Domain Adaptation for WSD (語義曖昧性の領域適応)

Hiroyuki Shinnou

WSD: Word Sense Disambiguation

語義曖昧性解消

bank
1. finace
2. land

- He sat on the bank of the river. → 2
- I deposited money in the bank. → 1

. . .
. . .
. . .

Supervised Learning

教師あり学習

$$D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$$

x_k : sentence including the target word (ex. bank)

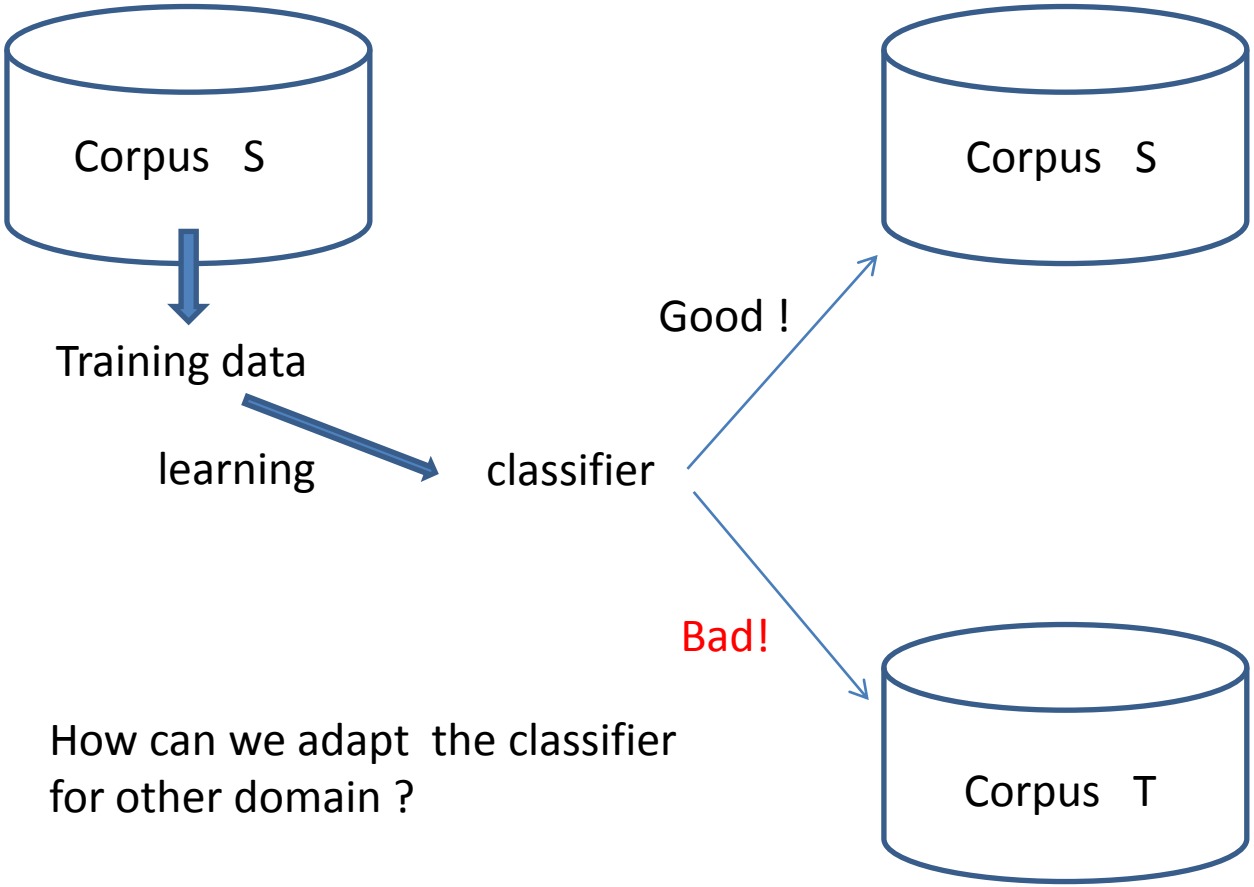
y_k : meaning number (ex. 1 or 2)

$$y = f(x)$$

Solution is to make above function f

Domain Adaptation

領域適応



MAP: Maximum a posteriori

事後確率最大化

$$\arg max\ P(y|x) = \arg max\ \frac{P(y)P(x|y)}{P(x)}$$
$$= \arg max\ P(y)P(x|y)$$

That is, problems of DA for WSD:

$$P_S(y) \neq P_T(y)$$ ← We agree !

$$P_S(x|y) \neq P_T(x|y)$$ ← Really ?

