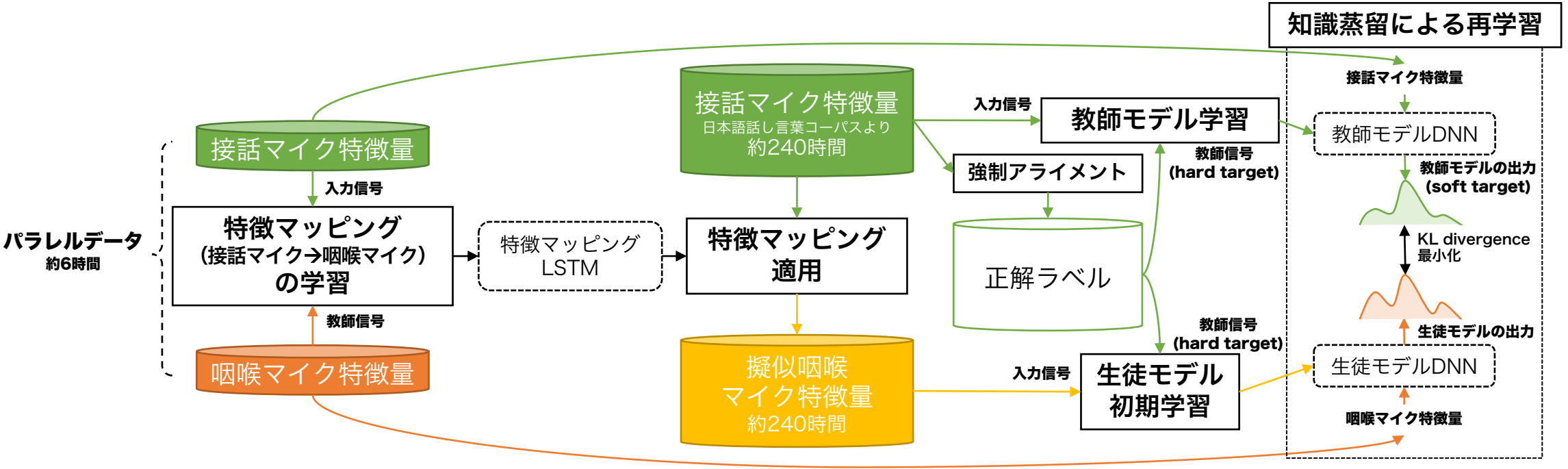
知識蒸留による再学習
パラレルデータ
約6時間
接話マイク特徴量
入力信号
特徴マッピング
(接話マイク→咽喉マイク)
の学習
教師信号
咽喉マイク特徴量
特徴マッピング
LSTM
接話マイク特徴量
日本語話し言葉コーパスより
約240時間
特徴マッピング
適用
擬似咽喉
マイク特徴量
約240時間
入力信号
教師モデル学習
強制アライメント
正解ラベル
教師信号
(hard target)
教師信号
(hard target)
入力信号
生徒モデル
初期学習
接話マイク特徴量
教師モデルDNN
教師モデルの出力
(soft target)
KL divergence
最小化
生徒モデルの出力
生徒モデルDNN
咽喉マイク特徴量