ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Modeling contextualized textual knowledge as a Long-Term Working Memory

Mauro Mazzieri, Sara Topi, Aldo Franco Dragoni, Germano Vallesi

Univiversità Politecnica delle Marche
Dept. of Computer, Management and Automation Engineering
via Brecce Bianche 22, 60100 Ancona - Italy
{m.mazzieri, s.topi, a.f.dragoni, g.vallesi}@univpm.it

Abstract.
A knowledge management system is more than an archive of textual documents; it provides context information, allowing to know which documents where used by people with a common goal. In the hypothesis that a set of textual documents with a common context can be assimilated to the long term memory of a human expert executor, we can use on them mining techniques inspired by the mechanic of human comprehension in expert domains. Text mining techniques for KM task can use a model of the long-term memory to extract meaningful keywords from the documents. The model acts as a dynamic and non-stationary dimensionality reduction strategy, allowing the clustering of context documents according to keyword presence, the classification of external documents according to local criteria, and a better understanding of document content and relatedness.

1 Introduction: the context actor hypothesis

A realistic hypothesis of the mechanics of human text comprehension is that, in expert domains, we can use a learned network of concept relations to automatically retrieve concepts related to the task at hand.

Textual sources in some local context, such as documents accessed by a group of people while performing a certain task, can be considered as a sort of collective expertise. As such, documents can be processed for mining purposes according to the mechanism of human text comprehension in expert domains.

A knowledge management system joining document management with workflow management allows to extract information about document used while executing a task or by actor with a common role; the task or the role are the context of the respective document groups.

Our claim is that a set of document in a local context can be considered as a *long term memory* about an *area of expertise* (i.e. the local context). We can consider an hypothetical ideal *context actor* representing the local context instead of a specific human actor.

We show how the formation of a scale-free concept network can be used as a dynamic and non-stationary dimensionality reduction strategy to allow the extraction of subset of concepts related to the task, allowing a fast categorization of the documents, a fast classification of external documents according to local criteria, and moreover a better understanding of document content and relatedness.

The paper is organized as follows. The human long-term working memory and its computer simulation are discussed in Section 2 and 3. The structure of a knowledge management system, and its relation with text mining techniques, is discussed in Section 4. Experimental result on the formation of memory models are in Section 5. Section 6 shows applications in a knowledge management software and Section 7 propose future works.

2 The long-term working memory

The human mind, in analyzing new information, generates dynamic structures that are adapted to the particular context of use. The information nodes form a network, and the meaning of a node is given by its position in the network. When involved in a cognitive task, not all nodes are used, but only those activated by the particular context.

The temporary storage of information being processed in any of a range of cognitive tasks is called Short-Term Working Memory (STWM). STWM is available under all conditions, but is severely capacity-limited: the focus is only on the last read sentence or paragraph. The Long Term Memory (LTM) is the opposite of STWM: it is very large, comprising much of what was previously learnt.

A realistic hypotheses about the mechanics of human text comprehension is that humans, in skilled activities, use, together with short-term working memory, some of the content of long term memory. According to the theory of the Long-Term Working Memory (LTWM) [1, 2], the difference in human text comprehension when made by an expert in its domain of expertise, compared to comprehension by a non-expert, lies in the role of the long term memory: outside the domain of expertise, comprehension relies only on the content of the short-term working memory. The long-term working memory can be defined as the subset of the long-term memory composed by concepts directly retrievable via cues in short-term working memory. Retrieval cues are stable memory structures, formed during training, linking nodes in the short-term working memory with other nodes in the long-term memory.

For the long term memory to be accessible, the retrieval cues must be previously formed: in a field where a previous expertise exists, the content of the long term memory can be used, via those cues in short term memory. If activation links between concept in the short-term and long-term memory are present, the retrieval of long-term memory content is fast and automatic. The need of previously retrieval cues explains why LTWM is available only in expert domains.

3 Memory model

An artificial simulation of the human memory involves the creation of an associative network of terms.