Covariate Shift

菊池君のテーマ

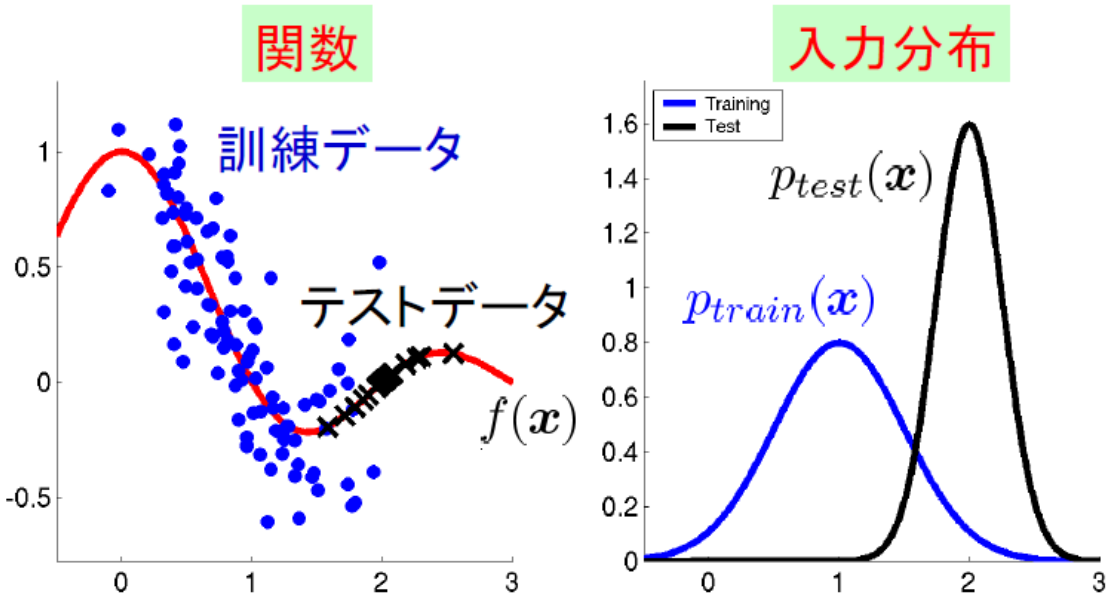
共変量シフト

$$P_S(y|x) = P_T(y|x)$$

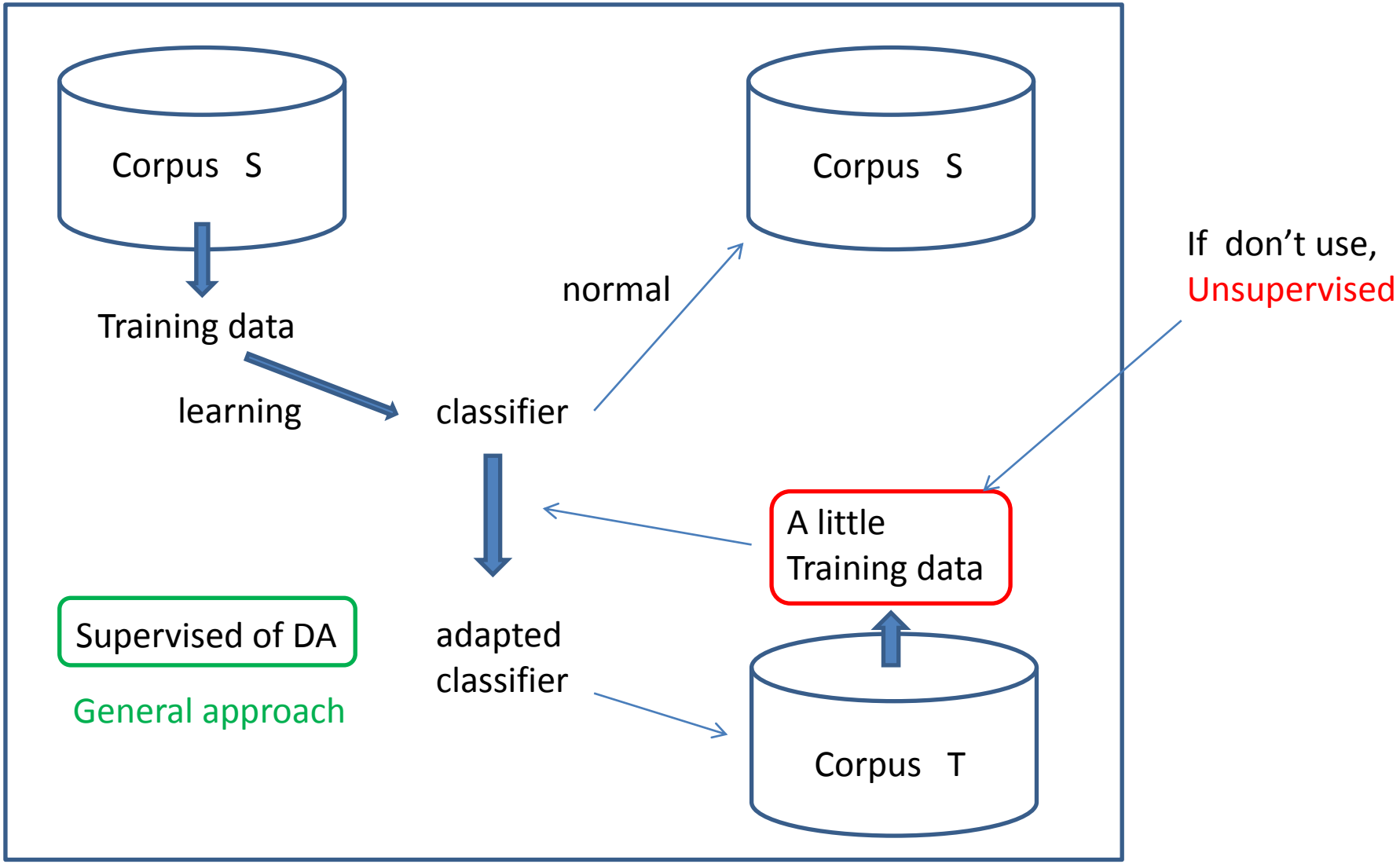
$$P_S(x) \neq P_T(x)$$

This is just
DA for WSD

http://sugiyama-www.cs.titech.ac.jp/~sugi/2008/NECsoft-MachineLearning4-jp.pdf

