Concept Induction via Fuzzy C-Means Clustering in a High Dimensional Semantic Space

Dawei Song[1] Guihong Cao[2] Peter Bruza[3] Raymond Lau[4]

[1] Knowledge Media Institute & Centre for Research in Computing
The Open University
Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
d.song@open.ac.uk

[2]Dept. d'Informatique et Recherche Operationnelle
Universite de Montreal
C.P. 6128, succursale CENTRE-VILLE, Montreal (Quebec), H3C 3J7 Canada
caogui@iro.umontreal.ca

[3]School of Information Technology
Queensland University of Technology
Brisbane, QLD 4001, Australia
p.bruza@qut.edu.au

[4] Department of Information Systems, City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong SAR
raylau@cityu.edu.hk

Abstract

Lexical semantic space models have recently been investigated to automatically derive the meaning (semantics) of information based on natural language usage. In a semantic space, a term can be considered as a concept represented geometrically as a vector, the components of which correspond to terms in a vocabulary. A primary way to perform reasoning in a semantic space is to categorize concepts in the space into a number of regions (i.e., groups). Such a process is referred to as concept induction, which can be realized by clustering objects in the space. The resulting groups can potentially form a basis for knowledge discovery and ontology construction. Conventional clustering algorithms, e.g., the K-Means method, normally produce crisp clusters, i.e., an object could be assigned to only one cluster. It is not always the case in reality. For example, a word "Reagan" may belong to both the cluster about administration of US government, and another one about the Iran-contra scandal. Therefore, a membership function is applied, which determines the degree to which an object belongs to different clusters. This chapter introduces a cognitively motivated semantic space model, namely Hyperspace Analogue to Language (HAL), and shows how a fuzzy C-Means clustering algorithm is used to concept categorization in the high dimensional semantic space. The experimental results indicate that applying fuzzy C-Means clustering over the HAL semantic space is promising in constructing semantically related groups of terms.

1. Introduction

A human encountering a new concept often derives its meaning via an accumulation of the contexts in which the concept appears. Based on this distributional characteristic of semantics, various lexical semantic space models have been investigated. The meaning of a word is captured by examining its co-occurrence patterns with other words in the language use (e.g., a corpus of text). There have been two major classes of semantic space models: document spaces and word spaces. The former represents words as vector spaces of text fragments (e.g. documents, paragraphs, etc) in which they occur. A notable example is the Latent Semantic Analysis (LSA) (Landauer and Dumais 1997). The latter represents words as vector spaces of other words, which co-occur with the target words within a certain distance (e.g., a window size). The strength of the association can be inversely proportional to the distance between the context and target words. The Hyper-space Analogue to Language (HAL) model employs this scheme (Lund and Burgess 1996). The dimensionality of semantic spaces is often very high, for example, Lund and Burgess (1996) constructed a 70,000x70,000 HAL vector space from a 300 million word textual corpus gathered from Usenet. The concepts occurring in the similar contexts tend to be similar to each other in meaning. For example, "nurse" and "doctor" are semantically similar to each other, as they often experience the same contexts, i.e., hospital, patients, etc. The similarity can be measured by the angle (Cosine) or Euclidean distance between two word vectors in the semantic space.

Semantic space models can be considered as computational approximations of the conceptual spaces advocated by (Gärdenfors 2000), which are built upon geometric structures representing concepts and their properties. At the conceptual level, information is represented geometrically in terms of a dimensional space. In this chapter, we propose to use HAL vectors to prime the geometric representation of concepts. HAL vectors are also interesting because semantic associations computed using these vectors correlate with semantic associations drawn from human subjects (Burgess et al. 1998). It has been shown that HAL vectors can be used to simulate semantic, grammatical and abstract categorizations (Burgess et al. 1998). Another advantage of the HAL approach is that it is automatic and computationally tractable.

