The 12 China Workshop on Machine Learning and Applications
MLA 2014, Xi'an
Nov. 8, 2014

Short Text Conversation Using Big Data and Deep Learning

Hang Li
Noah's Ark Lab
Huawei Technologies

Short Text Conversation

Human inputs a message
Computer returns a response

Alan Turing

One small step toward the goal of passing Turing test

Demo: Short Text Conversation


上海今天好熱，堪比新加坡。


上海今天热的不一般。


想去武当山 有想同游的么？


我想跟帅哥同游~哈哈

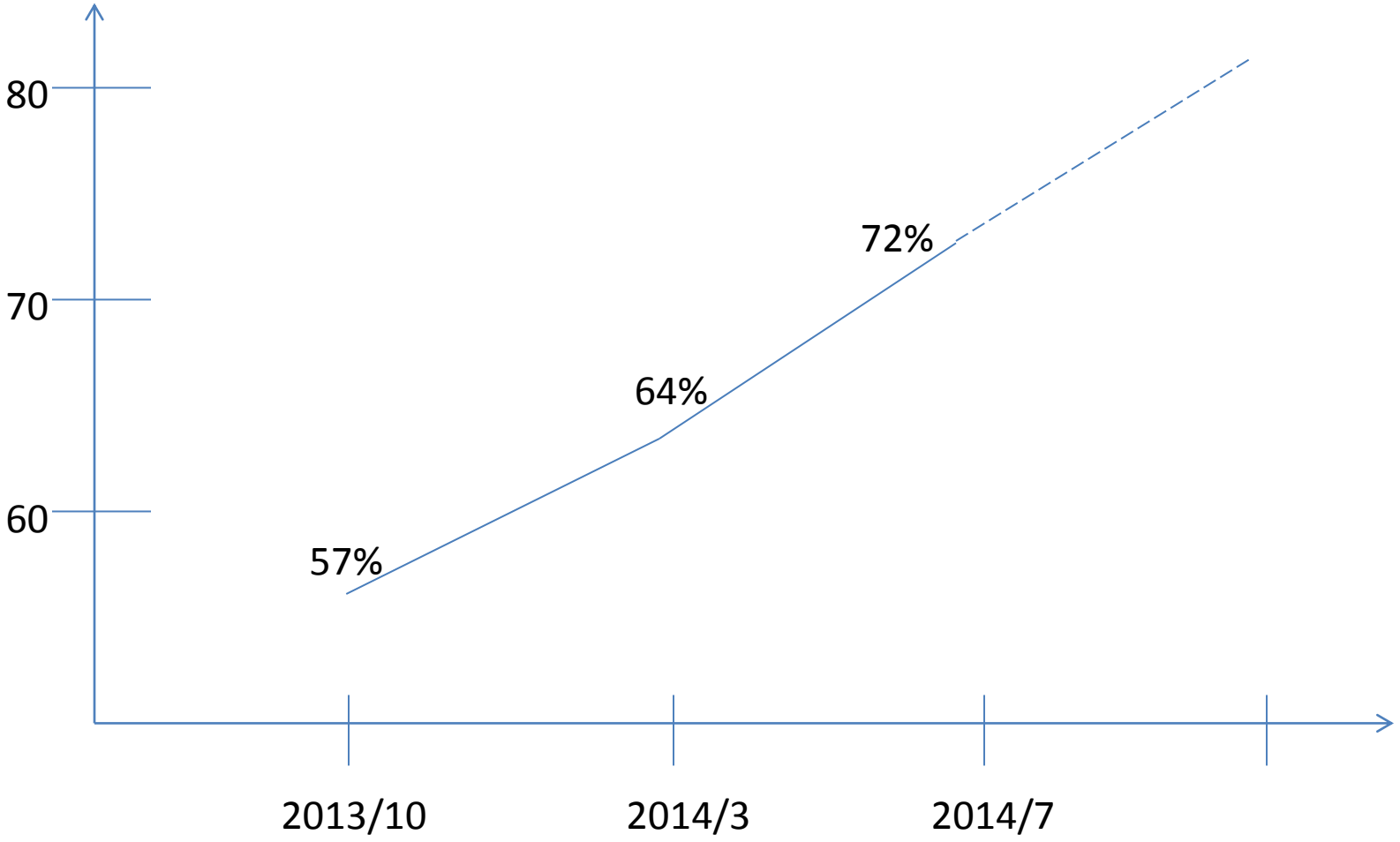

在家门口发现了一片野韭菜！


我看到一片一片的韭菜地.....


芝大综合排名升的好快呀，都第五了


芝大威武

Accuracy Improvement at Noah's Ark Lab



Outline

- Project: Intelligent Information Assistant
- Short Text Conversation
- Retrieval-based Short Text Conversation
- Learning to Match
- Deep Matching Models
- Summary

Project: Intelligent Information Assistant

Intelligent Information Assistant

External Information

Communications

Information Recommendation

Information Extraction

Information Management

Question Answering

Personal Information

Knowledge Base

Natural Language Conversation

Human Computer Interaction

Intelligent Help in Huawei Phones

手机服务首页

智能问答栏目

输入自动联想

Weibo Robot: 小诺_noah

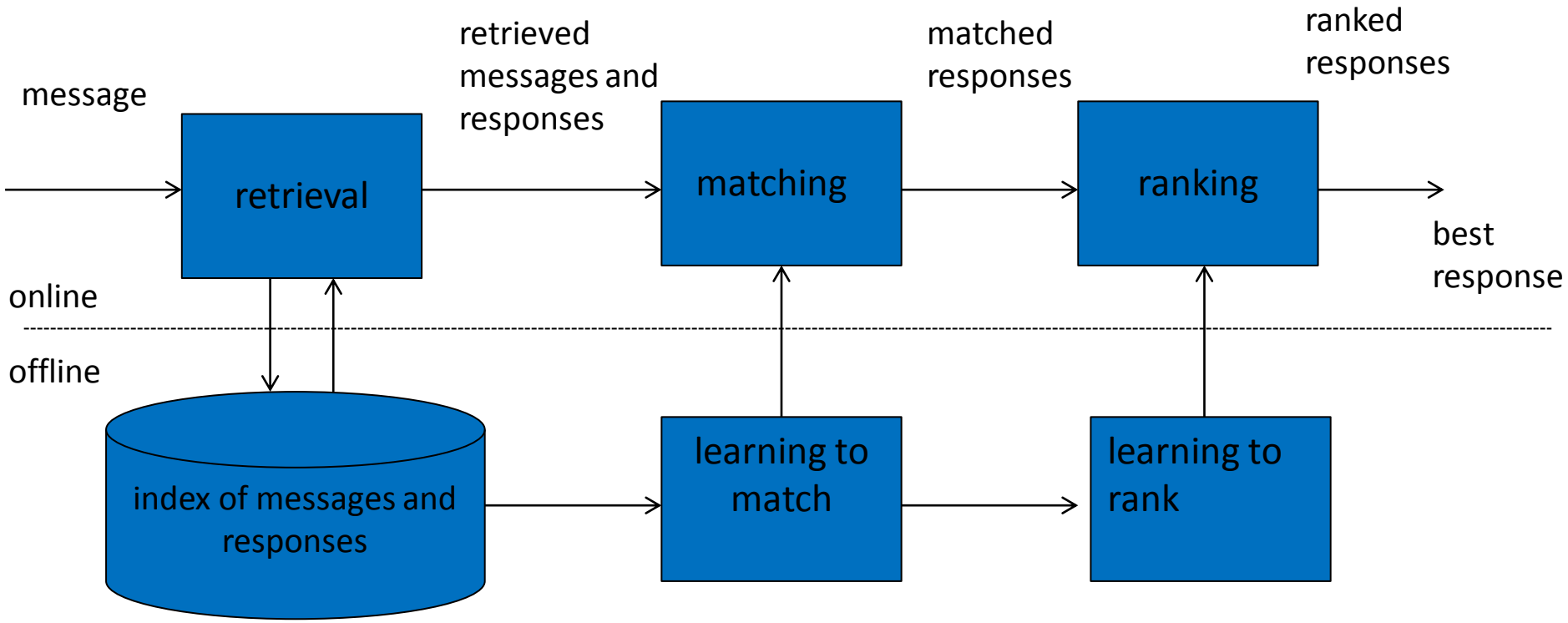
- Xiaonuo (Weibo robot version 1.0) released on Jan 1, 2013
- Persona: PhD student in NLP
- Number of followees = 523
- Number of followers = 1096
- Features Developed
  - Following people
  - Re-tweeting (forwarding tweets)
  - Generating short comments