A study of human brain activity using fMRI [3] shows that the functional network of correlated human brain sites has a small-world character [4]: they have a small average path length among nodes, comparable with those of random networks, but are highly clustered, orders of magnitude more than random networks. Moreover, the connectivity degree is scale-free [5]: the probability that a node as k links has a power-low tail distribution for large k, following $p(k) \approx k^{-\gamma}$. We expect that an artificially-built memory model must show similar characteristics.

The first memory models were built using Latent Semantic Analysis (LSA) [2], a statistical learning model. A matrix of word frequency in paragraphs is decomposed using Singular Value Decomposition (SVD); then, only the biggest singular values are retained. Words, sentences or whole concepts are expressed as vectors in the reduced space, and the relatedness of two concepts is given by the cosine of their vectors. This model is given by a static elaboration on the entire corpus.

An alternative memory model, proposed by Licata *et. al.* [6], use a scale-free graph as model. The model is incrementally built as new documents are read; the network allows each node (representing a word) to form a link with other M units. The probability for a word i to form a link with word j is given by [7]:

$$p_j = \frac{d_j \mu_j}{\sum_k d_k \mu_k}$$

where d_k is the degree (number of existing links) of node k and μ is a fitness value. The fitness valued is evaluated parsing the text into paragraphs and calculating the ratio of paragraphs containing both i and j over all paragraphs containing either i or j.

After the formation of the model of short-term memory, the long term memory is updated using its content: if a link is present is the STM, the weight of the corresponding link in the LTM is incremented. Finally, memory link weights for each words are normalized; the normalization has the effect of an exponential time discount of previously created links, thus emphasizing new links or recently refreshed links over old links.

4 Knowledge Management provides context to documents

A knowledge management system manages formal or tacit knowledge inside an organization [8]. For the present work, we are interested in two components of a knowledge management system: the document management system (DMS) and the work-flow management system (WFMS).

A document management system (DMS) track and store electronic documents and digitalized version of non-electronic documents. Usually a DMS can provide an extensive logging, including which user accessed which document and when. The structure of an enterprise document management system naturally follows the functional organization of the enterprise: documents can be organized in a hierarchical taxonomy, according to their subject, author, confidentiality, purpose.

A workflow management system (WMS) depicts organizational processes as a sequence of tasks to be completed. Each task is the work of a person, or a group of persons, or of a simple or complex mechanism. A generic entity able to complete a task is generically referred to as an *actor*.

The WMS allow a decomposition of knowledge in local contexts: according to a structural or functional division, actors can be classified according to common skills or similarity in completed tasks; conversely, task can be classified according to required skills or according to the actor completing them.

If DMS and WMF are closely integrated inside a knowledge management system, document usage data (from DMS) can be matched with the task or process currently executed (from the WMS), allowing a decomposition of the document base in overlapping local contexts.

We can define a context as composed of either the documents accessed by actors performing a task related to a certain process, or the documents accessed by actors with common skills; actor with common skill are usually grouped in a KMS into a *role*.

If we form a model of long-term memory using the content of documents related to a context, we expect to find a structure reflecting the hidden knowledge common to actors in role or performing a process.

5 Experimental results

To test the validity of the context actor hypothesis, we built a scale-free model of the long-term memory from some collection of text documents, then compared the properties of the resulting graphs.

The first model is built using the Italian section of Reuters Corpus, volume 2 (Multilingual) [9]. This model is used only to obtain a baseline; the news have a common structure and a uniform length, but are otherwise loosely related.

The other sets are extracted from the document archive of a knowledge management system. The knowledge management software[1] is used by a software development company. We extracted two sets: the first composed by documents recently accessed by employee in the administrative department, and the second composed by documents recently accessed by members of the software development team.

Figure 1 shows average path length, clustering coefficient and average node degree for test data. The average path length measures the average number of links in the shortest path between two words. The clustering coefficient measures the cliquishness of a typical neighborhood: if a word has k links, then there are at most $k(k-1)/2$ links among them; the coefficient measure the average fraction of those links that actually exists.

