Dual-Coding Theory and Connectionist Lexical Selection

Abstract

We introduce the bilingual dual-coding theory as the model for bilingual mental representations. Based on the theory, lexical selection neural networks are built for a connectionist transfer project.

Topic Area: Connectionist Machine Translation, Lexical Selection

Dual-Coding Theory and Connectionist Lexical Selection

Abstract

We introduce the bilingual dual-coding theory as the model for bilingual mental representations. Based on the theory, lexical selection neural networks are built for a connectionist transfer project.

1. Introduction

To build an artificial natural language processing system, it is important, as we believe, to understand how human processes languages. For machine translation, we should understand how the languages are represented in human mind, how the representations are mapped from one language to another, and how these representations and mappings are acquired by human.

The bilingual dual-coding theory [9] [10] in psycholinguistics partially answers the above problems. It depicts the verbal representations for two different languages as two separate but connected logogen systems, characterizes the translation precess as the activation along the connections between the logogen systems. It also claims that the acquisition of the connections are through the exposure to the two languages, without the explanation of how.

We have explored an information theoretical neural network that can acquire the verbal associations in the dual-coding theory. It is a detailed explanation of the learning mechanism of the bilingual dual-coding theory. On the other hand, it provides a learnable lexical selection sub-system for a connectionist transfer project.

2. The Dual-Coding Theory

One of the debates in psycholinguistics concerns the bilingual mental representations. The independence position assumes that bilingual memory is represented by two functionally independent storage and retrieval systems that interact only through translation process. The interdependence position hypothesizes that all information of languages exists in a common memory store. Researches on cross-language transfer and cross-language priming found evidence for both hypotheses [5] [2] [7] [4].

Dual-coding theory tries to explain the coexistence of independent and interdependent phenomena with separate but connected structures. The general dual-coding theory (for both monolinguals and bilinguals) hypothesizes that human represents language with dual systems --- the verbal system and the imagery system. The elements of the verbal system are logogens for words in a language. The elements of the imagery system, called "imagens" are connected to the logogens in the verbal systems via *referential connections*. Logogens in a verbal system are also interconnected to each other with *associative connections*. The bilingual dual-coding theory proposes an architecture in which a common imagery system is connected to two different verbal systems, and the two verbal systems are interconnected to each other via associative connections [Figure 1]. Unlike the within-language associations, which are rich and diverse, these between-language associations involve primarily translation equivalent terms that are experienced frequently together. The interconnections among the three systems explain the interdependent functional behavior. On the other hand, the different characteristics between within-language and between-language associations account for the independent functional behavior.

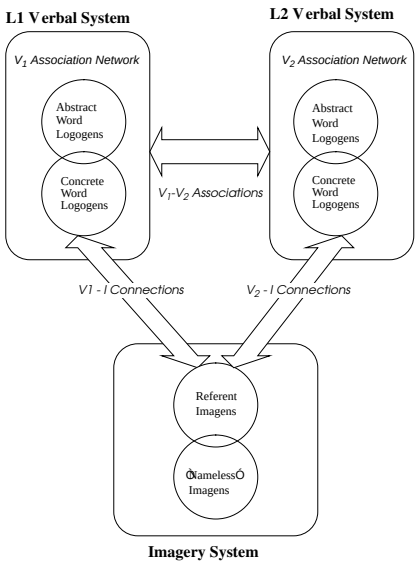

Figure 1: Bilingual Dual-Coding Representation

Together with the structural assumption, dual-coding theory proposes a parallel set of processing assumptions. Activation of connections between referentially related imagens and logogens is called *referential processing*. Naming objects and imaging to words are prototypical examples. Activation of associative connections between logogens is called *associative processing*. Lexical trans-

lation is an example of associative processing between two languages.

3. Connectionist Lexical Selection

3.1. Lexical Selection

Lexical selection is the task of choosing target-language words that accurately reflect the meaning of the corresponding source-language words. It plays an important role in machine translation [11].

The most straightforward method of lexical selection is dictionary lookup. However, since there is no one-to-one correspondence of words in two languages, this method is very restricted.

A common lexical selection practice involves an intermediate representation. It disambiguates the source language word to an entity in the intermediate representation, then maps from the entity to the target lexical entry. This intermediate representation may be Lexical Concept Structure [3] or interlingua concepts [8]. This engineering approach requires great effort in designing the representation and the mapping rules.

Currently, there are some efforts in statistical lexical selection. A target language word W_t can be selected with the posterior probability $\Pr(W_t|W_s)$ given the source language word W_s. Several target language lexical entries may be selected for a single source language word. Then the correct selections may be identified by the language model of the target language [1]. This approach is learnable. However, the accuracy is low. One reason is that it does not use any structural information of the languages.

We propose an approach based on the bilingual dual-coding theory: if we have a mechanism to implement the structures, processes and association acquisition of the bilingual dual-coding theory, we can use the between-language associative process to select target language lexical entries.

3.2. Information-Theoretical Networks

Connectionist network and its learning abilities provide us with an ideal mechanism to simulate the structures and processes in the dual-coding theory. The interconnected neurons, together with the process of local activation through the interconnections, are isomorphic to the interconnected logogens (or imagens) and local activation of associative (or referential) connections.

Information-theoretical network [6] is a neural network formalism that is capable of doing associations between two layers of representations. The associations can be obtained statistically according to the network's experiences.

