

# Adaptive Resonance Theory Based Two-Stage Chinese Name Disambiguation

Xin Wang, Yuanchao Liu, Xiaolong Wang, Ming Liu, Bingquan Liu

School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

Email: [xwang@insun.hit.edu.cn](mailto:xwang@insun.hit.edu.cn)

Received April 18, 2012; revised July 16, 2012, accepted July 23, 2012

## ABSTRACT

It's common that different individuals share the same name, which makes it time-consuming to search information of a particular individual on the web. Name disambiguation study is necessary to help users find the person of interest more readily. In this paper, we propose an Adaptive Resonance Theory (ART) based two-stage strategy for this problem. We get a first-stage clustering result with ART1 model and then merge similar clusters in the second stage. Our strategy is a mimic process of manual disambiguation and need not to predict the number of clusters, which makes it competent for the disambiguation task. Experimental results show that, in comparison with the agglomerative clustering method, our strategy improves the performance by respectively 0.92% and 5.00% on two kinds of name recognition results.

**Keywords:** Name Disambiguation; Adaptive Resonance Theory; Text Clustering; Natural Language Processing

## 1. Introduction

Searching for person is an important part of real-world entity searching activity on the World Wide Web. We randomly selected 10,000 of 21,426,941 non-empty queries from the query logs of June in 2008 provided by Sogou, a widely-used Chinese search engine. Statistical result shows that 13% of the queries include at least one person's name and 8.2% is exactly one name. It's common that different individuals share the same name. There are 290 thousand people share the single name “张伟” in China<sup>1</sup>. It's time-consuming to filter the search results among lots of namesakes to get information of the target individual and one will probably fail to recall the considerable web pages during the procedure [1]. With the development of related fields like object retrieval, entity linking and communities identification, name disambiguation attracts more researchers' attention.

Analogous with Web People Search Evaluation Campaign (WePS) focusing on the English names ambiguity, CIPS-SIGHAN Joint Conference on Chinese Language Processing (SIGHAN2010) features a bake-off on Chinese personal name disambiguation. This task defined as a problem of documents clustering. Each cluster in the result is corresponding to a real-world person entity.

During the process of manual disambiguation, such as distinguishing the search result of a person name, we have no idea about how many different individuals sharing the query name involved in the ranked pages. Besides,

the main steps of macroscopic process of thinking and learning during manual disambiguation could be described as follows:

- 1) Compare current web textual document with the representative patterns of known persons, usually some distinctive social attribute;
- 2) If the document is similar enough with a pattern of a person, namely can be considered as the same person based on human knowledge, we classify it into the documents' set standing for the very individual and update the pattern by integrating current document;
- 3) Otherwise, we treat it as a new individual, and construct a new category for her/him;
- 4) After disposing all the pages, we may check the result and combine some clusters probably relevant to the same person entity.

This paper presents an adaptive resonance theory based two-stage strategy for the name disambiguation problem. In the first stage, we use a ART1 model to get results of a high accuracy; then an agglomerative clustering algorithm is implemented to merge clusters sharing a high similarity. Our method is similar with the process of manual disambiguation and need not to predict the number of clusters, which makes it competent for the disambiguation task.

The remainder of this paper is organized as follows. In Section 2, we give an overview of previous studies on name disambiguation. In Section 3, we present the details of our method. Section 4 describes experimental results on the SIGHAN2010 task 3 data sets. Finally, we present

<sup>1</sup><http://news.cctv.com/society/20070725/105888.shtml>

our conclusions and future work suggestion in Section 6.

## 2. Related Work

Treating the problem as cross-document coreference resolving, Bagga and Baldwin [2] use a standard vector space model. First of all, their system produces coreference chains within each document. Then it outputs a summary of the document by extracting the sentences that related to the chains. Summaries whose cosine similarity is higher than a specific threshold will be agglomerated into the same cluster. Most works represent documents as vectors now, and improvements are achieved in the following aspects: extraction of feature, usage of web resources and application of algorithm.

Mann and Yarowsky [3] get a rich feature space by extracting the biographic facts e.g. occupation and birth place. Attribute extraction is a subtask of WePS2, consisting of extracting 18 kinds of “attribute values” for target individuals<sup>2</sup>. Ontology-driven attribute extraction from unstructured text also provides credible evidence for entity name disambiguation in specific field [4]. Due to the lack of biographical information on the web, Bollegala *et al.* [5] develop an automatically key phrases extraction method to enrich the feature.

