

Artificial Intelligence

# **Project 2 : Language Learning Using Hypernetworks**

2009. 5. 4.

Ko, Younggil ([ykko@bi.snu.ac.kr](mailto:ykko@bi.snu.ac.kr))

Kim, Seungyeon ([sykim@bi.snu.ac.kr](mailto:sykim@bi.snu.ac.kr))

Biointelligence laboratory

# Contents

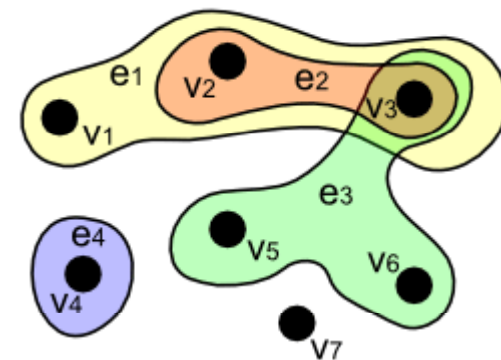
- Outline
- Language Learning using Hypernetworks
- Data set for Language Learning
- Tool for Hypernetworks
- Guide to Writing Reports
  - Style, mandatory contents, optional contents
- Submission guide / Marking scheme
- Demo on the Tool

# Outline

- Goal
  - Understand Hypernetworks & machine learning deeper
  - Practice research and technical writing
- Language Learning (sentence completion)
  - The problem is to predict each word in a sentence based on surrounding words
  - Consider each word as an attribute and train a Hypernetwork with sentences
- Data Set
  - Sentences from 'Friends' and BBC documentary

# Brief Introduction to Hypernetworks

- Hypernetworks
  - Representation and learning method based on weighted hypergraph
    - <http://en.wikipedia.org/wiki/Hypergraph>
  - Generate hyperedges with sampling and manage a library of weighted hyperedges
  - Learning strategy
    - (explained in pp. 5~8)



- Reference: [IEEE-CIM, 2008] B.-T. Zhang, Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory, *IEEE Computational Intelligence Magazine*, 3(3):49-63, 2008.

# Language Learning with Hypernetworks (1/8)

- Sentence Completion
  - We want to complete a sentence which has some missing words.

I'm	gonna	make	?	move
-----	-------	------	---	------

- How to complete the sentence?
  - Motivation: predict the blank based on the pattern of word co-occurrences in some specific corpus of sentences

# Language Learning with Hypernetworks (2/8)

- Goal : Sentence Completion
  - Train a hypernetwork to be able to recall any sentence in the given corpus
  - To complete sentences which contain missing words with the trained hypernetwork
  - Target function to be maximized for the sentence completion problem
    - It is the accuracy in basic
    - $f(x) := (\text{number of correct answers}) / (\text{number of whole tests})$
    - **A test** : try to guess a word in a sentence with a trained hypernetwork, and compare the result with the original sentence in dataset.
      - **How? (see the following page)**

# Language Learning with Hypernetworks (3/8)

- Learning steps
  - Sampling step
    - Randomly choose ( $n$ ) words in each sentence while preserving the order of words.
    - Repeat ( $m$ ) times to get samples per sentence
    - $\{(1,2,3,4,5,6),(n, m=3)\} :: (1,3,4), (2,3,5), (4,5,6)$
  - Weight update step
    - Guess each word with the current Hypernetwork.

# Language Learning with Hypernetworks (4/8)

- Learning steps
  - Weight update step(con't)
    - Weight update
      - Test every word with current Hypernet
      - If correct : do nothing
      - If incorrect
        - » Add a constant score (weight update rate) for every **related** hyperedges.
        - » If some hyperedges are not exist in the HN, regard their score as 0