In a conceptual space, a domain is defined as a set of integral dimensions in the sense that a value in one dimension(s) determines or affects the value in another dimension(s). For example, pitch and volume are integral dimensions representing a domain of "sound". Gärdenfors' and Williams (2001) state "the ability to bundle up integral dimensions as a domain is an important part of the conceptual spaces framework". The thrust of Gärdenfors' proposal is that concepts are dimensional objects comprising

domains. A domain, in turn, is a vector space with as basis a set of integral dimensions. Properties are represented as regions within a given domain.

By their very nature, conceptual spaces do not offer a hierarch of concepts as is often the case in ontologies, taxonomies, and the like. However, similar objects can be grouped due to their semantic similarity. By way of illustration, a set of feathered objects with wings leads to the grouping "bird". This facet of conceptual space is referred to as *concept induction* in this paper. One way of gaining operational command of concept induction is by means of clustering of objects in a semantic space. Clustering techniques divide a collection of data into groups or hierarchy of groups based on similarity of objects (Chuang and Chien 2005). A well known clustering algorithm is the K-Means method (Steinbach et al., 2000; Cimiano et al., 2005), which takes a desirable number of clusters, K, as input parameter, and outputs a partitioning of K clusters on the set of objects. The objective is to minimize the overall intra-cluster dissimilarity, which is measured by the summation of distances between each object and the centroid of the cluster it is assigned to. A cluster centroid represents the mean value of the objects in the cluster. A number of different distance functions, e.g., Euclidean distance, can be used as the dissimilarity measure.

Conventional clustering algorithms normally produce crisp clusters, i.e., one object can only be assigned to one cluster. However, in real applications, there is often no sharp boundary between clusters. For example, depending on the context it occurs, President "Reagan" could belong to a number of different clusters, e.g., one cluster about the US government administration, and another about the Iran-contra scandal. The latter reflects the fact that he was involved in the illegal arms sales to Iran during Iran-Iraq war. Therefore, a membership function can be naturally applied to clustering, in order to model the degree to which an object belongs to a given cluster. Among various existing algorithms for fuzzy cluster analysis (Höppner et al. 1999), a widely used one is the fuzzy C-Means (Hathaway et al. 2000, Krishnapuram et al. 2001, Höppner and Klawonn 2003, Kolen and Hutcheson 2002, etc.), a fuzzification of the traditional K-Means clustering.

The practical implication of the use of fuzzy clustering for conceptual induction is rooted in its ability to exploit the context sensitive semantics of a concept as represented in semantic space. There is a connection here with the field of text mining. Broadly speaking, text mining aims at extracting new and previously unknown patterns from unstructured free text (Hearst 2003, Perrin and Petry 2003, Srinivasan 2004). Conceptual space theory and its implementation by means of semantic space models introduced in this chapter provides a cognitively validated dimensional representation of

information based on the premise that associations between concepts can be mined in a principled way (Song and Bruza 2003).

The goal of this chapter is to introduce the construction of a high dimensional semantic space via the HAL model (Section 2) and address how a fuzzy C-Means clustering algorithm, presented in Section 3, can be applied to conceptual induction within a HAL space. Its effectiveness is illustrated by a case study in Section 4. Finally, we conclude the chapter and highlight some future directions in Section 5.

2. Constructing a High-Dimensional Semantic Space via Hyperspace Analogue to Language

In this section, we give a brief introduction to the Hyperspace Analogue to Language (HAL) model. Given a *n*-word vocabulary, the HAL space is a word-by-word matrix constructed by moving a window of length *l* over the corpus by one word increment ignoring punctuation, sentence and paragraph boundaries. All words within the window are considered as co-occurring with each other with strengths inversely proportional to the distance between them. Given two words, whose distance within the window is *d*, the weight of association between them is computed by $(l - d + 1)$. After traversing the whole corpus, an accumulated co-occurrence matrix for all the words in a target vocabulary is produced. HAL is direction sensitive: the co-occurrence information for words preceding every word and co-occurrence information for words following it are recorded separately by its row and column vectors. By way of illustration, the HAL space for the example text "The effects of spreading pollution on the population of Atlantic salmon" is depicted below (Table 1) using a 5 word moving window (l=5). Note that, for ease of illustration, in this example we do not remove the stop words such as "the", "of", "on", etc. The stop words are dropped in the experiments reported later. As an illustration, the terms "effects" appears ahead of "spreading" in the window and their distance is 2-word. The value of cell (spreading, effect) can then be computed as: 5-2+1 = 4.

