Triplex Transfer Learning: Exploiting both Shared and Distinct Concepts for Text Classification

Fuzhen Zhuang, Ping Luo, Changying Du,
Qing He, Zhongzhi Shi

中国科学院计算技术研究所
Institute of Computing Technology,Chinese Academy of Sciences

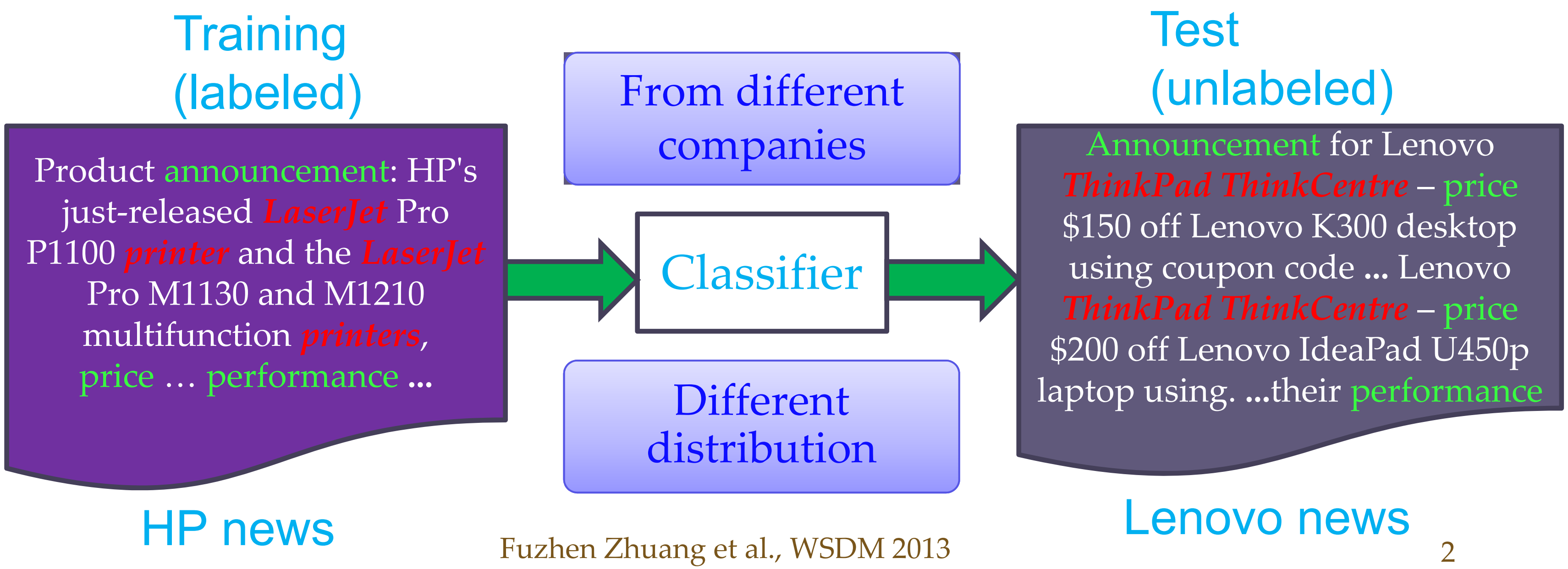
LABS hp

Introduction

- Many traditional learning techniques work well only under the assumption: Training and test data follow the same distribution