There are also many approaches focusing on the rapidly increasing resource on the web. Bunescu *et al.* [6] make a use of free online encyclopedia. They consider a Wikipedia item as a named entity, and generate a dictionary mapping the ambiguous documents to their possible named entity. Similarly, X. P. Han and L. Sun [7] convert a name disambiguation problem into an entity linking task. A generative model is proposed in their work to classify the contexts. To improve the estimation of similarity, Vu *et al.* [8] enrich the conventional feature by invoking the directories on the web. In their work, measurement of term weights and similarities is modified to take word frequency of each directory into consideration. X. P. Han and J. Zhao [9] extract professional taxonomy from the Freebase. With the assumption of one individual relevant to one profession, the authors classify documents into a real world professional taxonomic category. Then, the documents in the same professional category will be put into a same cluster. X. Zhou *et al.* [10] use exclusive character attributes contained in a taxonomic knowledge base to get more reliable similarity. Similarity will be reduced if the documents have exclusive attributes.

Various kinds of strategies have been applied in this field. Uncertainty of the number of individuals makes the clustering algorithms suitable for the problem. While most researchers implement an agglomerative clustering method [2,3,5], Pedersen *et al.* [11] use a divisive one to

differentiate documents. Fuzzy ant based clustering is inspired by the behavior of ants clustering dead nest mates into piles. E. Lefever *et al.* [12] propose an approach of high adaptability based on the fuzzy ant algorithms. Recent work including [13,14] respectively use a multi-stage strategy, in which reliable attributes are extracted to ensure the accuracy in the first stage and other features are involved to improve the low recall values of the result.

Social network based disambiguation is another main kind of clustering-based approaches [15-17]. Different from the biographic facts concerned method discussed above, social network based partition algorithms focus on the relationship between named entities. Firstly, these works extract co-occurrence of person names or link structure of the web pages. Then a graphical model (social network) is constructed based on the extraction result. The network will be divided into subgraphs relevant to different persons in the end.

Chinese name disambiguation has its own uniqueness. Chinese texts have no capital letters or space boundaries, which makes name recognition and word segmentation needed before clustering perform. Related works range from Chinese dependent feature extraction and knowledgebase utilization [10] to language independent algorithm application [17].

## 3. Two-Stage Disambiguation via ART

### 3.1. Adaptive Resonance Theory

Adaptive Resonance Theory (ART) model was proposed by S. Grossberg [18,19] as a result of long-term concentration on human cognitive information processing. Besides ART1, which is implicated in our research, some other models in the family [20-24] are studied by researchers of variable fields such as medical diagnosis [25], semiconductor fabrication [26] and intrusion detection [27].

The basic structure of ART network is shown in **Figure 1**. The ART network consists of attention subsystem and orienting subsystem. Comparison layer C is the underlying network of attention subsystem, where the input pattern is contrast enhanced and normalized before entering the recognition layer R. For each stored patterns in R, the network compares input pattern X with them. If the matching level of a certain pattern reaches the required attentional vigilance, *i.e.* produce good resonance, the network classifies X into the relevant category of the stored pattern, and modifies the prototype of the category to ensure the corresponding node win the competition when similar input patterns occur. The iteration is repeated over all nodes on output layer of R, but if no required category is found, a new pattern would be estab-

<sup>2</sup> <http://nlp.uned.es/weps/weps-2/weps2-task-guidelines>

lished in R through orienting subsystem.

Considering the name disambiguation as a cognitive process, the principle of the ART network is quite similar to the manual disambiguation process as described in Section 1. Moreover, similar to the biological memory, the memory capacity of the ART model can increase with the number of the input samples. Therefore, the ART network clustering process does not require a specific number of clusters as input, and this advantage makes it especially suitable for the disambiguation problem of uncertain number of namesakes.

### 3.2. Agglomerative Clustering in Stage Two

As explained in 3.1, the ART model is competent for pattern comparison, classification and new category construction. However, even if documents in one category have a high similarity with those in another one, meaning these two categories probably refer to the same person, an ART1 algorithm never merges them.

An agglomerative clustering approach is able to take the global similarity into consideration. To combine the advantages of both methods, we propose a two-stage strategy in this paper, which is shown in **Figure 2**. In the first stage, an ART1 model is implemented to classify the documents into categories that constructed automatically. Then the clusters sharing a high similarity are merged by an agglomerative method.