Table 1: Example of a HAL space

	the	effects	of	spreading	pollution	on	Population	Atlantic	salmon
the		1	2	3	4	5			
effects	5								
of	8	5		1	2	3	5		
spreading	3	4	5						
pollution	2	3	4	5					
on	1	2	3	4	5				
population	5		1	2	3	4			

atlantic	3		5		1	2	4		
salmon	2		4			1	3	5	

This table shows how the row vectors encode preceding word order and the column vectors encode posterior word order. For the purposes of this chapter, it unnecessary to preserve order information, so the HAL vector of a word is represented by the addition of its row and column vectors.

The quality of HAL vectors is influenced by the window size; the longer the window, the higher the chance of representing spurious associations between terms. A window size of eight or ten has been used in various studies (Burgess et al. 1998, Bruza and Song 2002, Song and Bruza 2001, Bai et al., 2005). Accordingly, a window size of 8 will also be used in the experiments reported in this chapter.

More formally, a concept[1] c is a vector representation: $c = < w_{cp_1}, w_{cp_2}, ..., w_{cp_n} >$ where $p_1, p_2, ..., p_n$ are called dimensions of c, n is the dimensionality of the HAL space, and w_{cp_i} denotes the weight of p_i in the vector representation of c. In addition, it is useful to identify the so-called *quality properties* of a HAL-vector. Intuitively, the quality properties of a concept or term c are those terms which often appear in the same context as c. Quality properties are identified as those dimensions in the HAL vector for c which are above a certain threshold (e.g., above the average weight within that vector). A dimension is termed a property if its weight is greater than zero. A property p_i of a concept c is termed a *quality property* iff $w_{cp_i} > \partial$, where ∂ is a non-zero threshold value. From a large corpus, the vector derived may contain much noise. In order to reduce the noise, in many cases only certain quality properties are kept. Let $QP_{\partial}(c)$ denote the set of quality properties of concept c. $QP_{\mu}(c)$ will be used to denote the set of quality properties above mean value, and $QP(c)$ is short for $QP_{\partial}(c)$.

HAL vectors can be normalized to unit length as follows:

$$w_{c_i p_j} = \frac{w_{c_i p_j}}{\sqrt{\sum_k w_{c_i p_k}^2}}$$

[1] The term "concept" is used somewhat loosely to emphasize that a HAL space is a primitive realization of a conceptual space

For example, the following is the normalized HAL vector for "*spreading"* in the above example (Table 1):

spreading = < the: 0.52, effects: 0.35, of: 0.52, pollution: 0.43, on: 0.35, population: 0.17 >

In language, word compounds often refer to a single underlying concept. As HAL represents words, it is necessary to address the question of how to represent a concept underpinned by more than a single word. A simple method is to add the vectors of the respective terms in a compound. In this article, however, we employ a more sophisticated concept combination heuristic (Bruza and Song 2002). It can be envisaged as a weighted addition of underlying vectors paralleling the intuition that in a given concept combination, some terms are more dominant than others. For example, the combination "GATT [2] Talks" is more "GATT-*ish"* than "talk-*ish*". Dominance is determined by the specificity of the term.

In order to deploy the concept combination in an experimental setting, the dominance of a term is determined by its inverse document frequency (*idf*) value. The following equation shows a basic way of computing the idf of a term *t*:

$idf(t) = \log(N/n)$ where *N* is the total number of documents in a collection and *n* is the number of documents which contain the term *t*.

