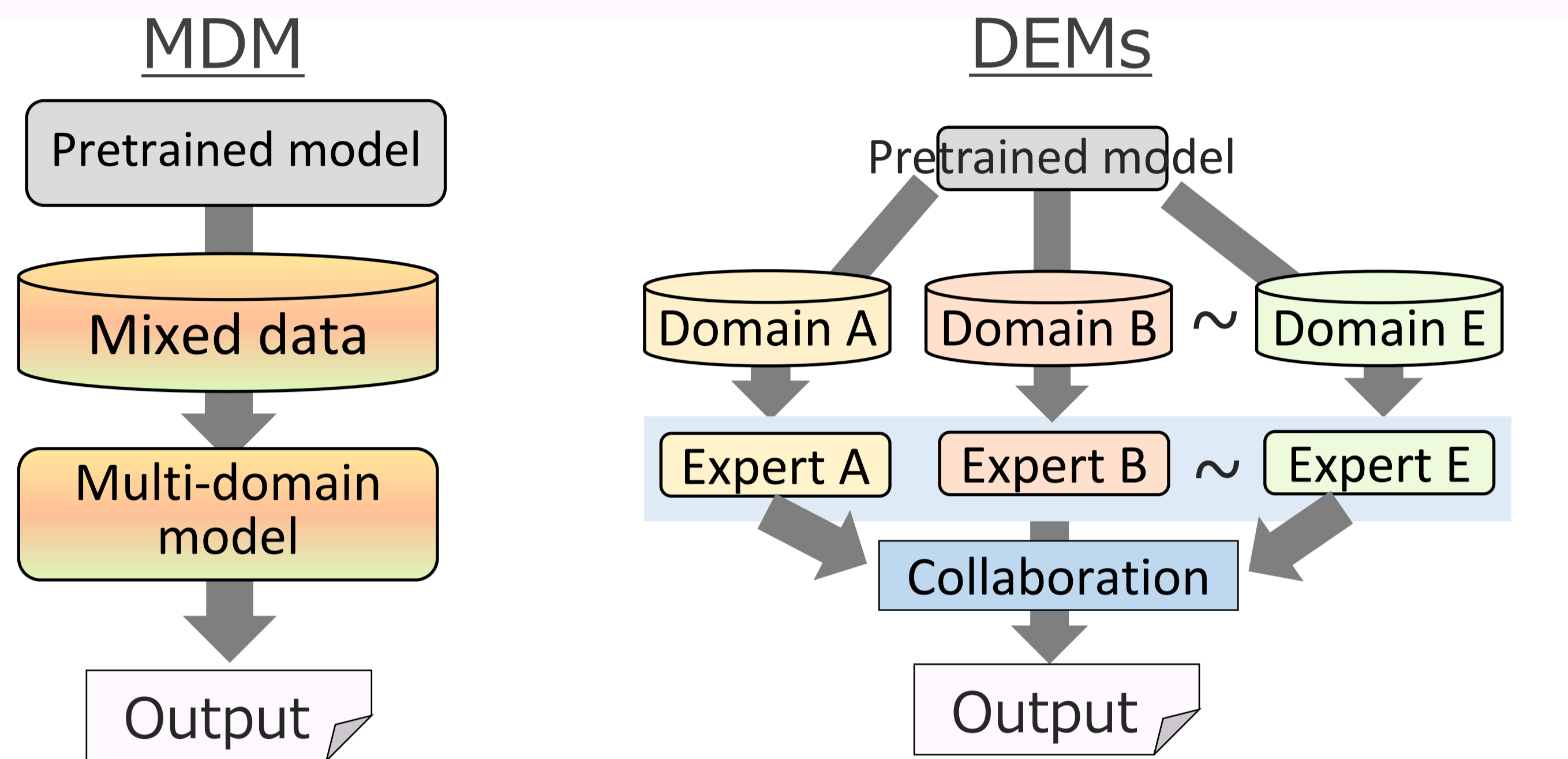
Ikumi Ito¹, Takumi Ito^{1,2}, Jun Suzuki^{1,3}, Kentaro Inui^{4,1,3}