# Language Learning with Hypernetworks (5/8)

- How to complete the sentence?
  - Hypernetworks : set of hyperedges with weights
  - Assume  $HN = \{ (1,2,3 : 3), (2,3,4 : 1), (1,2,4 : 5) \}$
  - Guess  $(1,2,?,4)$  : see following pages
    - $(1,2,3 : 3), (2,3,4 : 1)$  :  $\text{score}(3) = 4$
    - $(1,2,4 : 5)$  :  $\text{score}(4) = 5$
    - We assume unknown word as '4' with highest score, 5

# Language Learning with Hypernetworks (6/8)

- Sentence Completion

I'm	gonna	make	?	move
-----	-------	------	---	------

make	you	move
------	-----	------

make	Tom	move
------	-----	------

gonna	make	you
-------	------	-----

hyperedges in a library

# Language Learning with Hypernetworks (7/8)



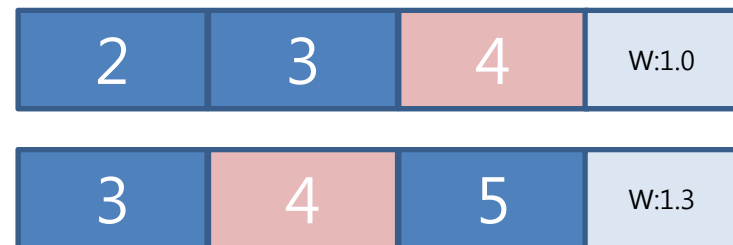
3 : 1.5  
4 : 1.0 + 1.3  
6 : 1.6

→ Answer → 4!!

# Language Learning with Hypernetworks (8/8)

- We can also analysis word associations.
- We can enumerate the associativity of words based on following features of hypernetworks
  - the co-occurrence of words in a hyperedge
  - the weight of the hyperedge

- For example
  - 4 is associated with
    - 2 with weight 1.0
    - 3 with weight 2.3(1.0+1.3)
    - 5 with weight 1.3



# Data Set (Friends & BBC)

- English sentences from movie subtitles
- Friends
  - Well known TV situation comedy
  - Captions from Season 1~10
  - 5,000 sentences
- BBC Documentary
  - Captions from three series on space, bird, and wild
  - 5,000 sentences

# Data Set (cont'd)

- Each sentence is translated to integer form based on dictionary file.
  - "This is not even a date"
  - → "33,34,35,36,27,37"
- Experiment with
  - friends\_training.csv, bbc\_training.csv
- Original sentence file
  - friends\_original.txt, bbc\_original.txt

# Tool For Hypernetworks

- Language Game (for this project)
  - Sentence Completion
  - Language Classification
  - Word Association
- Multimodal Game
  - Language to Image Generation
  - Image search using language query

# Report Contents – Mandatory (1/2)

- System description
  - Used software and running environments
- Result graphs and tables
  - Do several experiments and calculate average & standard deviations
- Analysis & discussion
  - Very Important



# Report Contents – Mandatory (2/2)

- Basic experiments
  - Draw learning curves for each training set
    - Graph type 1: accuracy vs. epoch, with *orders* of hyperedges ex) 3,4,5
    - Graph type 2: accuracy vs. epoch, with different weight update parameters ex) 0.01, 0.1, 0.5, 1, 2, 5, 10
      - Note : initial weight is assigned as 1.0 for each hyperedge
  - Comparison between two training sets
    - Graph type 3: learning curves for Friends and BBC sentences in one graph

# Report Contents – Optional

- Various experiments and analyses
  - Comparing learning curves
    - w/ various setting(varying order & update parameters & training data)
  - Comparing *word associations*
    - w/ different training set
  - Comparing sentence completion results
    - w/ different training set
    - Test ( 1,2,?,4) (**find out suitable queries**)
      - For training set A: ? = 3
      - For training set B: ? = 5, why?