## 4. Experimental Results and Analysis

### 4.1. Data Set and Corpus Aimed Preprocessing

SIGHAN2010 task 3 focuses on the Chinese name disambiguation problem. In this task, a manually clustered corpus contains 32 names is provided as train data. Each name in the corpus is relevant to a various number (from 100 to 400) of news from Xinhua News Agency and the total number of the documents is 6926. This corpus is used as the experimental data set in this paper.

The names given in the corpus is a substring of the news text, which means there are two situations where a person name recognition module is needed:

1) *The given name is the substring of the exact name of the person reported in the news (e.g. “张伟” is a substring of “张伟东”);*

2) *The given name is a polysemic word, which refers to another concept in the text rather than a person’s name (e.g. the name string “黄海” also means the yellow sea).*

In the light of this particularity of the corpus, we develop a preprocessing modular to divide documents into several sub-sets based on the different recognized names.

Then our method is performed on these sub-sets, before combining the results to get final clusters. If a

polysemic word referring to another concept is properly recognized, the relevant document will be put in a “discarded” cluster. In the experiments, we test our method on both machine person name recognition result and manual name tagged result.

### 4.2. Evaluation Metrics

There are two kinds of clustering scoring methods [28, 29] used in the evaluation. The formulas are shown in **Table 1**.

### 4.3. Features and Baseline System

We use classical textual features in our experiments. Firstly, word segmentation is executed. The nouns, verbs, adjectives and adverbs are extracted as features, after removing the stop words. Each document is represented as a vector consisting of Boolean values of existence of the feature words.

Baseline system implements a single-link agglomerative clustering based on cosine similarity with a standard vector space model. In the second stage of our method, we use the same approach to combine clusters.

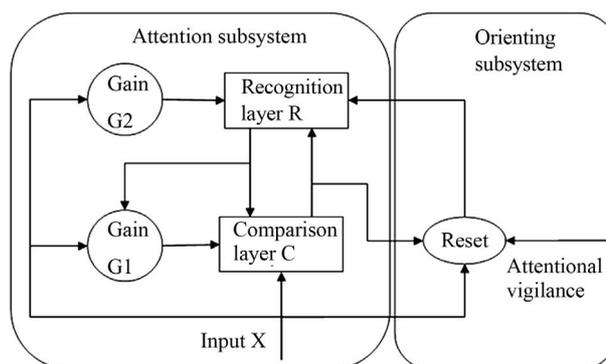


Figure 1. Structure of ART network.

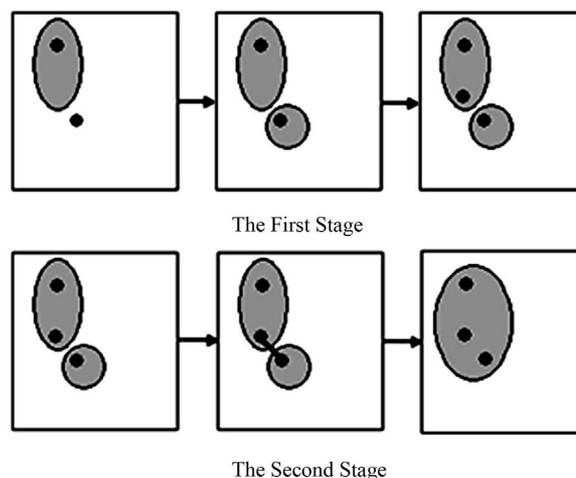


Figure 2. Two stages of our method.

**Table 1. The formulas of the two scoring algorithms for non-disjoint clustering.**

Scoring	Purity-Based	Modified B-Cubed
Precision (Purity)	$\frac{\sum_{S_i \in S} \max_{R_j \in R}  S_i \cap R_j }{\sum_{S_i \in S}  S_i }$	$\frac{\sum_{S_i \in S} \sum_{d \in S_i} \max_{R_j \in R; d \in R_j} \frac{ S_i \cap R_j }{ S_i }}{\sum_{S_i \in S}  S_i }$
Recall (Inverse-Purity)	$\frac{\sum_{R_i \in R} \max_{S_j \in S}  R_i \cap S_j }{\sum_{R_i \in R}  R_i }$	$\frac{\sum_{R_i \in R} \sum_{d \in R_i} \max_{S_j \in S; d \in S_j} \frac{ R_i \cap S_j }{ R_i }}{\sum_{R_i \in R}  R_i }$
F-Measure	$\frac{2 \times \text{Purity} \times \text{Inverse-purity}}{\text{Purity} + \text{Inverse-purity}}$	$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

#### 4.4. Evaluation Results

We test our approach on 16 different names, and the standard clustering result of the other names are used as training data. After parameter adjustment, we set the agglomerative similarity threshold to 0.2, and the ART attentional vigilance to 0.75.