The network has two layers. Each unit of a layer represents an element in the input or output of a training pattern, which might be a logogen or a word. Units in different layers are connected, while units in the same layers are not. The weight of the connection between unit i in one layer and unit j in the other layer is determined by the mutual information between the elements represented by the two units [6]:

$$w_{ij} = I(v_i, v_j) = log(Pr\{v_j \mid v_i\}/Pr\{v_i\})^1$$

Each layer also contains a bias unit, which is always activated. The weight of the connection between the bias unit in one layer and unit j in the other layer is determined by

$$w_{0j} = logPr(v_j)$$

Since the weights are determined statistically in an information-theoretical network, the network can learn in a single pass. And it can incrementally adapt to new experiences simply by adding new data to training samples and modifying the associations according to the changed statistics.

3.3. Lexical Selection as an Associative Process

A connectionist transfer project tried to map source language f-structures to target language f-structures [12]. Functionally, It had two sub-tasks: 1. finding the target (sub)structures, their phrasal categories and their corresponding source (sub)structures; 2. finding the head of a target (sub)structure. The second sub-task is a problem of lexical selection. It was done with a backpropagation network in [12], which tooks too much time for training.

We replaced the backpropagation networks in [12] with information-theoretical networks that simulate the associative process in the dual-coding theory. The networks had two layers of units. Each source (target) language lexical item was represented by a unit in the input (output) layer. One network was built for each phrasal category (NP, VP, AP, etc).

The networks worked in the following way: for a target-language f-structure to be generated, the transfer system knew its phrasal category and its corresponding source-language f-structure from the networks that per-

[1] Where v_i means the event that unit i is activated.

formed the sub-task 1 described above. It then activated the lexical selection network for that phrasal category with the input units that corresponded to the heads of the source language f-structure and its sub-structures. Through the connections between the two layers, the output units were activated, and the lexical item that corresponded to the most active output unit was selected as the head of the target f-structure. As an example, we show how the system selected the head "anmelden" for the German XCOMP sub-structure when it did the transfer from [$_{sentence}$ [$_{subj}$ I] would [$_{xcomp}$ [$_{subj}$ I] like [$_{xcomp}$ [$_{subj}$ I] to register [$_{pp-adj}$ for the conference]]]] to [$_{sentence}$ [$_{subj}$ Ich] werde [$_{xcomp}$ [$_{subj}$ Ich] [$_{adj}$ gerne] anmelden [$_{pp-adj}$ fuer der Konferenz]]][2]. Since the structure network determined that there was a VP sub-structure of XCOMP in the target structure whose corresponding input structure was [$_{xcomp}$ [$_{subj}$ I] to register [$_{pp-adj}$ for the conference]]], it activated the VP lexical selection network's input units for *I, register* and *conference*. By propagating the activation via the associative connections, the unit for *anmelden* were the most active output. Therefore, *anmelden* was chosen as the head of the XCOMP sub-structure.

The connection strength of the networks was assigned with the mutual information between two units, which was determined by the samples drawn from around 300 f-structures of the Conference Registration Conversation sentences. For the above example, the association ([I, register, conference] [anmelden]) was a sample drawn from the f-structures to train the VP network. The task had lexicon size of around 500 lexical items each for the source language (English) and the target language (German).

3.4. Preliminary Result

The accuracy of the lexical selection based on the dual coding theory was slight lower than the one in [12] (around 84% vs. around 96%) for the traininmg data. However, the generalization performance on the unseen inputs was better than [12] (around 70% vs. around 56%).

4. Summary

The lexical selection approach discussed here has two advantages. First, it is learnable. It does not require knowledge engineering efforts. Secondly, it is psycholinguistically well-founded --- the approach adopts a local activation processing model instead of symbol passing in typical symbolic systems; the approach is based on a underlying psycholinguistic theory; and knowledge in this approach is learned, but not imposed.

References

1. Peter F. Brown, et al. "A Statistical Approach to Machine Translation". *Computational Linguistics 16(2)* (1990), 73-85.

2. Annette M. de-Groot, and Gerard L. Nas, "Lexical representation of cognates and noncognates in compound bilinguals". *Journal of Memory and Language 30(1)* (1991).

3. Bonnie J. Dorr. Conceptual Basis of the Lexicon in Machine Translation. Tech. Rept. A.I. Memo No. 1166, Artificial Intelligence Laboratory, MIT, August, 1989.

4. Cheryland Frenck and Joel Pynte. "Semantic representation and surface forms: A look at across-language priming in bilinguals". *Journal-of-Psycholinguistic-Research 16(4)* (1987).

5. Linda D. Gerard and Don L. Scarborough. "Language-Specific Lexical Access of Homographs by Bilinguals". *Journal of Experimental Psychology: Learning, Memory, and Cognition 15,* 2 (1989).

6. A. L. Gorin and S. E. Levinson. Adaptive Acquisition of Language. Speech Research Department, AT&T Bell Laboratories, Murray Hill, 1989.

7. W. E. Lambert, J. Havelka and C. Crosby. "The influence of language acquisition contexts on bilingualism". *Journal of Abnormal and Social Psychology 56* (1958).

8. Sergei Nirenberg, Victor Raskin and Allen B. Tucker. The structure of Interlingua in TRANSLATOR. In *Machine Translation: Theoretical and Methodological Issues*, Sergei Nirenburg, Ed., Cambridge University Press, Cambridge, England, 1987.

9. Allan Paivio. *Imagery and verbal Processes.* Holt, Rinehart and Winston, Inc., New York, 1971.

10. Allan Paivio. *Mental Representations --- A Dual Coding Approach.* Oxford University Press, New York, 1986.

11. James Pustejovsky and Sergei Nirenburg. Lexical Selection in the Process of Language Generation. Proceedings of the 25th Annual Conference of the Association for Computational Linguistics, Standford University, Standford, CA, 1987, pp. 201-206.

12. Y. Wang and A. Waibel. Hybrid Symbolic/Connectionist Transfer. submitted to ACL '94, 1994.

[2]The f-structures are simplified here for the sake of conciseness