Retrieval-based Short Text Conversation

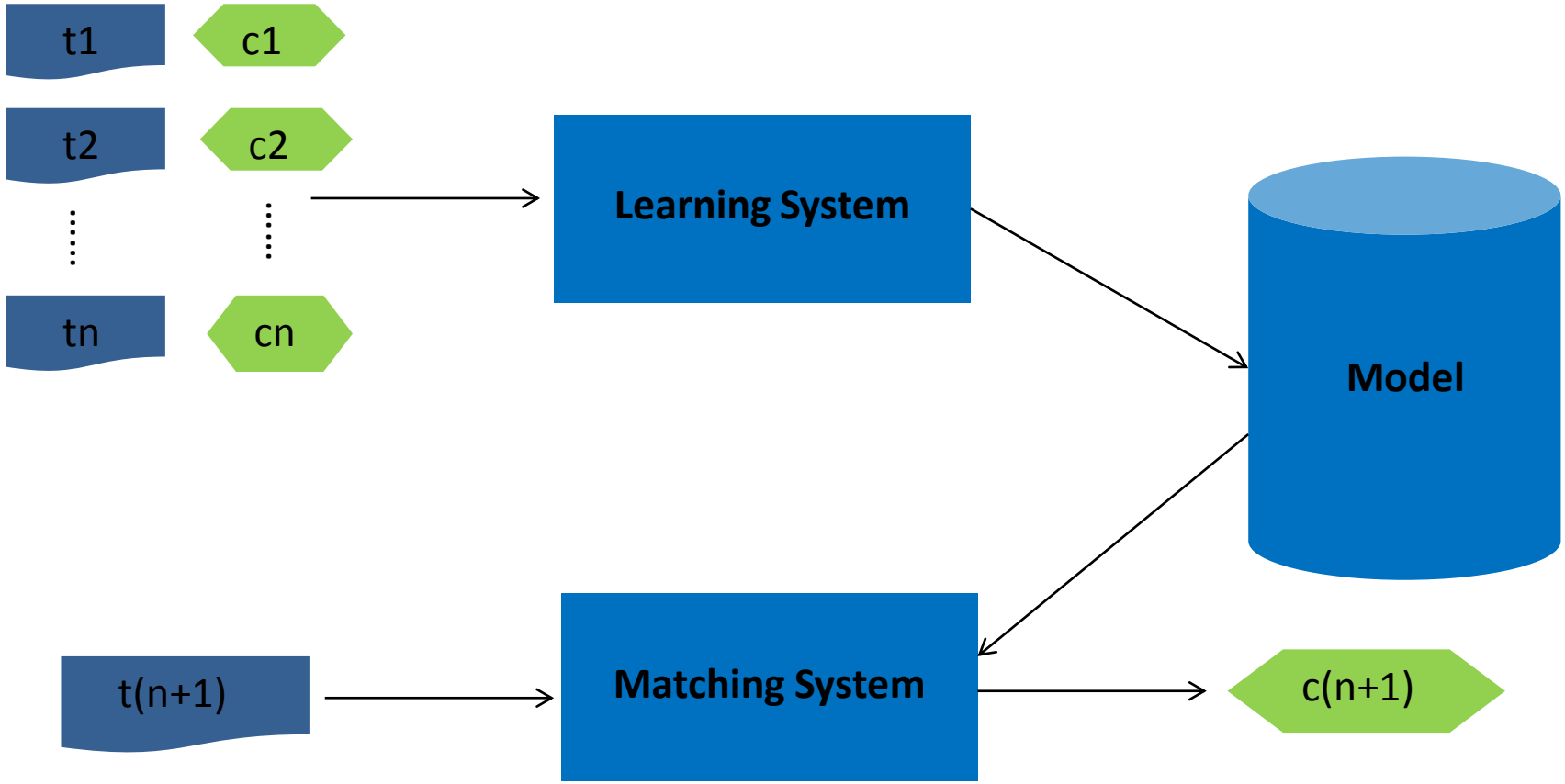
Massive Amount of Data Available

System of Short Text Conversation

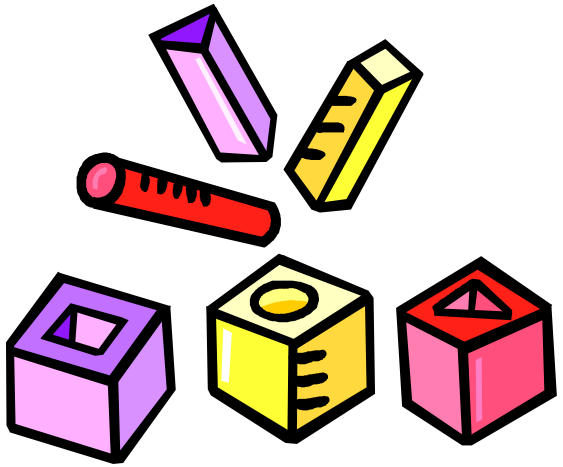
- Given message, find most suitable response
- Large repository of message-response pairs
- Take it as search problem



Learning to Match for Short Text Conversation



Learning to Match



Matching between Heterogeneous Data is Everywhere

- Matching between user and product (collaborative filtering)
- Matching between text and image (image annotation)
- Matching between languages (machine translation)
- Matching between receptor and ligand (drug design)
- Matching between people (dating)

Formulation of Learning Problem

- Learning matching function

$$f(x, y)$$

- Training data $(x_1, y_1, r_1), \cdots, (x_N, y_N, r_N)$
- Generated according to

$$x \sim P(X), \quad y \sim P(Y \mid X), \quad r \sim P(R \mid X, Y)$$

Formulation of Learning Problem

- Loss Function

$$L(r, f(x, y))$$

- Risk Function

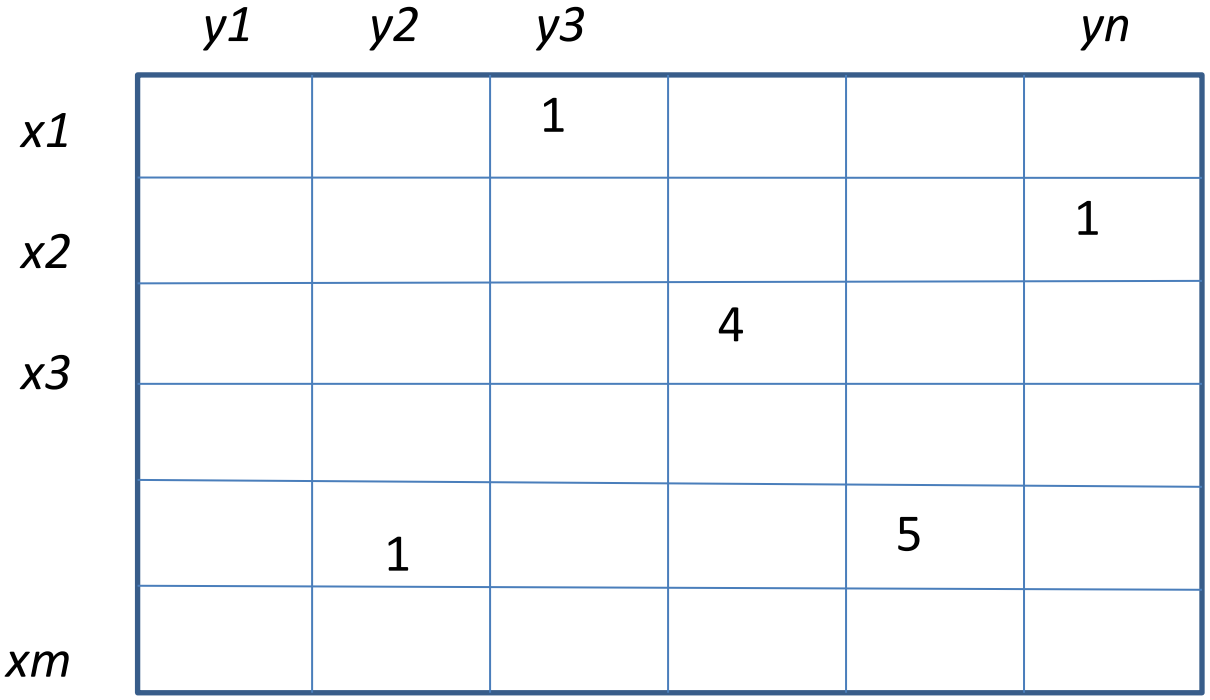
$$R(r, f(x, y)) = \int_{X \times Y \times R} P(x, y, r) L(r, f(x, y)) dP(x, y, r)$$