¹Tohoku University ²Langsmith Inc. ³RIKEN ⁴MBZUAI

Summary

- We evaluated two approaches in multi-domain machine translation:
 - a single **Multi-Domain Model (MDM)**
 - multiple **Domain Expert Models (DEMs)**
- The DEMs has the potential to outperform MDM
- We investigated the effective collaboration methods for DEMs
 - Minimum Bayes Risk** is the best way

Task Setting (Machine Translation)



- Mixed domain data amount: 1.15M
- A single multi-domain model
- 5 different domain data amount: 440k, 260k, 225k, 200k, 20k
- 5 expert models

Collaboration Methods for DEMs

- Ensemble (ENS)

$$p(y_{\leq t}|x) = \frac{1}{N} \sum_{i=1}^N p_i(y_{\leq t}|x)$$
- Quality Estimation (QE)

QE metric: MS-COMET-QE-22 [Kocmi+ '22]

$$\arg \max_{C_i \in \mathcal{C}} \text{QE}(Src, C_i)$$

each candidate
- Minimum Bayes Risk (MBR) [Kumar and Byrne '04]

MBR metric: MS-COMET-22 [Kocmi+ '22]

$$\arg \max_{C_i \in \mathcal{C}} \frac{1}{|C|} \sum_{j \neq i} \text{MBR}(Src, C_i, C_j)$$