It is shown in **Table 2** that performance (F1 value of B\_Cubed score) of our strategy improves by 0.92% in comparison with the baseline system, when taking the machine recognized names as input of the preprocessing modular. Furthermore, when taking the manual tagged names as input, our strategy achieves a significant improve by 3.00% and 5.00% in Purity score and B-Cubed score respectively. The corresponding results are shown in **Table 3**.

According to the scores indicated in **Tables 2** and **3**, there's large gap between the result of recognized name and tagged name. Precision of name recognition on test data is shown in **Figure 3**.

In **Figure 4** we show the detailed performance (Precision, Recall and F-measure) on each name in test data set. These names are representative for different situations as mentioned above, some refer to another concept, and some are substrings of longer names. Even if the given name is the exact name of persons, the number of person entities and documents distribution on these entities is variable. As indicated in **Figure 4** our approach is generally effective in all these cases.

#### 5. Conclusions and Further Work

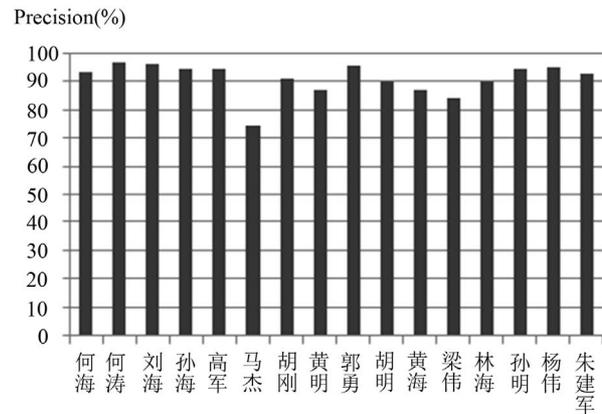
This paper proposes a two-stage strategy to simulate the process of manual name disambiguation. ART1 algorithm is used in pattern comparison, classification and new category construction, and then agglomerative clustering combines the similar clusters. Experimental results

**Table 2. F-scores of different strategies on machine name recognition result.**

F-scores (%)	Purity/Inverse-purity	B_Cubed
Baseline	83.97	77.21
Two-stage	84.35	78.13

**Table 3. F-scores of different strategies on manual name tagged result.**

F-scores (%)	Purity/Inverse-purity	B_Cubed
Baseline	87.69	81.96
Two-stage	90.69	86.96

**Figure 3. Precision of name recognition.**

show that our approach achieves a clearly improvement, increasing respectively 0.92% and 5.00%, on automatic name recognition result and manual name tagged result.

To represent a person entity more properly, a more targeted attribute extraction will be studied. Social information and text theme will be selected as feature in our further study. Analog patterns consisting of word