- Objective Function in Learning

$$\min_{f \in F} \sum_{i=1}^{N} L(r_i, f(x_i, y_i)) + \Omega(f)$$

Matching Problem: Instance Matching

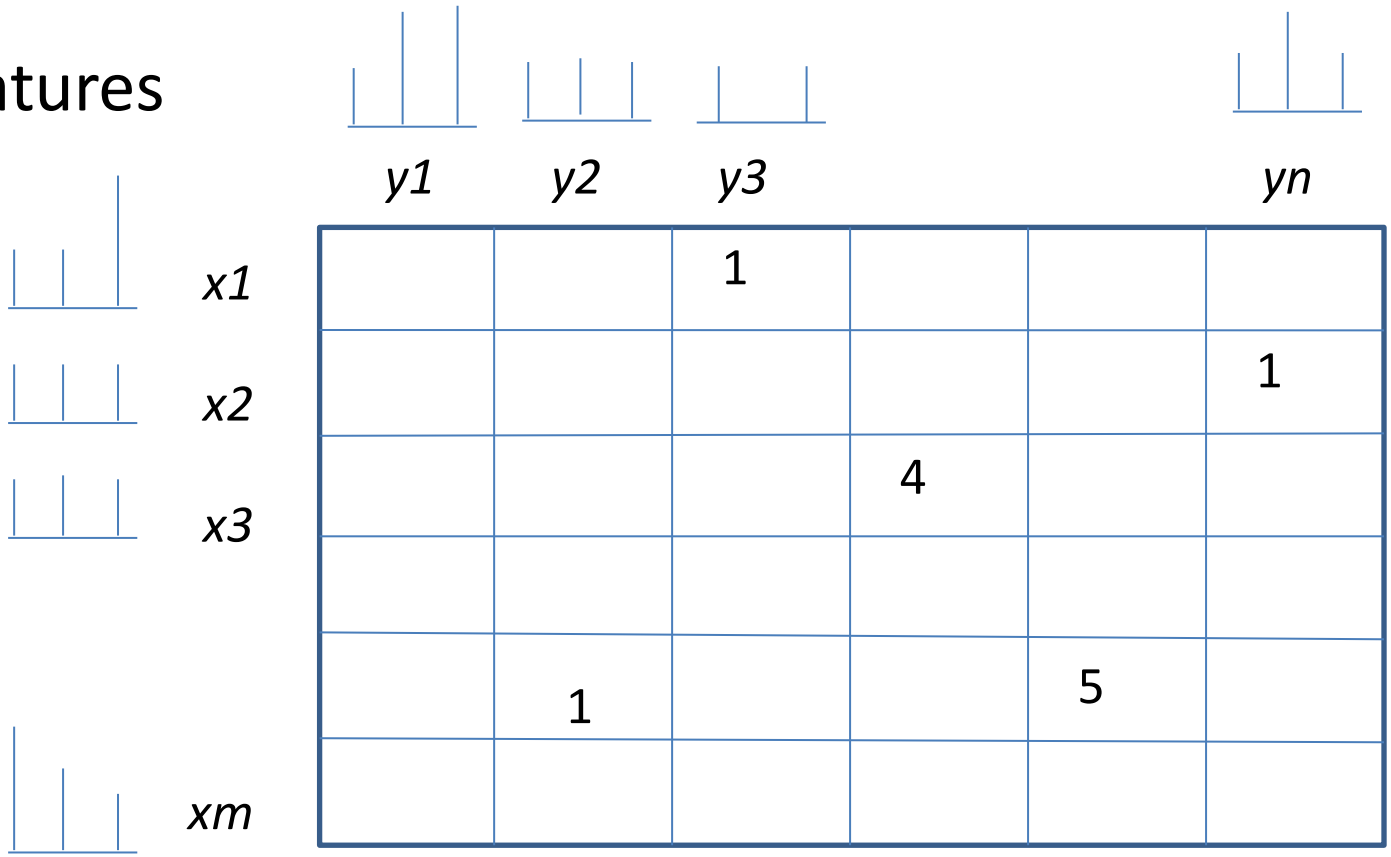
Instances



Can be represented as matching between nodes in bipartite graph

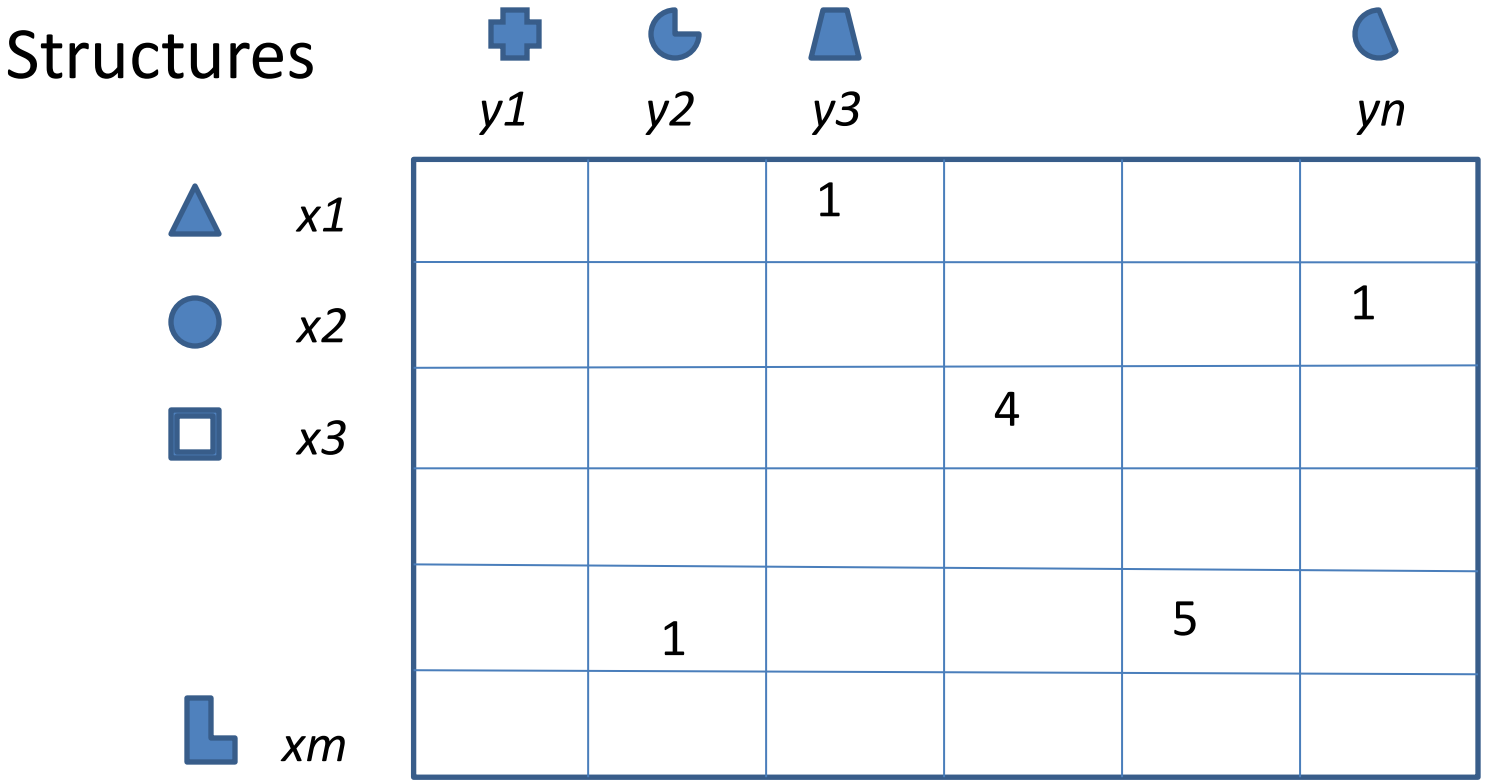
Matching Problem: Feature Matching

Features

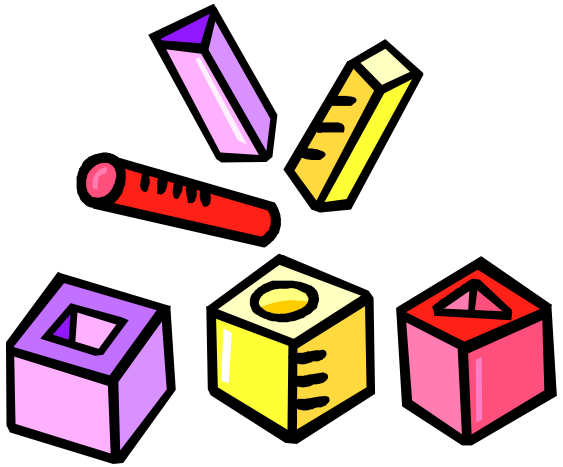


Can be represented as matching between objects in two spaces

Matching Problem: Structure Matching



Deep Matching Models