More specifically, the terms within a compound can be ranked according to its *idf*. Assume such a ranking of terms: $t_1,\ldots,t_m$ ($m > 1$). Terms t_1 and t_2 can be combined using the concept combination heuristic resulting in the combined concept, denoted as $t_1 \oplus t_2$, whereby t_1 dominates t_2 (as it is higher in the ranking). For this combined concept, its degree of dominance is the average of the respective *idf* scores of t_1 and t_2. The process recurs down the ranking resulting in the composed "concept" $((..(t_1 \oplus t_2) \oplus t_3) \oplus \ldots) \oplus t_m)$. If there is only a single term (m =1), its corresponding normalized HAL vector is used as the combination vector.

We will not give a more detailed description of the concept combination heuristic, which can be found in (Bruza and Song 2002). Its intuition is summarized as follows:

- Quality properties shared by both concepts are emphasized.

[2] General Agreement on Tariffs & Trade is a forum for global trade talks.

- The weights of the properties in the dominant concept are re-scaled higher
- The resulting vector from the combination heuristic is normalized to smooth out variations due to differing number of contexts the respective concepts appear in.

By way of illustration we have the following vector for the concept combination "GATT talks":

gatt ⊕ talks = < agreement: 0.282, agricultural: 0.106, body: 0.117, china: 0.121, council: 0.109, farm: 0.261, gatt: 0.279, member: 0.108, negotiations: 0.108, round: 0.312, rules: 0.134, talks: 0.360, tariffs: 0.114, trade: 0.432, world: 0.114,>

In summary, by constructing a HAL space from text corpus, concepts are represented as weighted vectors in the high dimensional space, whereby each word in the vocabulary of the corpus gives rise to an axis in the corresponding semantic space. The rest of this chapter will demonstrate how the fuzzy C-Means clustering can be applied to conceptual induction and how different contexts are reflected.

3. Fuzzy C-Means Clustering

As the focus of this chapter is not the development of a new clustering algorithm, the fuzzy C-Means algorithm we use in our experiment is adapted from some existing studies in the literature (Hathaway et al. 2000, Krishnapuram et al. 2001).

Let $X = \{x_1, x_2, ...x_n\}$ be a set of *n* objects in a *S-dimensional* space. Let $d(x_j, x_i)$ be the distance or dissimilarity between objects x_i and x_j. Let $V = \{v_1, v_2, ..., v_K\}$, each v_c be the *prototype* or *mean* of the *c-th* cluster. Let $d(v_c, x_i)$ be the distance or dissimilarity between the object x_i and the mean of the cluster that it belongs to.

The fuzzy clustering partitions these objects into *K* overlapped clusters based on a computed minimizer of the fuzzy within-group least squares functional:

$$J_m(U,V) = \sum_{c=1}^{K} \sum_{i=1}^{N} U^m(v_c, x_i) d(v_c, x_i) \qquad (1)$$

where the minimization is performed over all $v_c \in V$, and $U(v_c, x_i)$ is the membership function for the object x_i belonging to the cluster v_c.

To optimize (1), we alternate between optimization of $\bar{J}_m(U \mid V^*)$ over U with V^* fixed and $\bar{J}_m(V \mid U^*)$ over V with U^* fixed, producing a sequence $\{U^{(p)}, V^{(p)}\}$. Specifically, the *p+1st* value of $V = \{v_1, v_2, ..., v_K\}$ is computed using the *p-th* value of U in the right-hand side of:

$$v_c^{(p+1)} = \frac{\sum_{i=1}^{N} x_i * [U^{(p)}(v_c^{(p)}, x_i)]^m}{\sum_{i=1}^{N} [U^{(p)}(v_c^{(p)}, x_i)]^m} \qquad (2)$$

Then the updated *p+1st* value of *V* is used to calculate the *p+1st* value of U via:

$$U^{(p+1)}(v_k^{(p+1)}, x_i) = \frac{d(x_i, v_k^{(p+1)})^{-1/(m-1)}}{\sum_{c=1}^{K} d(x_i, v_c^{(p+1)})^{-1/(m-1)}} \qquad (3)$$