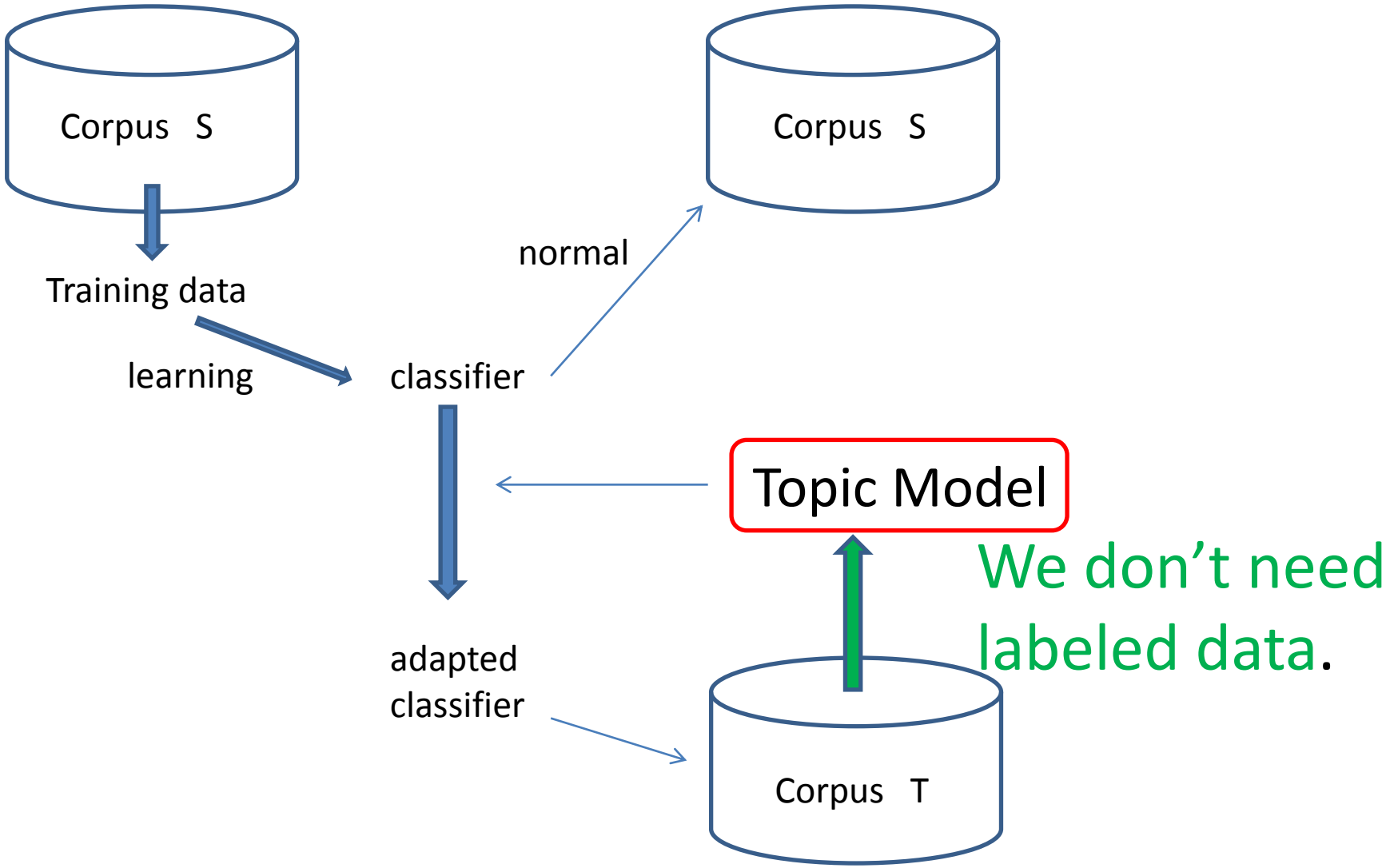


Unsupervised Learning of DA



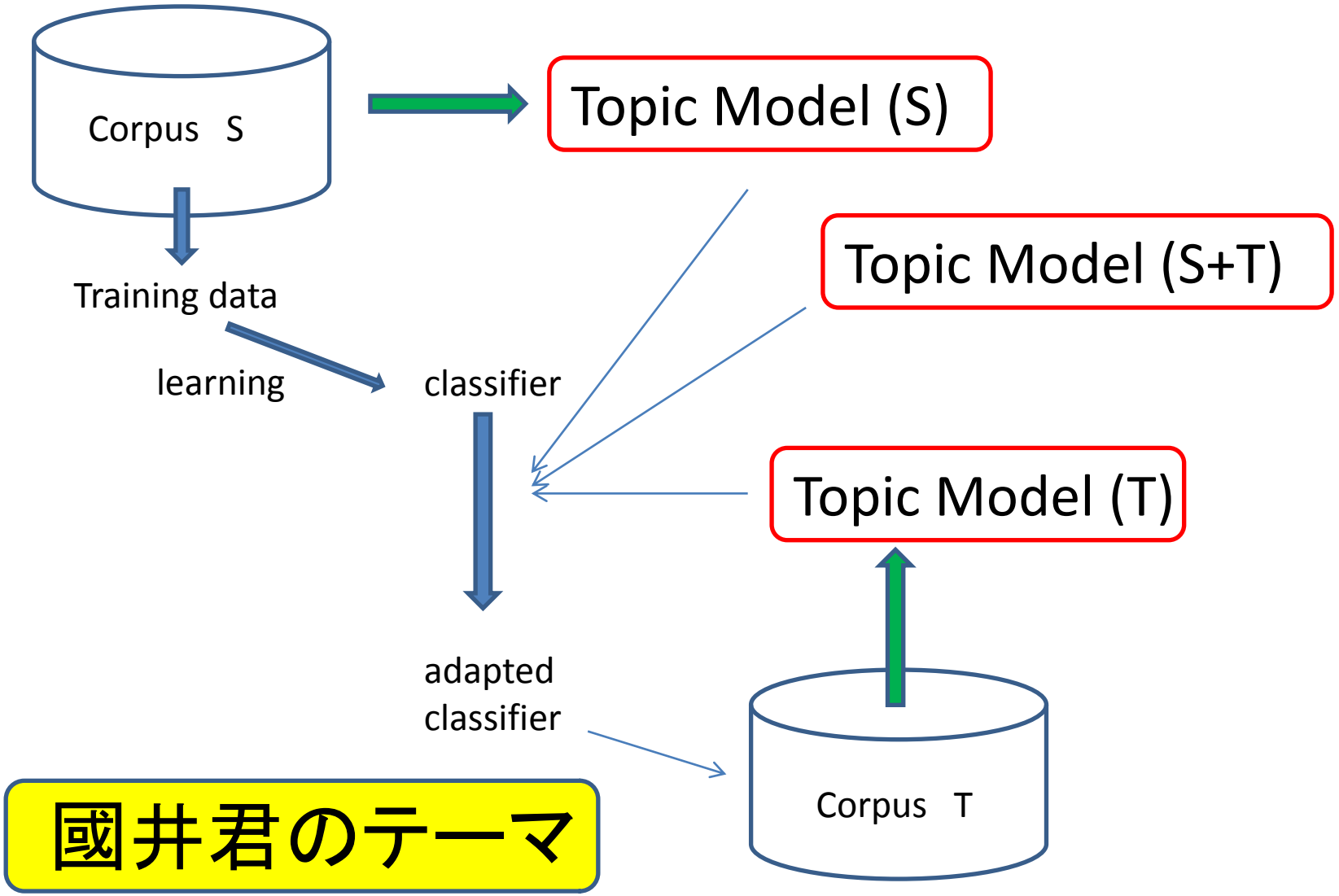
Topic Model

Unsupervised approach



My Theme 1

How do we use Topic model of Domain in WSD ?



國井君のテーマ

Daume III

http://www.umiacs.umd.edu/~hal/

My Theme 2