# Reports Style

- **English only**, Scientific journal-style
  - How to Write A Paper in Scientific Journal Style and Format  
<http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWsections.html>

<b>Experimental process</b>	<b>Section of Paper</b>
What did I do in a nutshell?	<a href="#"><u>Abstract</u></a>
What is the problem?	<a href="#"><u>Introduction</u></a>
How did I solve the problem?	<a href="#"><u>Materials and Methods</u></a>
What did I find out?	<a href="#"><u>Results</u></a>
What does it mean?	<a href="#"><u>Discussion</u></a>
Who helped me out?	<a href="#"><u>Acknowledgments (optional)</u></a>
Whose work did I refer to?	<a href="#"><u>Literature Cited</u></a>
Extra Information	<a href="#"><u>Appendices (optional)</u></a>

# Submission Guide

- Due date: May 27, 13:00
- Submit both 'hardcopy' and 'email'
  - Hardcopy submission to the office (301-417)
  - E-mail submission to [ykko@bi.snu.ac.kr](mailto:ykko@bi.snu.ac.kr)
    - Subject : [AI Project2 Report] Student number, Name
  - Length: report should be summarized within 12 pages.
  - If you build a program by yourself, submit the source code with comments
- Objective: NOT the accuracy and your programming skill, but your creativity and research ability.
- **Individual project! You have to do it by yourself.**

# Marking Scheme

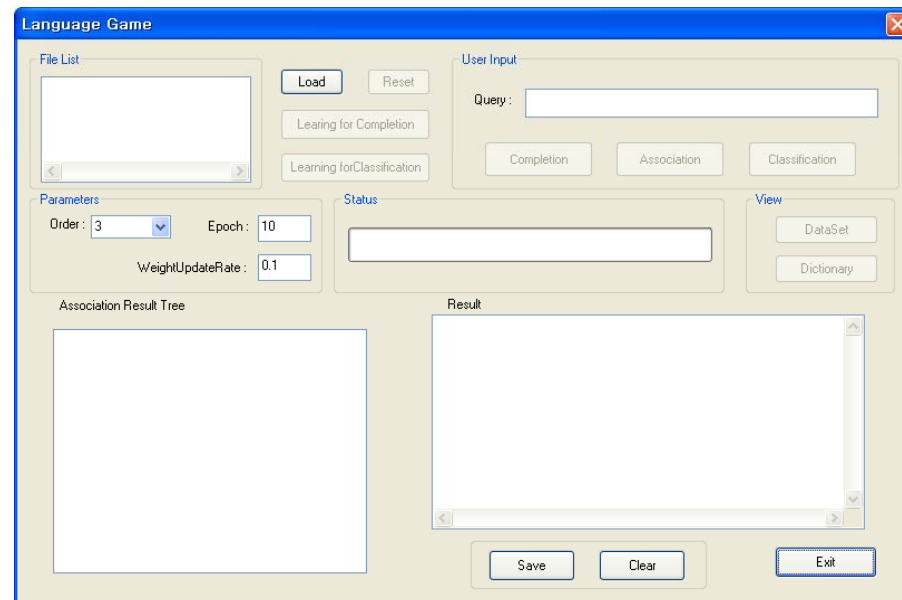
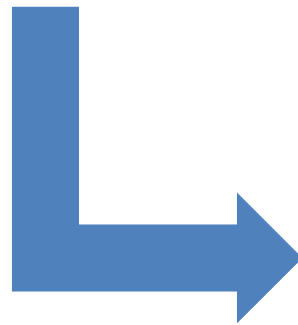
- 40 points for experiment & analysis
  - Extra 3 points per additional experiment
- 20 points for the report
- 6 points for overall organization
- Late work
  - (- 10%) per one day (8 points)
  - Maximum 7 days

# Demo – How to Start

- Execute: MemoryGame\_2.0.exe



Click!!