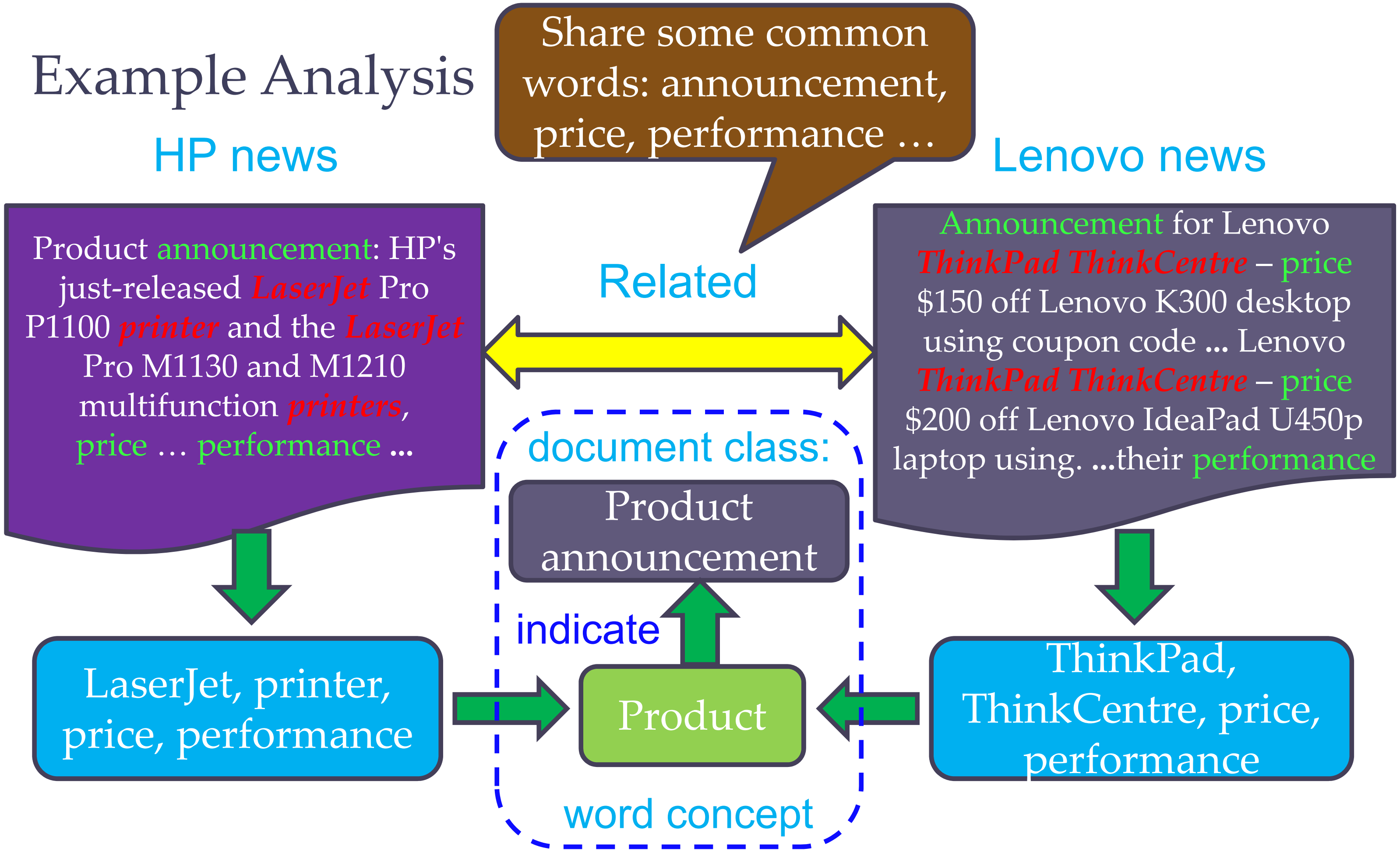
Fail !

Enterprise News Classification: including the classes "Product Announcement", "Business scandal", "Acquisition", … …



Motivation (1)

- Example Analysis



The classification based on concepts may be more appropriate for transfer learning

Motivation (2)

- Observations:
  - Different domains may use same key words to express the same concept (denoted as *identical concept*)
  - Different domains may also use different key words to express the same concept (denoted as *alike concept*)
  - Different domains may also have their own distinct concepts (denoted as *distinct concept*)
- The identical and alike concepts are used as the shared concepts for knowledge transfer learning
- In this study, we try to model these three kinds of concepts simultaneously for transfer learning text classification

Preliminary Knowledge

- Basic formula of matrix tri-factorization:

$$X_{m\times n} = F_{m\times k} S_{k\times c} G_{n\times c}^{\top}$$

where the input X is the word-document co-occurrence matrix

- **F** denotes concept information
- **S** the association between word concepts and document classes
- **G** denotes the document classification information

- Further divide the word concepts into three kinds:

$$X_{m\times n} = F_{m\times k} S_{k\times c} G_{n\times c}^{T}$$
$$= [F^1_{m\times k_1}, F^2_{m\times k_2}, F^3_{m\times k_3}] \begin{bmatrix} S^1_{k_1\times c} \\ S^2_{k_2\times c} \\ S^3_{k_3\times c} \end{bmatrix} G_{n\times c}^{\top}$$