Collaborators

Zhengdong Lu

Baotian Hu

Mingxuan Wang

Qun Liu

Qingcai Chen

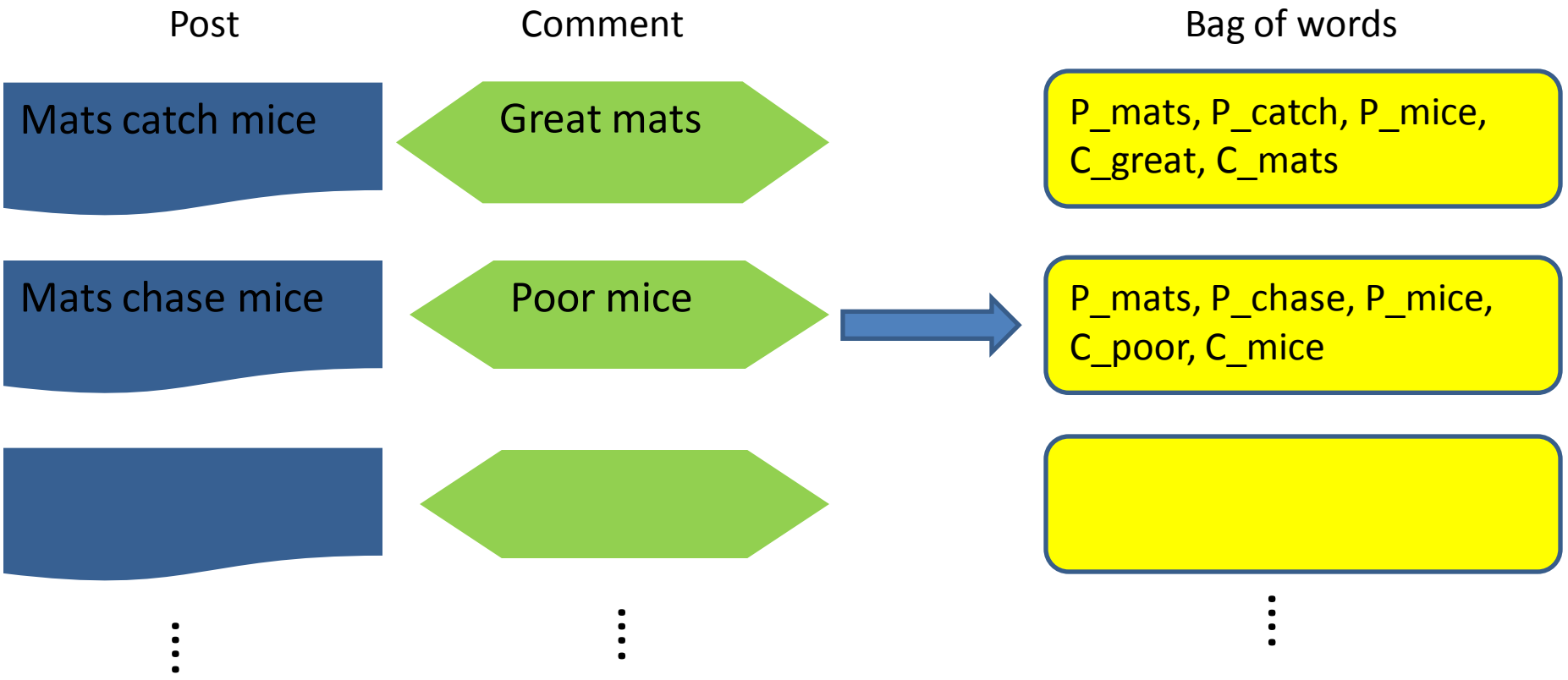
Model: Deep Match Topic

Model: Deep Match Topic

(Lu & Li, NIPS 2013)

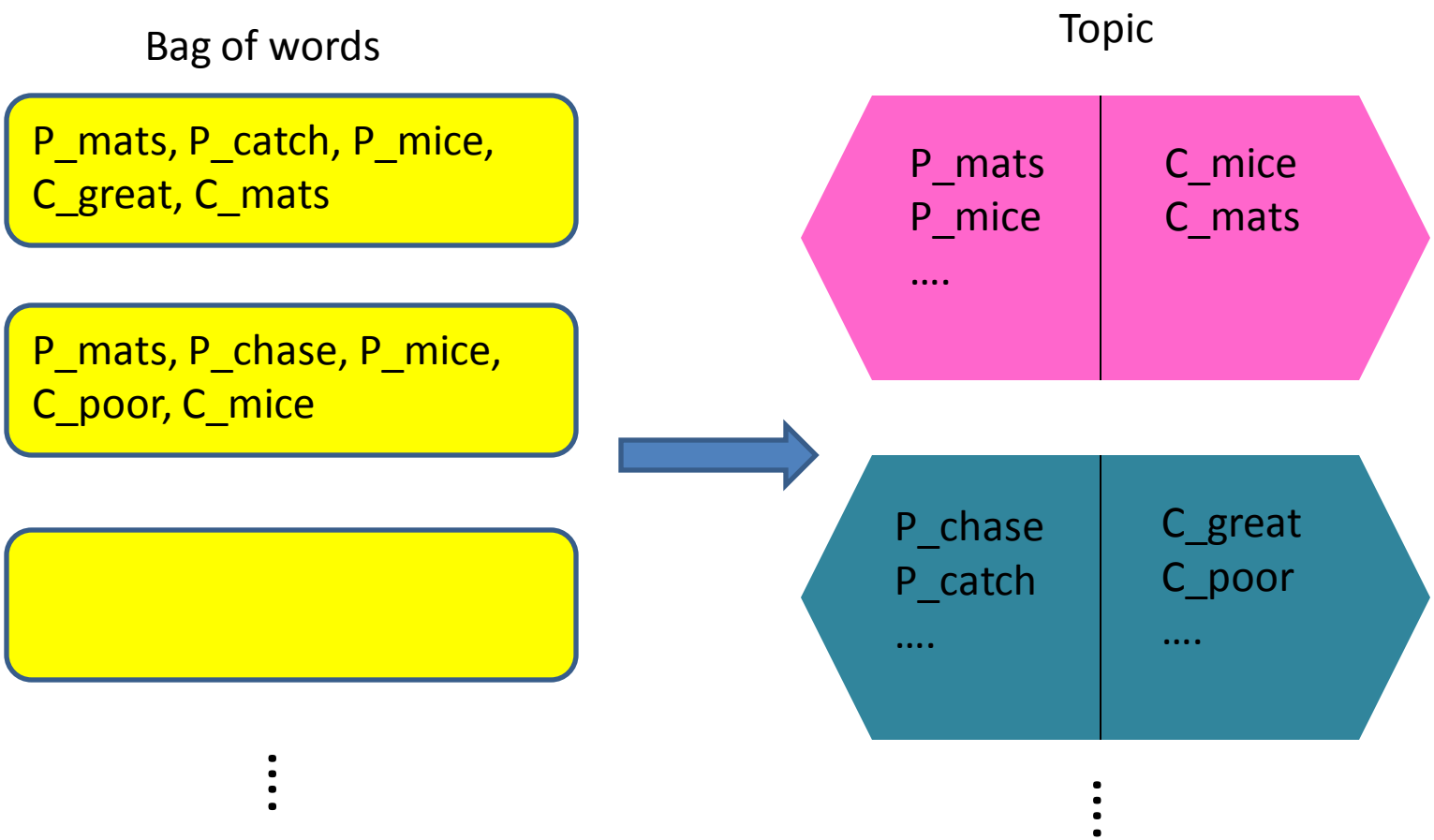
- Taking pairs of texts as input
- Learning topics of words in different granularities using LDA
- Taking topics in different granularities as hidden layers
- Constructing neural network using heuristics
- Learning parameters of neural network using back-propagation