A graph is scale-free if, compared to a random network with the same number of words, the average path length is comparable but the clustering coefficient is much greater. A random graph of N words would have an average path length of pN, where p is the probability that two random words have a link, and the

[1] josh, a product of itConsult srl, Fermignano (PU), Italy

data set	agv. path length	clustering coeff.	avg. degree
Italian Reuters corpus	2.35	0.72	18.64
Commercial	2.25	0.71	15.85
Technical	2.11	0.85	32.70
both	2.17	0.82	29.98

Fig. 1: Scale-free properties of LTM's graph model.

clustering coefficient is $1/N$. For comparison, there are 1598 unique words in the first 100 Reuters news, so the clustering coefficient of an equivalent random graph would be $\simeq 6.26 \cdot 10^{-3}$.

The set of technical documents are more strictly related than the document used by the administrative team, and this is reflected by the shortest path and greater clustering.

6 Current application

The memory model is currently used inside a knowledge management software. Text extracted from documents belonging to a context is first filtered, stripping suffixes [10] and removing an hand-built set of very common words (prepositions and conjunctions). Then, the long-term memory of the document's context is updated. Documents are represented by a feature vector in a reduced space composed by the most linked words.

Feature vectors are used to train a self-organizing map (SOM)[11], were nodes represents a set of documents. From the map, a hierarchy of categories is extracted according to the algorithm described in [12]: top level categories are selected as SOM nodes maximizing similarity with neighbors nodes; the other nodes are clustered around those dominating nodes. A recursive application yields a hierarchical taxonomy.

The local memory model and categorizations are batch elaborated: starting from the log of the KM software, models and categories related to documents recently added or changed are updated.

The same SOM is used to classify documents, even if they don't belong to the local context, according to their similarity with nodes representing clusters. The categorization of external document according to a local selection of keywords in local categories allows a personalized browsing of an external archive. Also, the classifier is used to post-process enterprise search results, allowing to show search results according to the local taxonomy.

We experimented also the use of the local taxonomy for an external web search. However, in this case, the representation of search result according to frequencies of local keywords is not a reliable representation, as many results contains few or any of the local keywords.

7 Future work

Our research is currently directed toward the building of a better memory model.

A LSA model built from a batch elaboration of document set would not entirely fit our purposes, as it completely ignores the order of the documents given as input, and require for each new document a global recalculation of the reduced space. The scale-free model, albeit incrementally built, as a time complexity $O(n^2)$, where n is the number of words in the dictionary, for the memorization of a new documents, as require the matrix of relative word co-occurrence to be recalculated. Also, it allows the extraction of a reduced space only by selecting the most linked words.

For those reasons, we are developing a dynamical version of the LSA. The models emulates an auto-associative neural network, whose hidden layer are a internal representation of the archive. The model is incrementally built as new document arrives, using an incremental version of the SVD. We are comparing the outcome of the different models. Finally, we are looking for a way to deliberately approximate this model with a bias towards recently added rows.

References

[1] K. A. Ericsson and W. Kintsch. Long-term working memory. *Psychological Review*, 102:211–245, 1995.

[2] W. Kintsch, V. L. Patel, and K. A. Ericsson. The role of long-term working memory in text comprehension. *Psychologia*, 42:186–198, 1999.

[3] V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian. Scale-free brain functional networks. *Physical Review Letters*, 94(1):018102, Jan 2005.

[4] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 4 June 1998.

[5] A. L. Barabàsi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 15 October 1999.

[6] I. Licata, G. Tascini, L. Lella, A. Montesanto, and W. Giordano. Scale free graphs in dynamic knowledge acquisition. In *Systemics of Emergence: Research and Development*, chapter 9, pages 615–624. Springer US, January 16 2006.

[7] G. Bianconi and A. L. Barabási. Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86(24), 2001.

[8] I. Nonaka and H. Takeuchi. *The Knowledge-Creating Company*. Oxford University Press, 1996.

[9] D. D. Lewis, Y. Yang, T. Rose, and F. Li. Rcv1: A new benchmark collection for text categorization research. *J. of Machine Learning Research*, 5:361–397, 2004.

[10] M. F. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[11] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Series in Information Sciences*. Springer-Verlag, Berlin, Germany, 2nd edition, 1997.

[12] H. C. Yang and C. H. Lee. Automatic category generation for text documents by self-organizing maps. In *Proc. of the Int. Joint Conf. on Neural Networks*, volume 3, pages 581–586, Piscataway, NJ, 2000. IEEE.