OPERATIONS ON CONVERSATIONAL MIND-GRAPHS

Jayanta Poray and Christoph Schommer
Department of Computer Science and Communication, University of Luxembourg
6, Rue Coudenhove-Kalergi, L-1359, Luxembourg City, Luxembourg

Keywords: Adaptive information management, Learning, Associative memories, Linguistic processing, Graph operations.

Abstract: Mind-graphs define an associative-adaptive concept of managing information streams, like for example words within a conversation. Being composed of vertices (or cells; representing external stimuli like words) and undirected edges (or connections), mind-graphs adaptively reflect the strength of simultaneously occurring stimuli and allow a self-regulation through the interplay of an artificial ‘fever’ and ‘coldness’ (capacity problem). With respect to this, an interesting application scenario is the merge of information streams that derive from a conversation of k conversing partners. In such a case, each conversational partner has an own knowledge and a knowledge that (s)he shares with other. Merging the own (inside) and the other’s (outside) knowledge leads to a situation, where things like e.g. trust can be decided. In this paper, we extend this concept by proposing extended mind-graph operations, dealing with the *merge* of sub-mind-graphs and the *extraction* of mind-graph skeletons.

1 INTRODUCTION

Today, the exchange of textual information by electronic devices is very popular. It ranges from simple short-term messages to collections of long-term conversations, which have obtained by several months or years. And in fact, the produce of textual information within a conversation is a non-deterministic process, which requires a linguistic preparation of the texts, and a computational finesse, if the generated information is to be accumulated or summarised with regards to the content. As one of the most promising research topic in the next years, the exploration of chats inside social networks belongs to this category (Tuulos and Tirri, 2004).

Some research works have been done in the field of information accumulation, but the handling of a dynamic conversation within an adaptive framework has mostly been solved by associative graphs and the representation of information within these graphs. For example, a text summarisation method *LexRank* has been suggested by (Radev, 2004), where each vertex corresponds to the extracted topic from the input text and connection to the relation between several topics. In (Poray and Schommer, 2009) it has been shown that each conversing person can receive an understanding of its partner, if all incoming textual stimuli are linguistically processed and then put to an associative framework (mind-graph): the idea is that strong and weak connections – which emerge depending on the intensity and frequency of the signals – then finally lead to even such associative mind-graphs, which do not only reflect a textual conversation but moreover support a mental representation of the conversing partner.

As per the general continuation of this approach, a refinement concerns the categorisation of the information – which occurs during a conversation – into several categories, which we call a) *known*, b) *mutual*, and c) *unknown* information (Figure 1). Category a) refers to information that is already aware by a person before a conversation takes place and that is already inside the associative mind-graph; b) refers to a common information between several conversing partners: it evolves over time and is then sent to the associative framework. Finally, c) *Unknown* information refers to information, which is not aware by a person before the conversation.

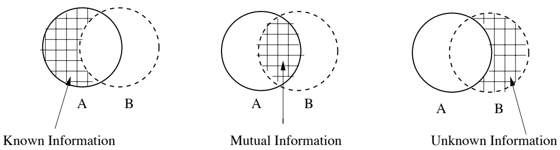

Figure 1: Three types of information during textual conversation (explanation see in text).

In this context, it is fair to confront the mind-graph

Poray J. and Schommer C..
OPERATIONS ON CONVERSATIONAL MIND-GRAPHS.
DOI: 10.5220/0003749905110514
In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence* (ICAART-2012), pages 511-514
ISBN: 978-989-8425-95-9

framework to the Extended Mind Theory, which has been presented by (among others, but mainly) (Clark and Chalmers, 1998). Here, it is suggested that external entities should be handled separately in the form of an *active externalism* and that mostly the internal and external information can not be separated, but considered as a *coupled system*.

2 RELATED WORK

The role of machine learning for automated text classification (or categorization) is discussed (among others) in (**?**) and (Ikonomakis et al., 2005). They clearly point out that an intersection of research fields like information extraction, text retrieval, summarization, question-answering et cetera exist particularly in those cases, where the inductive process of learning has been motivated by the texts. With respect to the set-up of the relation among social communities, (Ziegler and Golbeck, 2006) show how today's online communities allow their users to find the co-relation to measure trust and interests.