Where $m \in (1, +\infty)$ is the fuzzifier. The greater *m* is, the fuzzier the clustering is. Krishnapuram et al. (2001) recommend a value between 1 and 1.5 for *m*. In addition, the following constraint holds:

$$\forall i \;\; i = 1,2,...,N \qquad \sum_{c=1}^{K} U(v_c, x_i) = 1 \qquad (4)$$

For the sake of efficiency in large datasets, an alternative method is to use the top $L (L < N)$ objects in the cluster, and the objects are sorted based on descending membership value:

$$v_c^{(p+1)} = \frac{\sum_{i=1}^{L} x_i * [U^{(p)}(v_c^{(p)}, x_i)]^m}{\sum_{i=1}^{N} [U^{(p)}(v_c^{(p)}, x_i)]^m} \qquad (5)$$

If the dissimilarity is inner product induced, i.e. Square Euclidean measure defined later in section 3.1, it can be proved mathematically that computing V and U iteratively

according to Equation (2) and (3) satisfies the necessary conditions for optima of $J_m(U \mid V)$ (Bezdek 1981).

The traditional K-Means clustering algorithm, namely the hard C-Means clustering, is a special case of fuzzy C-Means clustering by simply replacing (3) with:

$$q = \arg\min_c d(v_c, x_i) \qquad U^{(p+1)}(v_c, x_i) = \begin{cases} 1 & if \quad c=q \\ 0 & if \quad c \neq q \end{cases}$$

The fuzzy C-Means clustering algorithm is detailed as follows:

Fuzzy C-Means Algorithm:

Fix the number of clusters K and Max_iter; Set iter = 0;
Pick initial means V={v_1, v_2, …, v_K} from X;

Repeat

Compute memberships U(v_c, x_i) for c = 1, 2, …, K and
i = 1, 2, …, N by using Equation (3) (A)

Store the current means, V^{old} = V;

Re-compute the new means v_c for c = 1, 2, …, K by using Equation (2) (B)
Iter = Iter + 1;

Until

Iter = Max_iter or

The absolute value of increment of the objective function $|\Delta J(U,V)| < \varepsilon$, where ε is some prescribed tolerance.

3.1 Dissimilarity Measures

Several measures can be employed to compute the dissimilarity between two objects (x_j, x_i), as well as between an object and the mean (v_c, x_i). The most frequently used approach is the Lp norm distance, which is defined as follows (Hathaway et al. 2000):

$$d(v_c, x_i) = \left(\sum_{j=1}^{S} | x_{i,j} - v_{c,j} |^p \right)^{1/p}$$

where $p \in [1, +\infty)$ and S is the dimensionality of the vectors. This is a generalized dissimilarity measure. By way of illustration, Euclidean distance corresponds to the case when $p = 2$:

$d(v_c, x_i) = \|x_i - v_c\| = \sqrt{\sum_{j=1}^{S} (x_{i,j} - v_{c,j})^2}$, it is used in (Bobrwoski and Bezdek 1991)

If $p = 1$, Manhattan dissimilarity results:

$d(v_c, x_i) = \sum_{j=1}^{S} | x_{i,j} - v_{c,j} |$.

Moreover, if $p = \infty$:

$d(v_c, x_i) = \underset{j=1}{\overset{S}{Max}} | x_{i,j} - v_{c,j} |$.

Hathaway et al (2000) have shown that $p = 1$ or 2 offers the greatest robustness for outlier handling.