pseudo reference

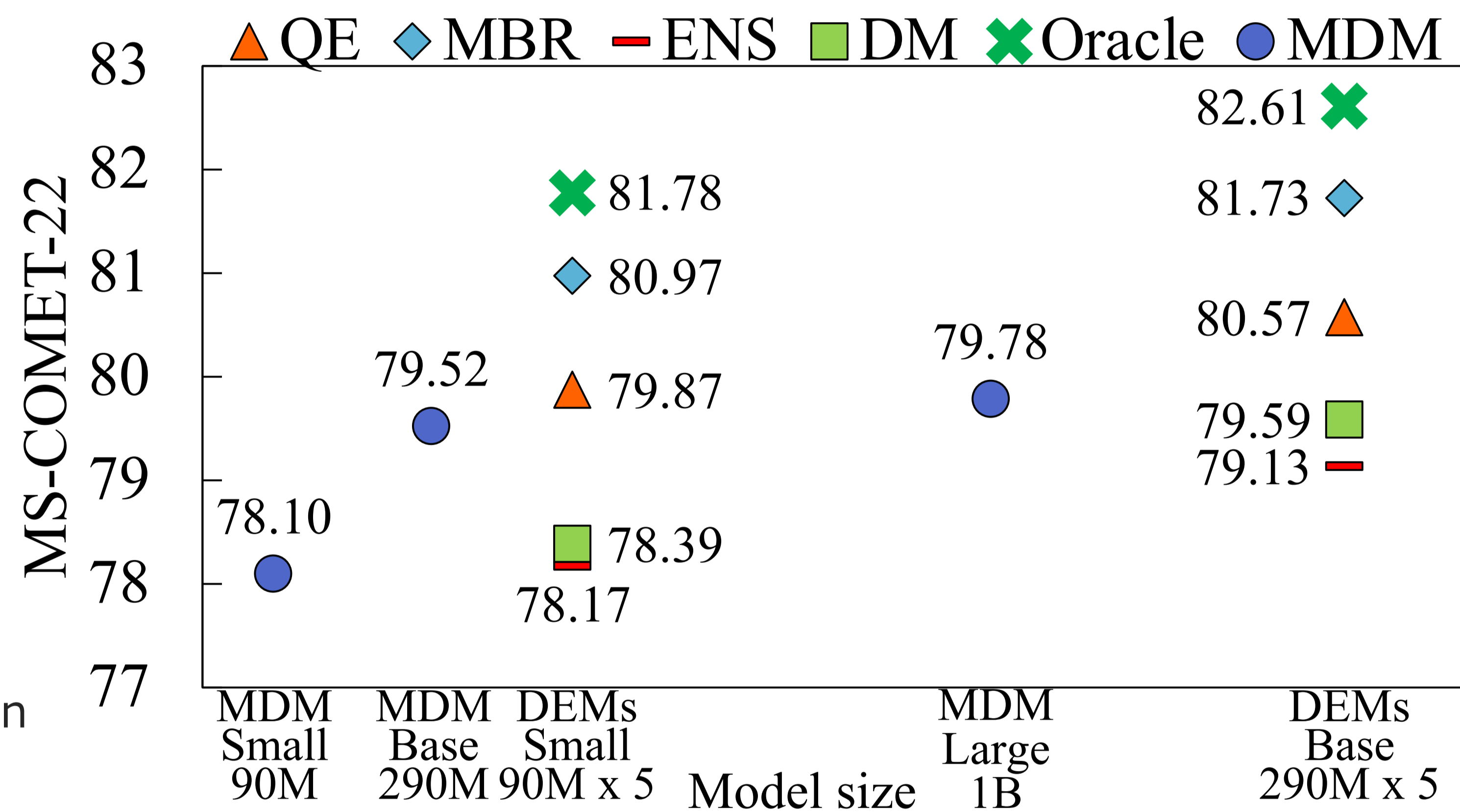
Results

- 90M x 5 DEMs (MBR) outperformed 1B MDM
- MBR > QE > ENS

😊 Selection from candidates generated by each expert

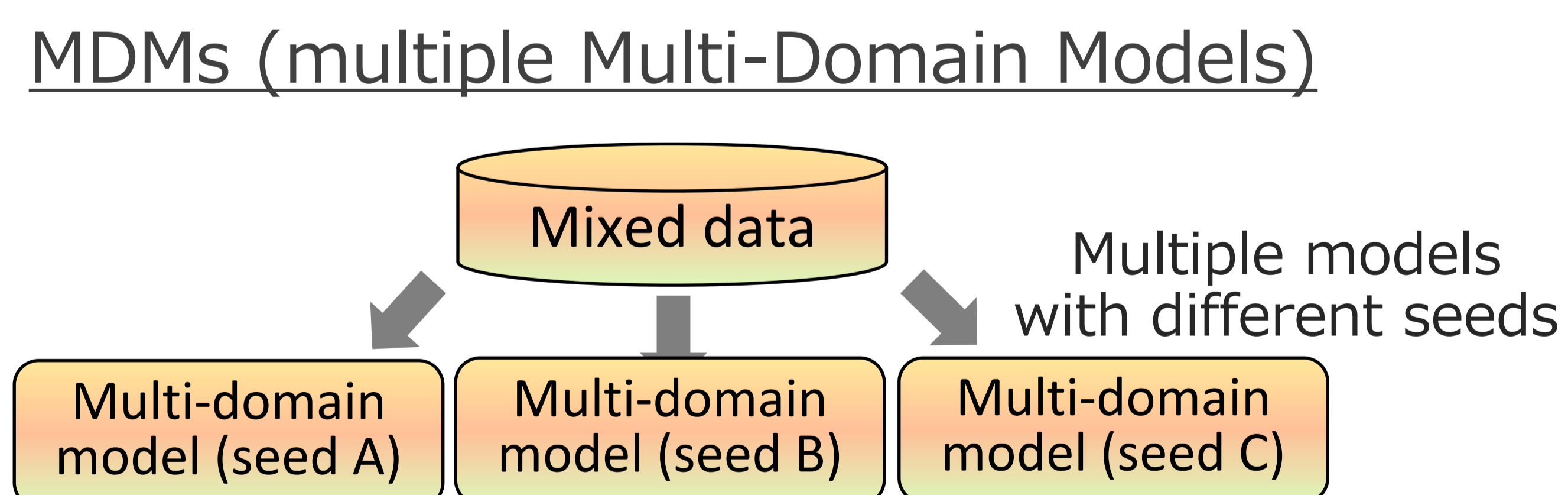
😞 ENS is equal to or less than MDM

Domain Match (DM): Select the MT model according to the input domain
Oracle: Select an output with the highest evaluation metric score



Analysis: Success Factors for DEMs

Domain-based data separation is effective



	AVG	QE	MBR	ENS
DEMs (90M x 5)	77.20	79.87	80.97	78.17
MDMs (90M x 5)	+0.95	-1.26	-2.34	-0.02

DEMs > MDMs
though training data amount was DEMs < MDMs

Multi-domain capability of the selection metric is important

Change to metrics trained on fewer domains:

MS-COMET-QE-22 → wmt22-cometkiwi-da [Rei+ '22]

MS-COMET-22 → wmt22-comet-da [Rei+ '22]
(15 domains) (news domain)

😞 QE and MBR with COMET have smaller gains than with MS-COMET