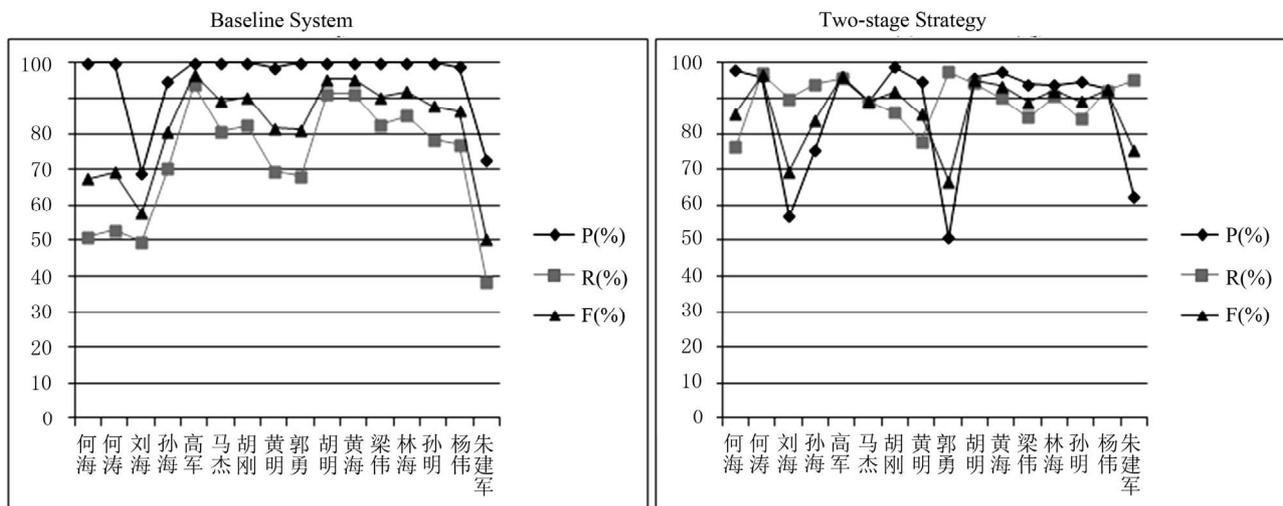


Figure 4. Performance comparison of baseline and ART based two-stage strategy.

frequency weights and the importance weights will be used as input, when other ART models are testified in experiments.

## 6. Acknowledgements

This work was supported by Key Laboratory Opening Funding of China MOE-MS Key Laboratory of Natural Language Processing and Speech. Query log data is provided By R&D Center of SOHU.

## REFERENCES

- [1] J. Artilles, S. Sekine and J. Gonzalo, "Web People Search: Results of the First Evaluation and the Plan for the Second," *Proceeding of the 17th International Conference on World Wide Web*, Beijing, 21-25 April 2008, pp. 1071-1072. [doi:10.1145/1367497.1367661](https://doi.org/10.1145/1367497.1367661)
- [2] A. Bagga and B. Baldwin, "Entity-Based Cross-Document Coreferencing Using the Vector Space Model," *Proceedings of the 17th International Conference on Computational Linguistics*, Montreal, 10-14 August 1998, pp. 79-85. [doi:10.3115/980451.980859](https://doi.org/10.3115/980451.980859)
- [3] G. S. Mann and D. Yarowsky, "Unsupervised Personal Name Disambiguation," *Proceedings of the 7th Conference on Natural Language Learning at HLT-NAACL 2003*, Edmonton, 31 May 2003, pp. 33-40.
- [4] J. Hassell, B. Aleman-Meza and I. Arpinar, "Ontology-Driven Automatic Entity Disambiguation in Unstructured Text," *Proceedings of the 5th International Semantic Web Conference*, Springer Berlin/Heidelberg, 2006, pp. 44-57.
- [5] D. Bollegala, Y. Matsuo and M. Ishizuka, "Disambiguating Personal Names on the Web Using Automatically Extracted Key Phrases," *European Conference on Artificial Intelligence*, 2006, pp. 553-557.
- [6] R. C. Bunescu and M. Pasca, "Using Encyclopedic Knowledge for Named entity Disambiguation," *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, The Association for Computer Linguistics, 2006, pp. 9-16.
- [7] X. Han and L. Sun, "A Generative Entity-Mention Model for Linking Entities with Knowledge Base," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, 2011, pp. 945-954.
- [8] Q. M. Vu, A. Takasu and J. Adachi, "Improving the Performance of Personal Name Disambiguation Using Web Directories," *Information Processing & Management*, Vol. 44, No. 4, 2008, pp. 1546-1561. [doi:10.1016/j.ipm.2007.11.001](https://doi.org/10.1016/j.ipm.2007.11.001)
- [9] X. Han and J. Zhao, "CASIANED: Web Personal Name Disambiguation Based on Professional Categorization," *2nd Web People Search Evaluation Workshop (WePS 2009) at 18th WWW Conference*, Madrid, 2009.
- [10] X. Zhou, C. Li, M. Hu and H. Wang, "Chinese Name Disambiguation Based on Exclusive Character Attributes," *Proceedings of the 6th China National Conference on Information Retrieval (CCIR2010)*, Harbin, 2010, pp. 333-340.
- [11] T. Pedersen, A. Purandare and A. Kulkarni, "Name Discrimination by Clustering Similar Contexts," *Computational Linguistics and Intelligent Text Processing*, Vol. 3406, 2005, pp. 226-237. [doi:10.1007/978-3-540-30586-6\\_24](https://doi.org/10.1007/978-3-540-30586-6_24)
- [12] E. Lefever, T. Fayruzov, V. Hoste and M. De Cock, "Clustering Web People Search Results Using Fuzzy Ants," *Information Sciences*, Vol. 180, No. 17, 2010, pp. 3192-3209. [doi:10.1016/j.ins.2010.05.018](https://doi.org/10.1016/j.ins.2010.05.018)
- [13] M. Ikeda, S. Ono, I. Sato, M. Yoshida and H. Nakagawa, "Person Name Disambiguation on the Web by Two-Stage Clustering," *2nd Web People Search Evaluation Workshop (WePS 2009) at 18th WWW Conference*, Madrid, 2009.
- [14] H. Ding, T. Xiao and J. Zhu, "A Mutil-Stage Clustering Approach to Chinese Person Name Disambiguation," *Proceedings of the 6th China National Conference on Information Retrieval (CCIR2010)*, Harbin, 2010, pp. 316-324.

