

RoBERTa training (avg_FIRST): cross-entropy loss