Representing Posts and Comments as Bags of Words



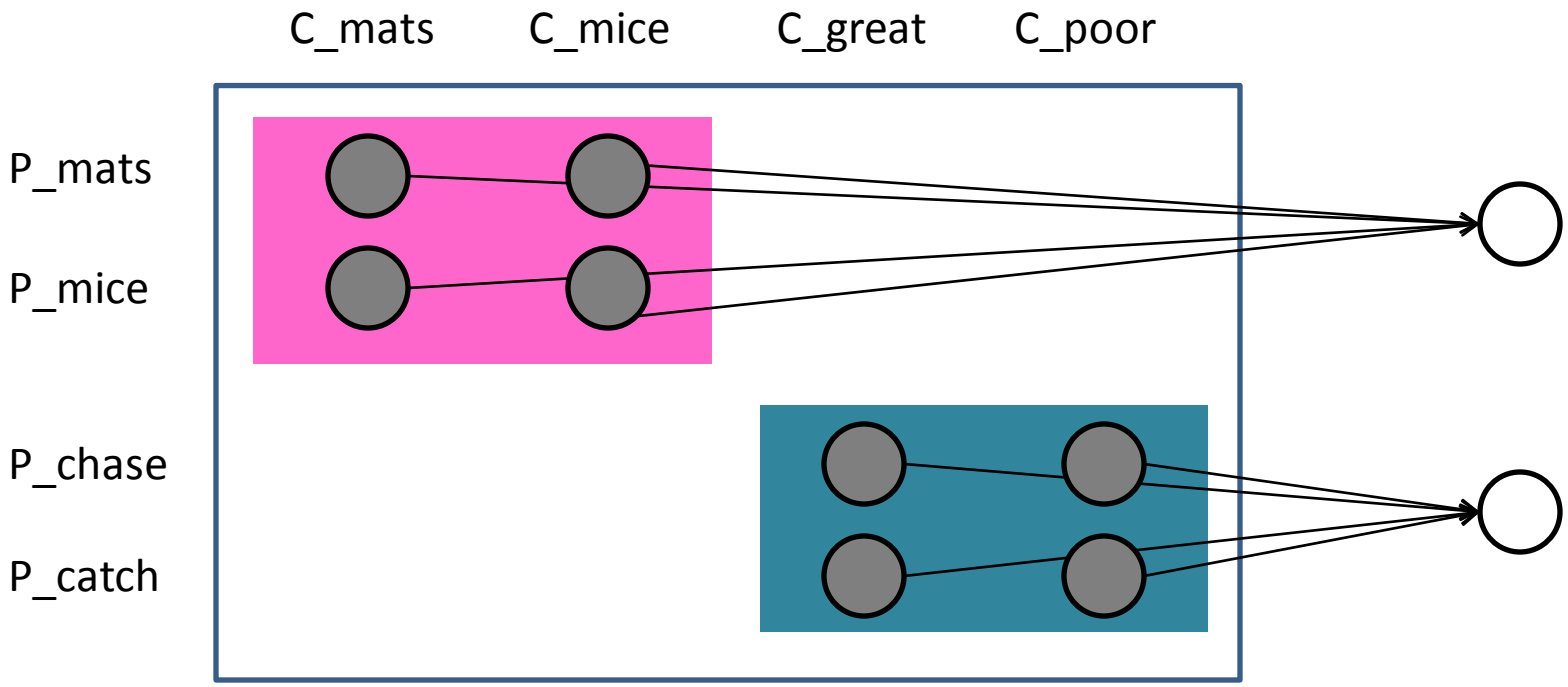
Words in post and comment are viewed as different words

Constructing Topics Using Latent Dirichlet Allocation



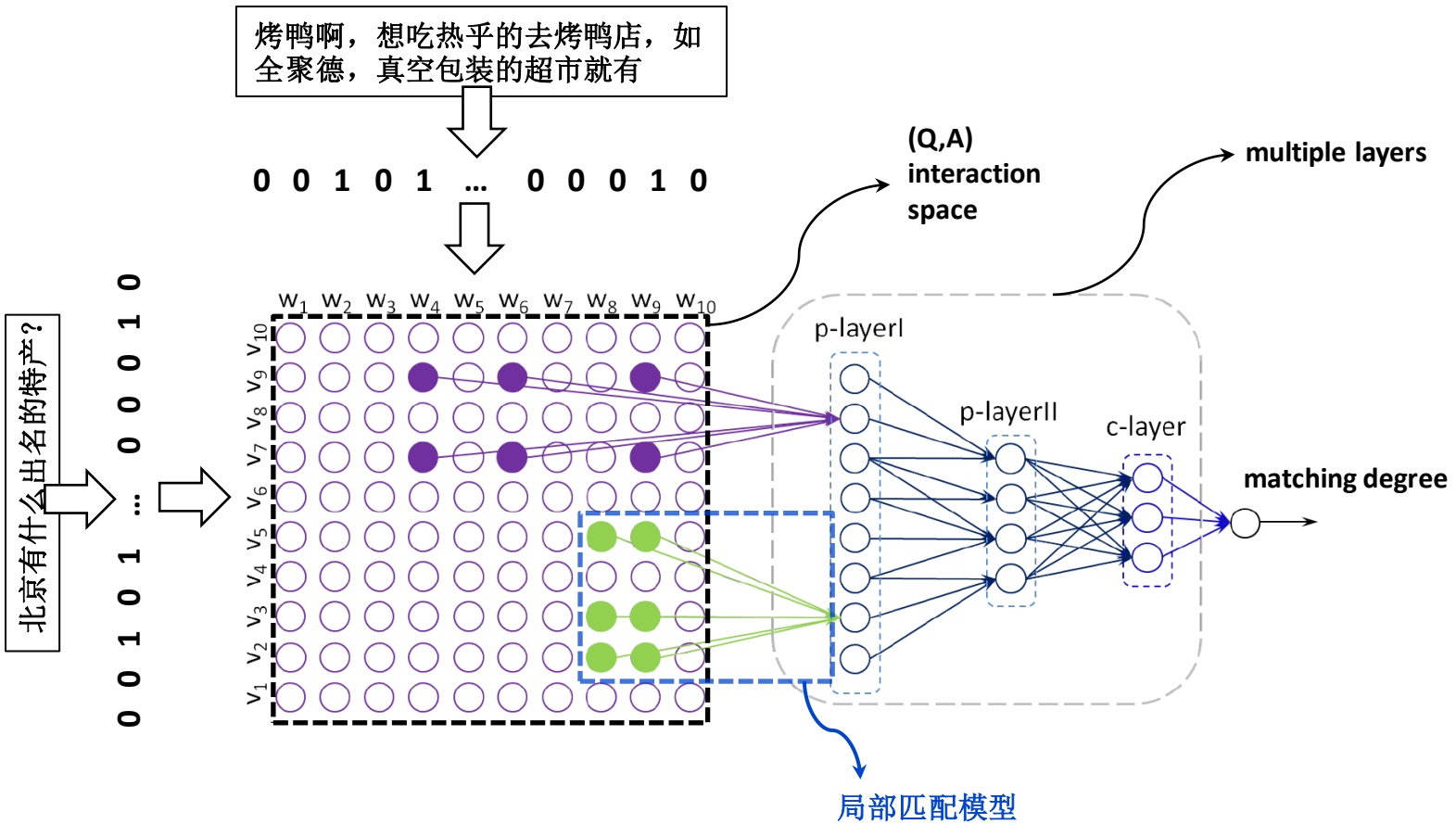
Construct Neural Network Using Heuristics