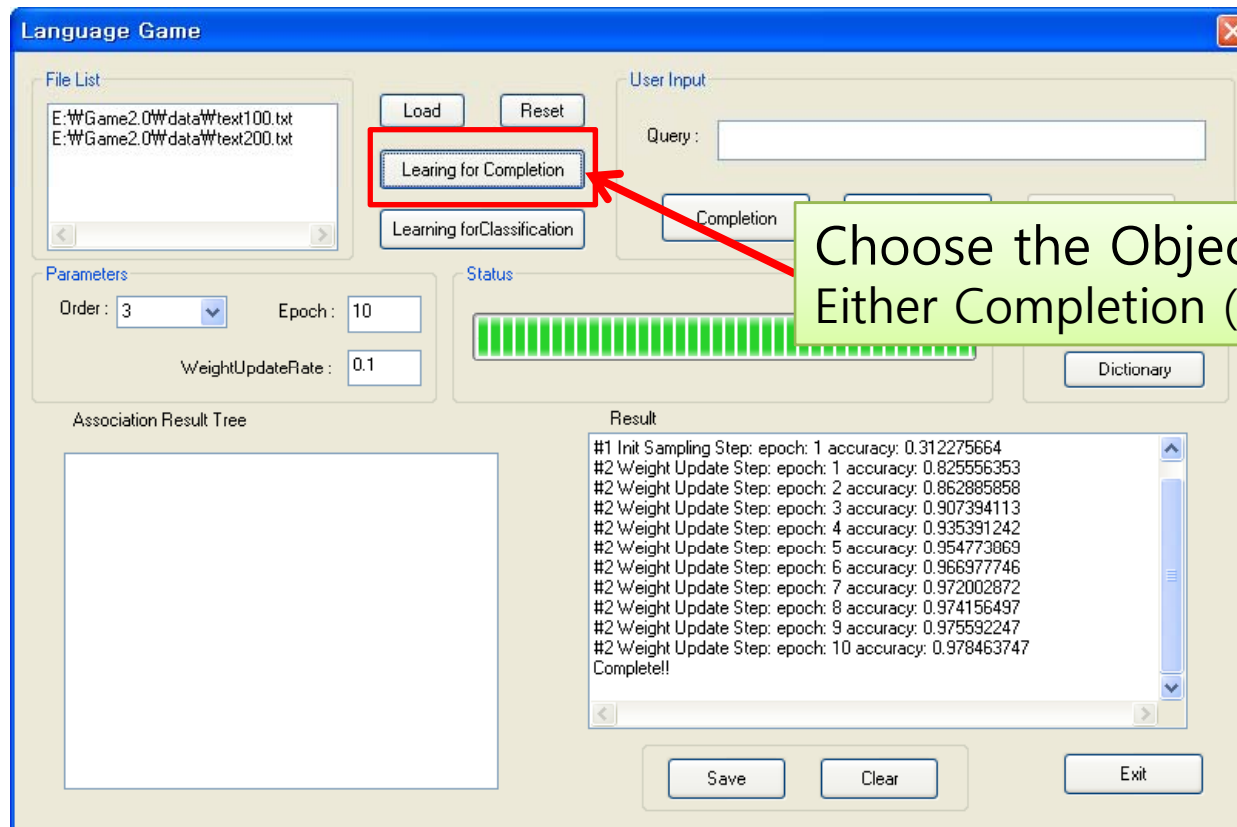
# Setting data and parameters

The screenshot shows the 'Language Game' application window. The interface includes a 'File List' area, a 'Load' button, a 'Parameters' section with 'Order' (3), 'Epoch' (10), and 'WeightUpdateRate' (0.1) fields, and a 'Result' area. Red boxes and arrows highlight the 'Load' button and the 'Parameters' section. A green callout box points to the 'Load' button, and another points to the 'Parameters' section.