- [15] R. Bekkerman and A. McCallum, "Disambiguating Web Appearances of People in a Social Network," *Proceedings of the 14th International Conference on World Wide Web*, Beijing, 2005, pp. 463-470.
- [16] B. Malin, "Unsupervised Name Disambiguation via Social Network Similarity," *SIAM International Conference on Data Mining*, Newport Beach, 2005, pp. 93-102.
- [17] J. Lang, et al., "Person Name Disambiguation of Searching Results Using Social Network," *Chinese Journal of Computers*, Vol. 7, 2009, pp. 1365-1374. [doi:10.3724/SP.J.1016.2009.01365](https://doi.org/10.3724/SP.J.1016.2009.01365)
- [18] S. Grossberg, "Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors," *Formal Aspects of Computing*, Vol. 23, No. 3, 1976, pp. 121-134.
- [19] S. Grossberg, "Adaptive Pattern Classification and Universal Recoding: II. Feedback, Expectation, Olfaction, Illusions," *Formal Aspects of Computing*, Vol. 23, No. 4, 1976, pp. 187-202
- [20] G. A. Carpenter and S. Grossberg, "ART 2: Stable Self-Organization of Pattern Recognition Codes for Analog Input Patterns," *Applied Optics*, Vol. 26, No. 23, 1987, pp. 4919-4930. [doi:10.1364/AO.26.004919](https://doi.org/10.1364/AO.26.004919)
- [21] G. A. Carpenter, S. Grossberg and D. B. Rosen, "ART 2-A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition," *Neural Networks*, Vol. 4, No. 4, 1991, pp. 493-504. [doi:10.1016/0893-6080\(91\)90045-7](https://doi.org/10.1016/0893-6080(91)90045-7)
- [22] G. A. Carpenter and S. Grossberg, "ART 3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures," *Neural Networks*, Vol. 3, No. 2, 1990, pp. 129-152. [doi:10.1016/0893-6080\(90\)90085-Y](https://doi.org/10.1016/0893-6080(90)90085-Y)
- [23] G. A. Carpenter, S. Grossberg and J. H. Reynolds, "ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network," *Neural Networks*, Vol. 4, No. 5, 1991, pp. 565-588. [doi:10.1016/0893-6080\(91\)90012-T](https://doi.org/10.1016/0893-6080(91)90012-T)
- [24] G. A. Carpenter, S. Grossberg and D. B. Rosen, "Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System," *Neural Networks*, Vol. 4, No. 6, 1991, pp. 759-771. [doi:10.1016/0893-6080\(91\)90056-B](https://doi.org/10.1016/0893-6080(91)90056-B)
- [25] P. J. G. Lisboa, "A Review of Evidence of Health Benefit from Artificial Neural Networks in Medical Intervention," *Neural Networks*, Vol. 15, No. 1, 2002, pp. 11-39. [doi:10.1016/S0893-6080\(01\)00111-3](https://doi.org/10.1016/S0893-6080(01)00111-3)
- [26] F. Chen and S. Liu, "A Neural-Network Approach to Recognize Defect Spatial Pattern in Semiconductor Fabrication," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 13, No. 3, 2000, pp. 366-373. [doi:10.1109/66.857947](https://doi.org/10.1109/66.857947)
- [27] D. Dasgupta and H. Brian, "Mobile Security Agents for Network Traffic Analysis," *Proceedings of DARPA Information Survivability Conference and Exposition*, Vol. 2, 2001, pp. 332-340. [doi:10.1109/DISCEX.2001.932184](https://doi.org/10.1109/DISCEX.2001.932184)
- [28] J. Artiles, J. Gonzalo and S. Sekine, "The SemEval-2007 WePS Evaluation: Establishing a Benchmark for the Web People Search Task," *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007, pp. 64-69.
- [29] A. Bagga and B. Baldwin, "Algorithms for Scoring Co-Reference Chains," *The Linguistic Co-Reference Workshop at the 1st International Conference on Language Resources and Evaluation (LREC'98)*, 1998, pp. 563-566.