Four paired words in pink are connected to a hidden node



Four paired words in blue are connected to another hidden node

Architecture of Deep Matching Model



Examples

Local Model 1: (特产, 土产, 味道, …) || （豆腐, 烤鸭, 甜, 野味, 糯米…）

Local Model 2: (路程, 安排, 地点, …) || （距离, 安全, 隧道, 高速, 机票…）

Experimental Results

- Data: 12,000 post-comment pairs from Weibo
- Cross validation in terms of P@1
- Conclusion: Deep Match works better than linear model

Model	MAP	P@1
P2R	0.565	0.489
P2R + P2P	0.621	0.567
P2R + MATCH	0.575	0.513
P2R + P2P + MATCH	0.621	0.574

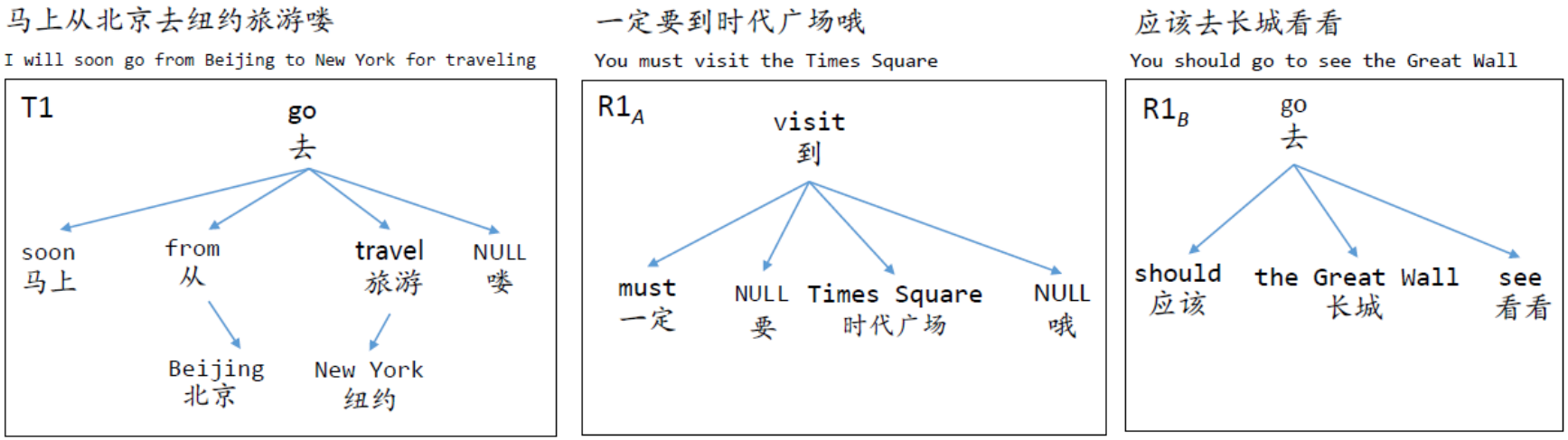
Model: Deep Match Tree

Model: Deep Match Tree
(Wang, Lu, Li, & Liu; to appear)

- Taking pairs of dependency trees as input
- Mining frequent matching patterns from pairs of dependency trees
- Taking matching patterns as input layer of neural network
- Training weights of neural network using back propagation

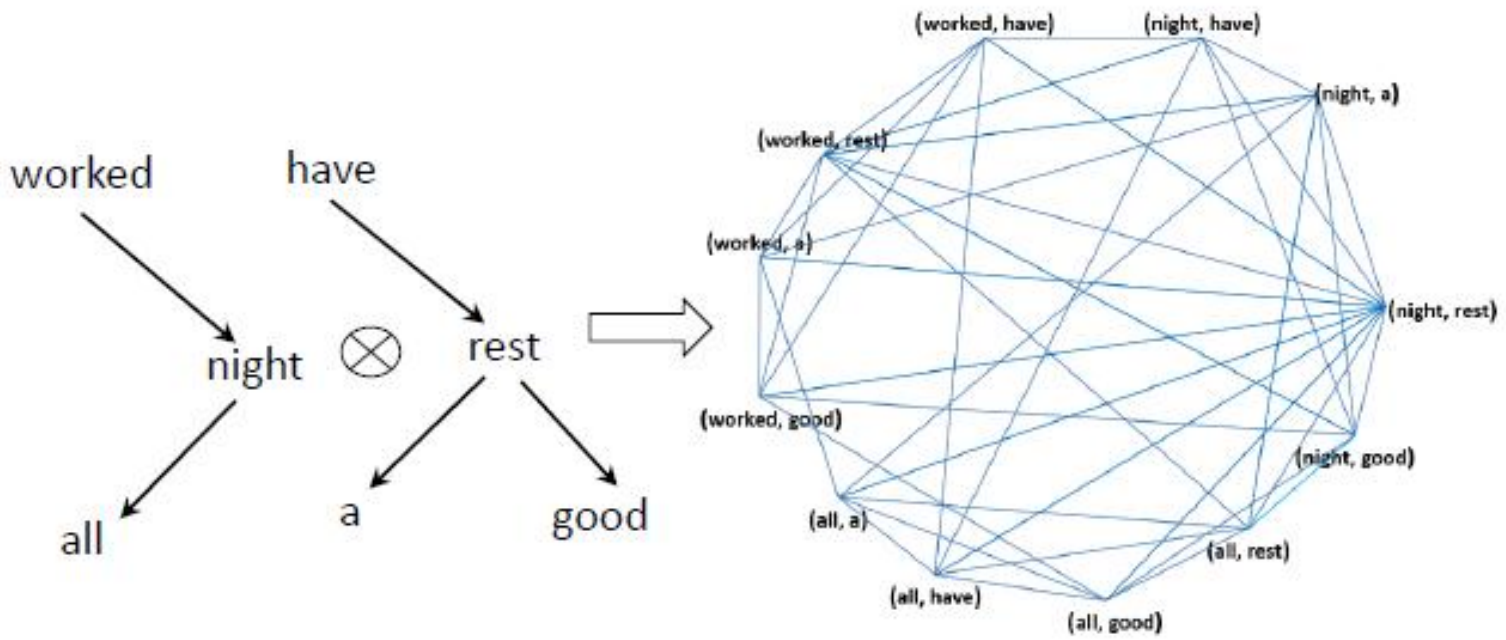
Representing Sentence with Its Dependency Tree

Lexical and syntactic information of sentence is represented in its dependency tree



Constructing Product of Trees

- Use two dependency-trees to create product of graph
- Represent interaction between two sentences
- A sub-graph represents a matching pattern
- Find high frequency patterns using pattern mining technique