In addition, other widely used dissimilar measures are:

Squared Euclidean: $d(v_c, x_i) = \|x_i - v_c\|^2 = \sum_{j=1}^{S} (x_{i,j} - v_{c,j})^2$

Cosine based dissimilarity: $d(v_c, x_i) = e^{-Sim(v_c, x_i)}$, where $Sim(v_c, x_i)$ is defined as:

$$Sim(v_c, x_i) = \frac{\sum_{j=1}^{S} x_{i,j} * v_{c,j}}{\sqrt{\sum_{j=1}^{S} x_{i,j}^{\ 2} \sum_{j=1}^{S} v_{c,j}^{\ 2}}}$$

3.2 Initialization

Initialization is vital to the performance of Fuzzy *C-Means* algorithm. Though we stated in the beginning of section 3 that the algorithm satisfies the necessary conditions for optima of the objective function $J_m(U|V)$, the Fuzzy *C-Means* algorithm is not guaranteed to find the global minimum. Different initialization procedures will produce slightly different clustering results. Nevertheless, appropriate initialization will make the algorithm converge fast. If the *K* means are initialized randomly, it is desirable to run the algorithm several times to increase the reliability of the final results. We have experimented with two different ways of initializing the *K* means. The first way is to pick all the means candidates randomly. This method is referred to as *Initialization 1*. The second way is to pick the first candidate as the mean over all the items in the space X, and then each successive one will be the most dissimilar (remote) item to all the items that have already been picked. This makes the initial centroids evenly distributed. We refer to this procedure as Initialization 2.

Initialization 2 for Fuzzy C-Means Clustering

Fix the number of means K > 1;

Compute the first mean $v_{1:}$

$$v_1 = \frac{\sum_{i=1}^{N} x_i}{N}$$

Set V = { v_1 }, iter = 1;

Repeat

iter = iter + 1;

$$v_{iter} = \max_{\substack{1 \le j \le N \\ x_j \notin V}} (\min_{1 \le k \le |V|} d(x_j, v_k)) \quad then \quad V = V \cup \{v_{iter}\}$$

Until

iter = K;

For a given data set, the initial produced by Initialization 2 is fixed. In our experiments, Initialization 2 outperforms Initialization 1 consistently.

4. Word Clustering on a HAL Space – A Case Study

This experiment aims to illustrate the effectiveness of the fuzzy C-Means approach for clustering concepts (words) represented as HAL vectors.

HAL Space Construction

We applied the HAL method to the Reuters-21578 collection, which consists of new articles in the late 1980s. The vocabulary is constructed by removing a list of stop words and also dropping some infrequent words which appears less than 5 times in the collection. The size of final vocabulary is 15415 words. The window size is set to be eight. A too small window leads to loss of potentially relevant correlations between words, whereas a too large window may compute irrelevant correlations. We think window size of eight is reasonable since precision is our major concern. Previous studies in HAL (Lund and Burgess 1996; Song and Bruza 2003) have also employed a window size of eight in their experiments.

Table 2: The *Iran* vector.

Iran	
Dimension	Value
arms	0.64
iraq	0.28
scandal	0.22
gulf	0.18
war	0.18
sales	0.18
attack	0.17
oil	0.16
offensive	0.12
missiles	0.10
reagan	0.09
...	...

HAL vectors are normalized to unit length. As an example, the following is part of the cosine-normalized HAL vector for "*Iran"* computed derived from applying the HAL method to the Reuters-21578 collection. This example demonstrates how a word is represented as a weighted vector whose dimensions comprise other words. The weights represent the strengths of association between "Iran" and other words seen in the context of the sliding window: the higher the weight of a word, the more it has lexically co-occurred with "Iran" in the same context(s). The dimensions reflect aspects which were relevant to the respective concepts during the mid to late eighties. For example, Iran was involved in a war with Iraq, and President Reagan was involved in an arms scandal involving Iran.

Data

The following twenty words were selected from the vocabulary to prime the clustering process: *airbus, boeing, plane, Chernobyl, nuclear, disaster, computer, nec, japan, ibm, contra, industry iran, iraq, scandal, war, president, reagan, white, house*.