In association with the representation of the text based models, graphs are proven very useful. As an example (Haghighi et al., 2005) has developed a novel graph matching model for sentence inference from texts. Many related approaches regarding the graph representation for texts and documents have been proposed since last few years (e.g., (Schenker et al., 2003) and (Hensman, 2004)). Recently, (Jin and Srihari, 2007) used a novel graph based text representation model capable to capture *a) term order b) term-frequency c) term co–occurrance,* and *d) term context* in a document; then test has been performed for a specific text mining task. The state–of–the–art of our proposed graph similarity based text representation model is mostly motivated by these research efforts.

3 OPERATIONS ON MIND-GRAPH

With respect to the life-cycle process of a textual information, raw conversational text data is treated firstly as the linguistic pre-processer. This includes a tokenization, the elimination of stop-words, a resolution of pronouns, and others. A temporary storage space, termed as Short Term Memory (STM), which contains this filtered information, is then used. For each set of conversational text (document), information is represented as an undirected graph ((Jin and Srihari, 2007)), called mind-graphs (g=$\{V,E\}$), which on their way represent a pre-processed conversational text by a set of vertices (V) and a set of weighted edges (E).

3.1 Inside and Outside

A mind-graph assimilates textual components and assigns each component to a vertex. Components that occur together are consequently bidirectionally connected and logistically managed inside a *STM* (Short Term Memory). It is connected to the *LTM* (Long Term Memory), which administers those mind-graphs that have proven a stability over time ((Poray and Schommer, 2010)). Concerning the "self" (*Inside*) and the "other" (*Outside*) as mentioned in section 1, Figure 2 reflects this situation, where each conversational partner keeps an own information as *Inside* and newly obtained information as *Outside*.

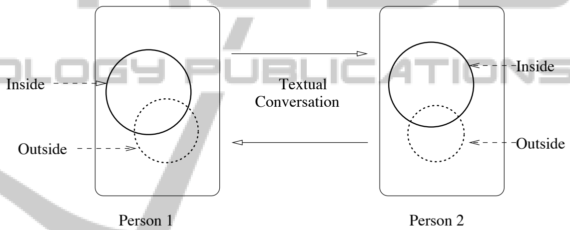

Figure 2: During textual conversation each person store their information as *Inside* and *Outside*.

3.2 Merge

In a conversation between two or more individuals, the merge of the (prepared) stimuli is one of the essential task and consequence. Here, the grade of similarity between more than one type of information (see Section 1) has to be considered. In an extended situation of Figure 1, a merge (function) actually represents the similarity (intersection) between two types of information.

Definition 1. A merge function μ for two mind-graphs $g \in G$ and $g^{'} \in G$ (where G is the set of mind-graphs) is defined as a *one-to-one* mapping among them, estimating the maximum common (similar) attributes among two mind-graphs.

$$\mu(g,g^{'}) : g \to g^{'}$$

This similarity measurement using merge function reflects the amount of mutual information between *Inside* and *Outside*.

A mind-graph $g = (V,E,\lambda,\Delta)$ consists of a set of vertices (V) and a set of edges (E). Here, $\lambda : V \to L_V$ represents the identifier of a vertex, such that

$\lambda(m) \neq \lambda(n), \forall m, n \in V, m \neq n$ and $\Delta : E \rightarrow \mathbb{R}^+$, where Δ is the number of traffic observed in the labeled graph g. Here, each mind-graph can be considered as the assemble of the different sub-graphs. Therefore, for n different mind-graphs $g^1, \ldots, g^n$, a mutual similarity is given by their *maximum common sub-mind-graph* ($mcs(g^1, \ldots, g^n)$. Following (Bunke and Shearer, 1998), the *distance* (d) between these two mind-graphs is then defined as follows:

$$d(g^1, \ldots, g^n) = 1 - \frac{|mcs(g^1, \ldots, g^n)|}{max\{|g^1|, \ldots, |g^n|\}}$$

As an example (see Figure 3), for two mind-graphs g and $g^{'}$ having the number of vertices $|g|$, $|g^{'}|$ respectively, also $|mcs(g, g^{'})|$ represent the number of vertices for their maximum common sub-mind-graph $mcs(g, g^{'})$. Since $|g| = 6$, $|g^{'}| = 4$ and $|mcs(g, g^{'})| = 3$, $d(g, g^{'}) = 0.5$. Having the similarity as the complement of the the distance d between g and $g^{'}$, the relation is then

$$\mu(g, g^{'}) = 1 - d(g, g^{'})$$

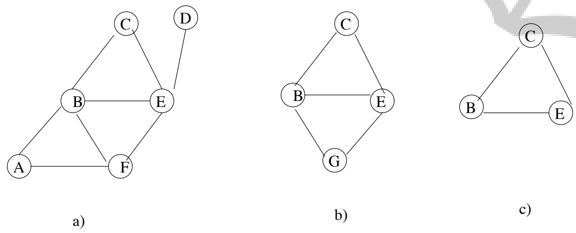

Figure 3: Mind-graphs a) g and; b) $g^{'}$; and their c) maximal common sub-graph, $mcs(g, g^{'})$.