Large Scale Graph Mining

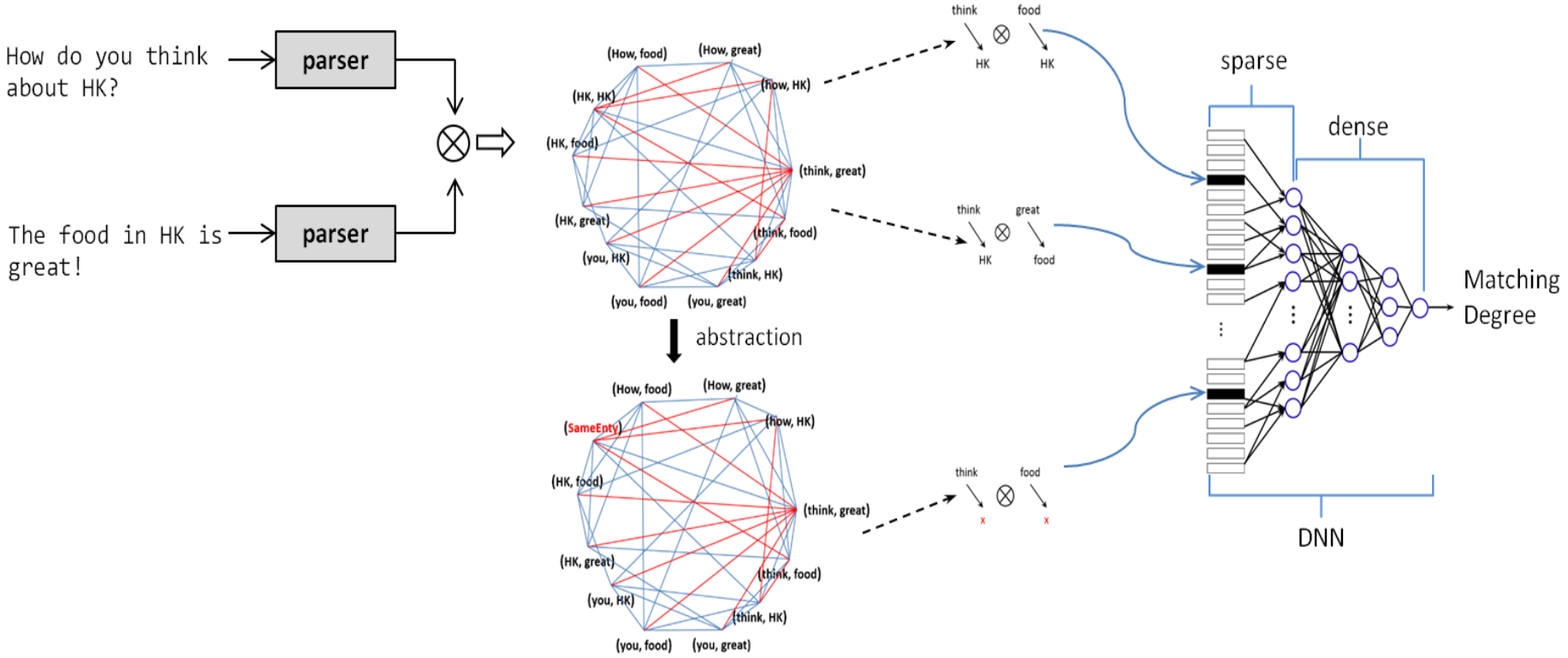
- Large-scale graph mining for finding high frequency sub-graphs (matching patterns)
- Lexical and syntactic information is incorporated in the patterns

Patterns without abstractions
exam ⊗ score
Information theory ⊗ Shannon
thank→present ⊗ happy→birthday
win→game ⊗ trying→keep
out-of-control→prices ⊗ regulation
work→weekend ⊗ rest

Patterns without abstractions
hope→win→x ⊗ support→x
how about→x ⊗ like→x
gift→x ⊗ happy→x
recommend→x ⊗ x→nice
pretty good→x ⊗ fine→also→x'

Deep Match Tree

- Constructing deep neural network, with first layer corresponding mined patterns



Experimental Results

- Retrieval-based Conversation
- 5 million post-comment pairs for mining of patterns
- Data: 12,000 labeled post-comment pairs from Weibo
- Cross validation in terms of P@1
- Conclusion: Deep Match Tree outperforms baselines

Model	P@1
BASELINE	0.574
+DEEPMATCH	0.587
+WORDEMBED	0.579
+TRANSLATION	0.585
+DEEPMATCH$_{tree}$	0.608

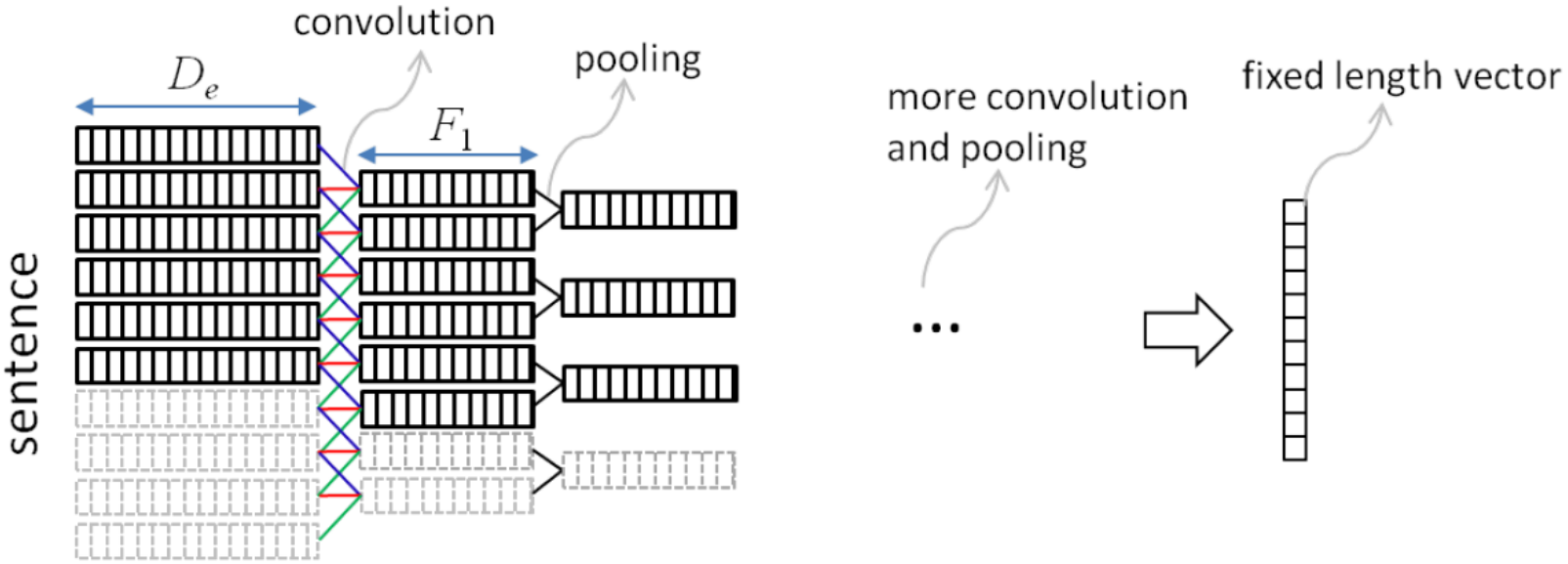
Model: CNN Match

Model: CNN Match
(Hu, Lu, Li, & Chen; NIPS 2014)

- Taking pairs of sentences as input
- Representing content of sentences and matching of sentences using Convolutional Neural Network
- No linguistic knowledge is needed

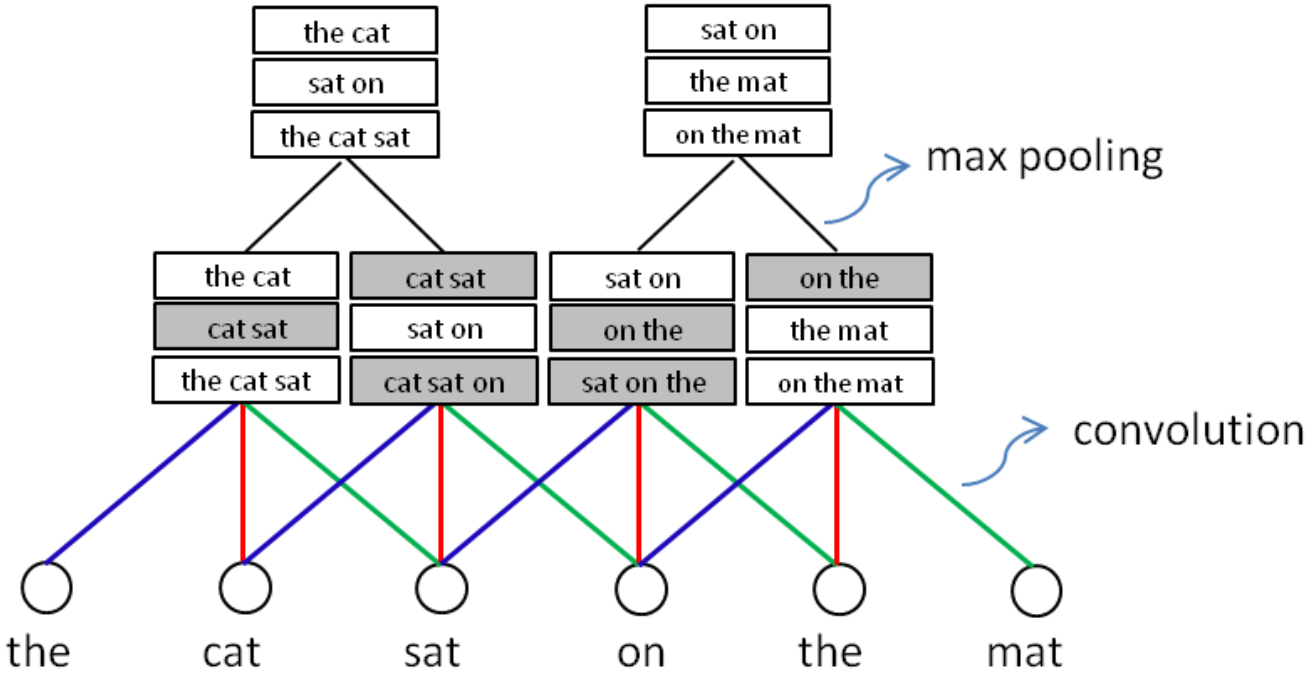
Sentence Model Using Convolutional Neural Network (CNN)