Table 3: Handcrafted Result

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
Airbus	Chernobyl	Computer	White	**Iran**
Boeing	Disaster	Nec	House	Scandal
Plane	Nuclear	Ibm	President	**Contra**
Industry		**Industry**	**Reagan**	**Reagan**
			Iraq	**War**
			War	**Reagan**
			Iran	**Iraq**
			Japan	
			Industry	

Table 3 summaries a manual clustering of the above words. These words involve approximately the following contexts in the Reuters collection:

(1) Aircraft manufacturers;

(2) Chernobyl nuclear leaking disaster in the Soviet Union;

(3) Computer companies;

(4) The roles of the White House (i.e., Reagan government) in the middle 1980s (dealing with Iran-Iraq war and trade war against Japan);

(5) The Iran-contra scandal (President Reagan was involved in the illegal arms sales to Iran during the Iran-Iraq war).

Note there is some overlap between clusters. For example, cluster 4 shares "industry" with clusters 1 and 3; it also shares "reagan" and "iran" with cluster 5, etc.

Fuzzy Clustering of HAL Vectors

In order to find the best performing parameter settings for the fuzzy C-Means clustering, we have developed a test bed on which a series of prior studies have been conducted. Cosine combined with fuzzifier 2.0 and Initialization 2 was finally chosen after some initial pilot studies. When the membership value of a word belonging to a cluster is greater than a prior probability (0.2 for this experiment), it is output as a member in the cluster. The following table lists the result of fuzzy C-Means clustering (the number following each word is the membership value of the word belonging to the corresponding cluster).

Table 4: Clustering Result of Fuzzy C-Means Algorithm

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
Airbus: 0.914	Chernobyl: 0.966	Computer: 0.921	White: 0.861	Iraq: 0.869
Boeing: 0.851	**Disaster: 0.302**	Nec: 0.906	House: 0.793	Scandal: 0.814
Plane: 0.852	Nuclear: 0.895	Ibm: 0.897	President: 0.653	**Contra: 0.776**
			Reagan: 0.708	**Iran: 0.725**
			Japan: 0.558	**War: 0.584**
			Industry: 0.494	**Reagan: 0.213**
			Disaster: 0.488	
			War: 0.331	
			Iran: 0.221	
			Contra: 0.203	

We also conducted experiments with K-Means algorithm on the same data and the best performing result (via Cosine-based dissimilarity function) is depicted below:

Table 5: Clustering Result from K-Means Algorithm

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Airbus	Chernobyl	Computer	Contra	Nuclear
Boeing	Disaster		House	
IBM			Iran	
Industry			Iraq	
Japan			President	
Nec			Reagan	
Plane			Scandal	
			War	
			White	

Discussions:

Table 4 shows that the fuzzy clustering results basically reflect the underlying contexts described in Table 3, particularly the overlap between Reagan government, Iran-Iraq war and Iran-Contra scandal.

However, the K-Means clustering result presented in Table 5 is less ideal: Cluster 1 contains the words related to industry, either plane-manufacturing or IT; "Nuclear" is separated from the "Chernobyl Disaster"; "Computer" forms a singular cluster; Cluster 4 contains terms related to politics.

In short, the results from the case study suggest that Fuzzy K-Means clustering of word "meanings" in a HAL space is promising.

5. Conclusions and Future Work

In this chapter, we have introduced a cognitively motivated model, namely Hyper-space Analogue to language (HAL), to construct a high dimensional semantic space. The HAL space can be used to realize aspects of Gärdenfors' conceptual space theory dealing with the geometrical representation of information. Within the conceptual space, concepts can be categorized into regions reflecting different contexts. A fuzzy C-Means algorithm has been investigated in detail for concept induction with the advantage that an acknowledged weakness, namely, the crispiness of the traditional K-Means method is overcome. We present a case study on word clustering in a HAL space which is constructed from the Reuters-21578 corpus. The case, though preliminary, suggests that the fuzzy C-Means algorithm is encouraging.

The work presented in this article can potentially be extended to other areas, such as query expansion of information retrieval, web page clustering, etc. Furthermore, we will conduct formal evaluation of the algorithm based on larger collections in the future.

Acknowledgement

This chapter is an extended version of our previous work (Cao et al. 2004).