F^1, identical concepts; F^2, alike concepts; F^3, distinct concepts

Problem Formulation

- Input: s source domain $X_r(1\leq r\leq s)$ with label information, t target domain X_r $(s+1\leq r\leq s+t)$
- Triplex transfer learning framework based on matrix tri-factorization (TriTL for short)

$$\min_{F_r,S_r,G_r} \mathcal{L} = \sum_{r=1}^{s+t} ||X_r - F_r S_r G_r^\top||^2$$

$$= \sum_{r=1}^{s+t} ||X_r - [F^1, {F^2}_r, {F^3}_r] \begin{bmatrix} S^1 \\ S^2 \\ {S^3}_r \end{bmatrix} G_r^\top||^2$$

$$s.t. \sum_{i=1}^{m} {F^1}_{[i,j]} = 1, \sum_{i=1}^{m} {F^2}_{r[i,j]} = 1,$$

$$\sum_{i=1}^{m} {F^3}_{r[i,j]} = 1, \sum_{j=1}^{c} G_{r[i,j]} = 1.$$

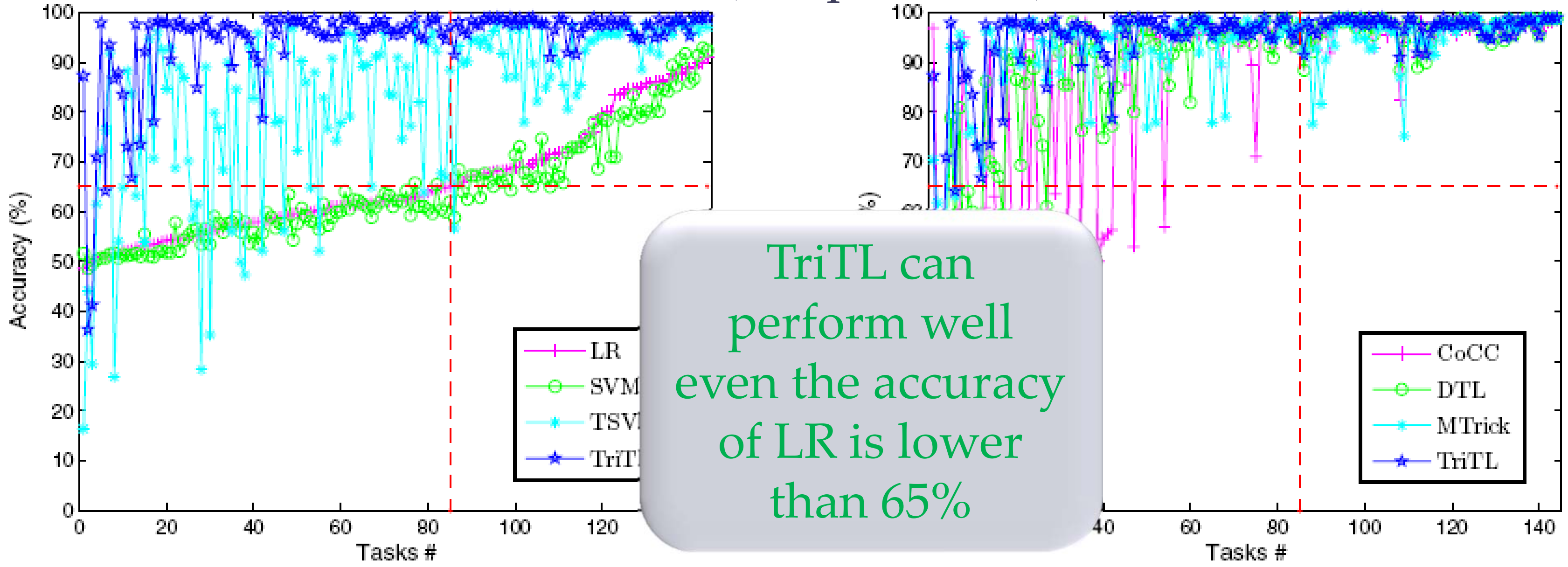
F^1, S^1 and S^2 are shared as the bridge for knowledge transfer across domains

The supervision information is integrated by G_r $(1\leq r\leq s)$ in source domains

We develop an alternatively iterative algorithm to derive the solution and theoretically analyze its convergence

Experimental Results

- Comparisons among TriTL, DTL, MTrick, CoCC, TSVM, SVM and LR on data set *rec vs. sci* (144 problems)



TriTL can perform well even the accuracy of LR is lower than 65%

Data Set		LR	SVM	TSVM	CoCC	DTL	MTrick	TriTL
rec vs. sci	*Lower*	57.41	56.78	75.73	79.69	84.29	90.44	**92.23**
	Higher	75.77	73.48	91.66	96.18	96.56	95.53	**97.19**
	Total	65.57	64.20	82.81	87.02	89.75	92.70	**94.43**

Conclusions

- Explicitly define three kinds of word concepts, i.e., identical concept, alike concept and distinct concept
- Propose a general transfer learning framework based on nonnegative matrix tri-factorization, which simultaneously model the three kinds of concepts (TriTL)
- Extensive experiments show the effectiveness of the proposed approach, especially when the distinct concepts may exist

Thank you!

Q. & A.

Fuzhen Zhuang' Homepage:
http://www.intsci.ac.cn/users/zhuangfuzhen/index.htm