- Representing content of sentence using CNN



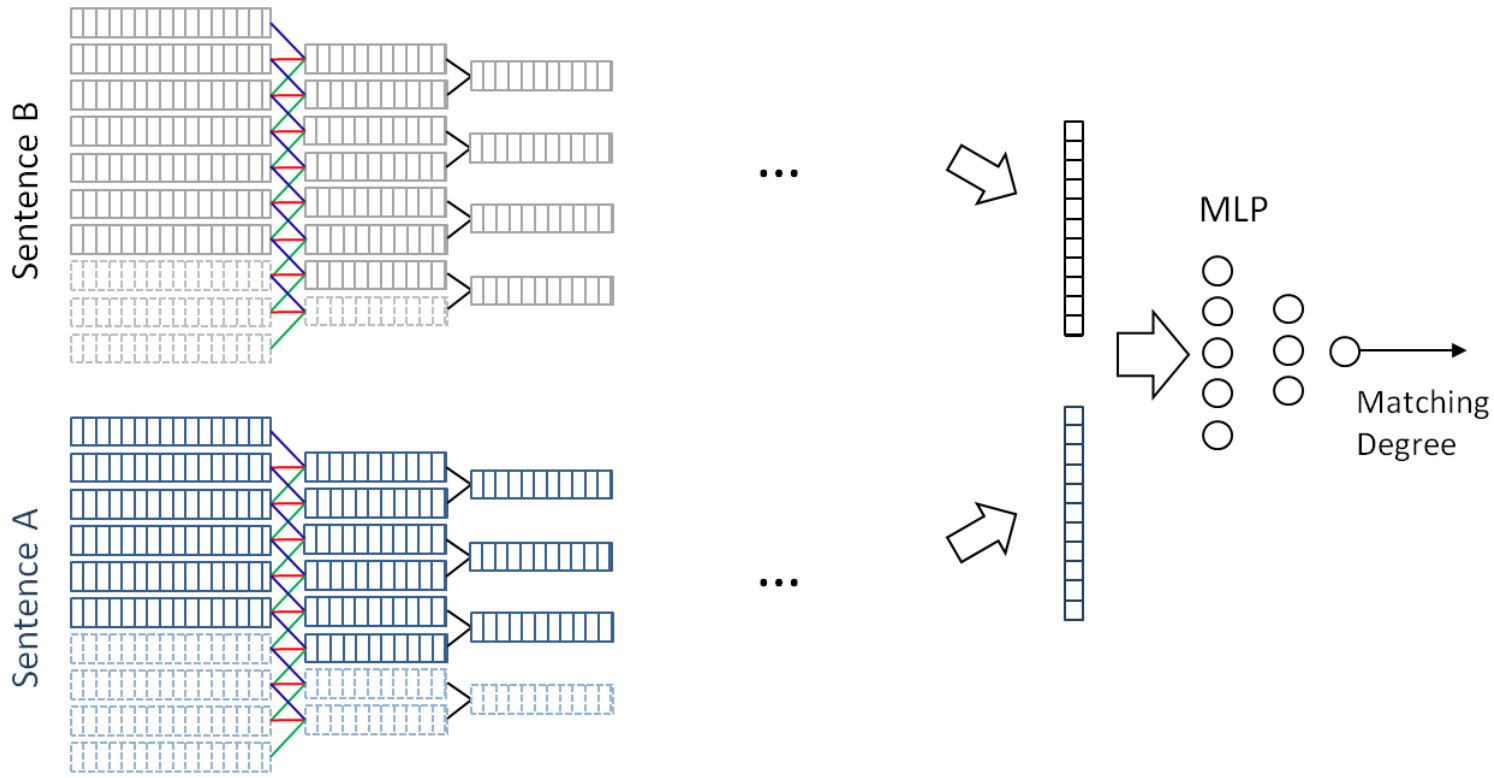
Advantage of Using CNN

- Sliding windows: possible sequences of words for composition
- Convolution: composition of words
- Pooling: selection of word sequences for composition



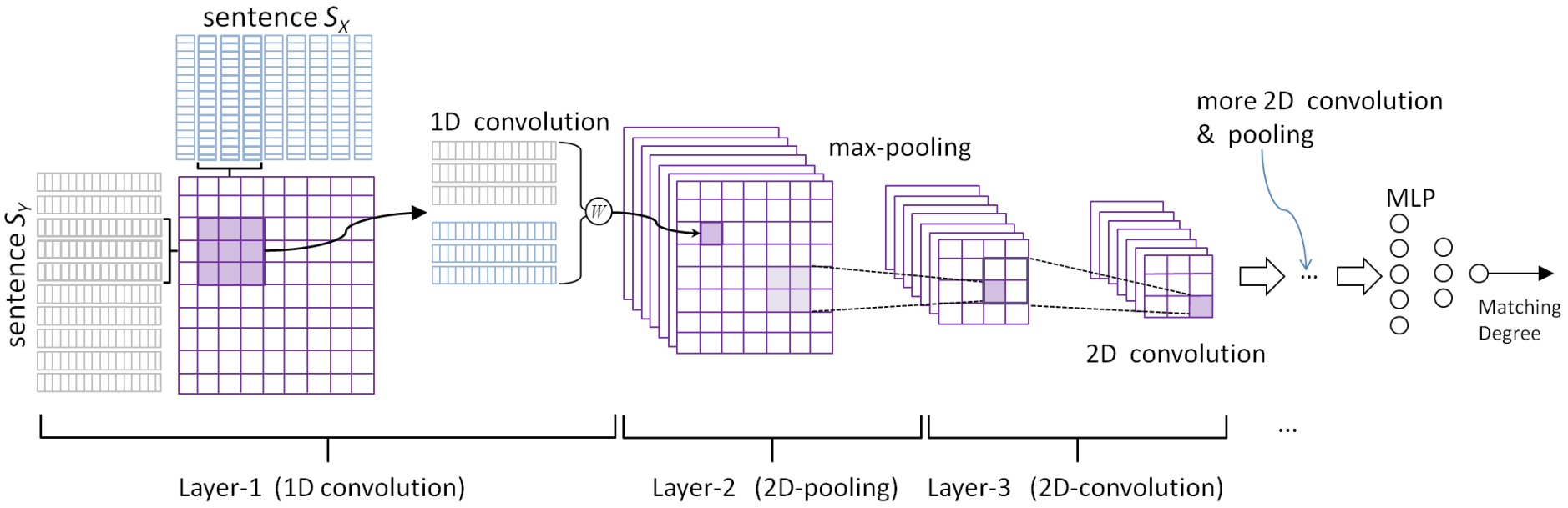
CNN Match Arc I:

- First represent two sentences, and then match



CNN Match Arc II:

- Represent and match two sentences simultaneously
- Two dimensional convolution and pooling



Advantages of CNN Match

- Order of words in two sentences are considered in the model
- Structures of two sentences can be captured
- Shared parameters improve efficiency in training
- Arc-II takes Arc-I as special case

Experimental Results

- Retrieval-based conversation
- Training data: 4 million pairs
- Testing data: 450*k* pairs
- Conclusion: Arc-II > Arc-I > baselines

Model	P@1(%)
Random Guess	20.00
DEEPMATCH	49.85
WORDEMBED	54,31
SENMLP	52.22
SENNA+MLP	56.48
ARC-I	59.18
ARC-II	**61.95**

Summary

Summary

- Noah's Ark Lab is working on intelligent information assistant
- Short Text Conversation (STC) is challenging yet interesting task
- Current approach = retrieval-based STC
- Learning to match is fundamental problem
- Big data and deep learning are powerful tools
- Several deep matching models proposed

Thank You!

hangli.hl@huawei.com