Distance between Source Domain and Target Domain

DA for WSD is very different from other DA tasks.

The difficulty depends on not a domain but a target word.

related

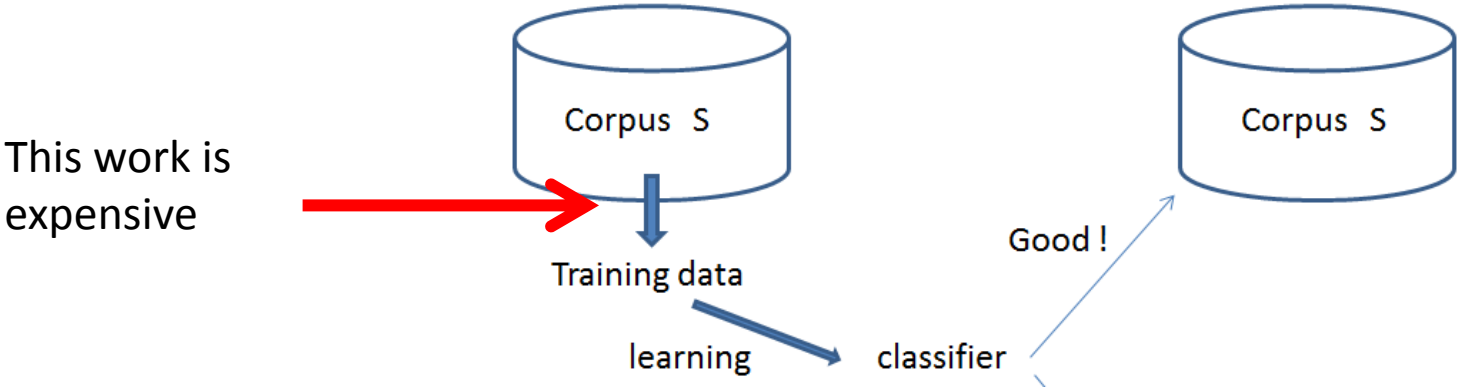
As to this difficulty, we should change the method.

How do we measure that distance?

吉田君のテーマ

Active Learning

http://hunch.net/~active_learning/



Select effective data for learning automatically

小野寺君のテーマ

Your research titles

國井：領域別トピック素性の混合利用による
語義曖昧性解消

菊池：共変量シフトの問題として見た
語義曖昧性解消の領域適応

小野寺：語義曖昧性解消の領域適応ための
密度比による能動学習（仮）

吉田：語義曖昧性解消の領域適応を対象とした
単語別領域間距離測定（仮）