**Load Training File**  
- friends\_training.csv  
or BBC\_training.csv

**Set learning parameters**  
- Order  
- Epoch  
- WeightUpdateRate

# Do Learning - Completion



Warning: it takes much time



# Test (sentence completion)

## Sentence Completion

The screenshot displays the 'Language Game' application interface. The 'User Input' section contains a text box with the query 'i realize i am ? naked', which is highlighted by a red box and a red arrow pointing from a green callout box labeled 'Query with only one blank'. Below the input are buttons for 'Completion', 'Association', and 'Classification'. The 'Parameters' section shows 'Order: 3', 'Epoch: 10', and 'WeightUpdateRate: 0.1'. The 'Status' section features a progress bar. The 'Result' section shows a list of accuracy metrics for each epoch, followed by a red box containing the completed sentence: 'i realize i am totally naked', with a red arrow pointing from a blue callout box labeled 'Result'. The 'File List' section shows two text files: 'E:\Game2.0\data\text100.txt' and 'E:\Game2.0\data\text200.txt'. The 'Association Result Tree' is currently empty.

# Learning for Completion

## Word Association

The screenshot shows the 'Language Game' software interface. The 'User Input' section has a 'Query' field containing the word 'realize'. A red box highlights this field, with a red arrow pointing to it from a green callout box that says 'Query with only one word'. Below the query field are 'Completion' and 'Association' buttons. The 'Association' button is highlighted with a red box, and a red arrow points from it to the 'Association Result Tree' section. The 'Association Result Tree' shows a list of words with their associated weights: 'i: 4.100000000', 'a: 1.500000000', 'and: 1.100000000', 'phone: 1.000000000', 'there's: 1.000000000', 'totally: 1.000000000', and 'am: 0.500000000'. A red box highlights this list, with a red arrow pointing from it to a green callout box that says 'Association result' and 'i - word' and '1.900000 - weight sum (strength)'. The 'Result' window shows the text 'Completion!' and 'Sentence Completion:'. The 'Parameters' section shows 'Order: 3', 'Epoch: 10', and 'WeightUpdateRate: 0.1'. The 'Status' section shows a progress bar. The 'View' section has 'DataSet' and 'Dictionary' buttons. The 'File List' section shows two files: 'E:\Game2.0\data\text100.txt' and 'E:\Game2.0\data\text200.txt'. The 'Load' and 'Reset' buttons are also visible. The 'Learning for Completion' and 'Learning for Classification' buttons are also present. The 'Save', 'Clear', and 'Exit' buttons are at the bottom.

# Reset

The screenshot shows the 'Language Game' application window. The 'Reset' button is highlighted with a red box, and a red arrow points from a callout box to it. The callout box contains the text 'Remove trained model and File list'. The interface includes a 'File List' panel, a 'User Input' section with a 'Query' field containing 'realize', and buttons for 'Completion', 'Association', and 'Classification'. A 'Status' section features a green progress bar and a 'WeightUpdateRate' field set to '0.1'. The 'Association Result Tree' shows a tree with a single node 'realize'. The 'Result' panel displays the output: 'Complete!', 'Sentence Completion', 'Answer for completion: i realize i am totally naked', and 'Words Association' with a list of words: 'i', 'a', 'and', 'phone', 'there's', 'totally', and 'am'. At the bottom, there are 'Save', 'Clear', and 'Exit' buttons.

# Save result

Language Game

File List

User Input

Query : realize

Completion Association Classification

Load Reset

Learning for Completion

Learning for Classification

Parameters

Order : 3 Epoch : 10

WeightUpdateRate : 0.1

Status

View

DataSet

Dictionary

Association Result Tree

Result

Complete!!

-----Sentence Completion-----

Answer for completion :

i realize i am totally naked

-----Words Association-----

Word : i

Word : a

Word : and

Word : phone

Word : there's

Word : totally

Word : am

Save Clear Exit

If you want to save result, "result.txt" in same the folder

# Warning!!

- Program path can not have any Korean or other language except English
  - "C:\Documents and Settings\바탕 화면\MemoryGame2.0"
    - Not accepted.
  - "C:\Documents and Settings\MemoryGame2.0"
    - It's OK.
- Current program does not allow making new training files
  - Dictionary file is fixed.
  - If you want to, make dictionary file too.
- If you have any question about the program, visit the office 301-417 (Tel. 880-1835)
  - Youngkil, Ko (ykko@bi.snu.ac.kr)