References

Bai J, Song D, Bruza PD, Nie, JY and Cao G 2005 Query expansion using term relationships in language models for information retrieval. In *Proc. of the 14th Int. Conf. on Information and Knowledge Management (CIKM'2005)*, pp. 688-695.

Bezdek JC 1981 *Pattern Recognition with Fuzzy Objective Function Algorithms.* New Yok: Plenum.

Bobrowski J and Bezdek JC 1991 C-Means clustering with the L_1 and L_∞ norms. *IEEE Trans. Syst. Man, Cybern.* 21, 545-554.

Bruza PD and Song D 2002 Inferring query models by computing information flow. In *Proc. of the 12th Int. Conf. on Information and knowledge Management (CIKM2002)*, pp. 260-269.

Burgess C, Livesay L and Lund K 1998 Explorations in context space: words, sentences, discourse. In Quantitative Approaches to Semantic Knowledge Representation (ed. Foltz PW), *Discourse Processes*, 25(2&3), 179-210.

Cao G, Song D and Bruza PD 2004 Fuzzy K-means clustering on a high dimensional semantic space. In *Proceedings of the 6th Asia Pacific Web Conference (APWeb'04)*, LNCS 3007, pp. 907-911.

Chuang SL and Chien LF 2005 Taxonomy generation for text segments: a practical web-based approach. *ACM Transactions on Information Systems (TOIS)*, 23(4), 363-396.

Cimiano P, Hotho A and Staab S 2005 Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research* 24, 305-339.

Deerwester S, Dumais ST, Furnas GW, Landauer TK and Harshman R 1990 Indexing by latent semantics analysis. *Journal of the American Society for Information Science* 41(6), 391-407.

Gärdenfors P 2000 *Conceptual Spaces: The Geometry of Thought*. MIT Press.

Gärdenfors P and Williams M 2001 Reasoning about categories in conceptual spaces. In *Proceedings of 14th International Joint Conference of Artificial Intelligence (IJCAI'2001),* pp. 385-392.

Hathaway RJ, Bezdek JC and Hu Y 2000 Generalized fuzzy C-means clustering strategies using Lp norm distances. *IEEE Transactions on Fuzzy Systems* 8, 576–582.

Hearst M 2003 What is text mining? http://www.sims.berkeley.edu/~hearst/text-mining.html

Höppner F, Klawonn F, Kruse R and Runkler T 1999 *Fuzzy Cluster Analysis.* John Wiley & Sons.

Höppner F and Klawonn F 2003 A contribution to convergence theory of fuzzy c-means and derivatives. *IEEE Transactions on Fuzzy Systems* 11(5), 682-694.

Kolen J and Hutcheson T 2002. Reducing the time complexity of the fuzzy c-means algorithm. *IEEE Transactions on Fuzzy Systems* 10(2), 263-267.

Krishnapuram R, Joshi A, Nasraoui O and Yi Y 2001. Low-complexity fuzzy relational clustering algorithm for web searching. *IEEE Transactions on Fuzzy Systems* 9(4), 595-607.

Landauer T and Dumais S 1997 A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2), 211-240.

Lund K and Burgess C 1996 Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research Methods, Instruments, & Computers* 28(2), 203-208.

Perrin P and Petry F 2003 Extraction and representation of contextual information for knowledge discovery in texts. *Information Sciences* 151, 125-152.

Srinivasan P 2004 Text mining: generating hypotheses from MEDLINE. *Journal of the American Society for Information Science and Technology* 55(5), 396-413.

Song D and Bruza PD 2003 Towards context sensitive informational inference. *Journal of the American Society for Information Science and Technology* 52(4), 321-334.

Song D and Bruza PD 2001 Discovering information flow using a high dimensional conceptual space. In *Proceedings of the 24th Annual International Conference on Research and Development in Information Retrieval (SIGIR'01*), 327-333.

Steinbach M, Karypis G and Kumar V 2000 A comparison of document clustering technique. In *KDD'2000 Workshop on Text Mining*. Available online: www.cs.cmu.edu/~dunja/KDDpapers/Steinbach_IR.pdf