3.3 Extraction of Skeletons

As described in (Poray and Schommer, 2009), a *skeleton* is a mind-graph with strong connections (threshold) between its vertices. The extraction procedure follows an algorithm, which is described in the following. It follows two steps. In a first step, the graph *potential* for the weighted mind-graph is computed whereas in the second step the actual extraction is done:

(Step1)

Require: The set of weighted nodes $\{\omega_1, \ldots, \omega_m\}$
Require: The set of weighted connections $\{\omega_{ij}\}, i, j \in \{1, \ldots, m\} \& i \neq j$
$\alpha_p \leftarrow \frac{1}{m} \sum_{i=1}^{m} \omega_i$
$\beta_p \leftarrow \frac{1}{n} \sum_{k=1}^{n} \omega_{ij}$, where, $n = |\omega_{ij}|$, $i, j \in \{1, \ldots, m\}$ and $i \neq j$
Compute the graph potential : $\delta_p \leftarrow \frac{\alpha_p + \beta_p}{2}$

and (Step 2)

Require: The skeleton threshold δ_s
if $\delta_p \geq \delta_s$ **then**
Mind-graph $g_p \leftarrow$ skeleton
if $g_p \leftarrow$ skeleton **then**
$f(g_p) : g_p(\text{STM}) \rightarrow g_p(\text{LTM})$ fiers
end if
end if

First, α_p, which denotes the average value of all weighted nodes ω_i and β_p, which denotes the average value of all weighted connections ω_{ij}, are considered. Thereafter, the graph potential δ_p, which is the overall weighted average of that graph is obtained. Then (in second step), the graph potential δ_p is compared with some pre-defined skeleton threshold δ_s and finally as per its grade, the skeletons are identified.

3.4 Mind-graph Normalization

Sometimes, the mind-graphs need to be managed properly, such that the *complexity* of the graph always keep below a certain threshold value and maintain its healthy status. To keep the mind-graph in a consistent (normalized) state, it is advisable to consider only the connections, which do not a threshold value. The process of *decomposition, join* and *selection* is motivated to resolve this issue.

As an example in figure 4 the mind-graph *G1* is estimated as the "over graph threshold value" with too many complex connections. First this is decomposed into five subgraphs *A,B,C,D* and *E*. Suppose among these five sub-graps *B* and *C* are again pointed as "over graph threshold value". Therefore, only rest three sub-graphs are taken for *join*. Similarly, in the *selection* phase the mind-graphs *AD* and *DE* are considered. Again the mind-graph *AE* not selected as it is in "over graph threshold value" state.

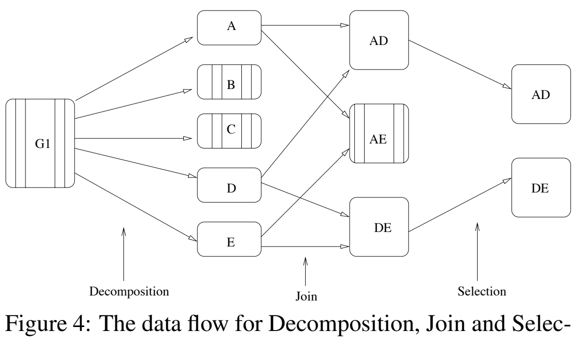

Figure 4: The data flow for Decomposition, Join and Selection for the mind-graph.

The algorithm for this described technique is presented below, where only the candidate mind-graphs g_r with "below graph threshold value" are stored inside the graph storage stack $\tilde{G}$.

Require: The mind-graph g_p with its set of weighted nodes $\tilde{N} = \{\omega_1, \ldots, \omega_m\}$ and their weighted connections $\tilde{C}$.
Ensure: $\tilde{G} \leftarrow \emptyset$

Compute the graph potential δ_p of the input graph g_p
for $r = 2$ to m **do**
  $\tilde{N} \leftarrow \{\omega_i\}$, where $|\tilde{N}| = r$ and $i \in \{1, \ldots, m\}$
  $\tilde{C} \leftarrow \{\omega_{ij}\}$, where $i, j \in \{1, \ldots, m\} (i \neq j)$
  $\delta_r \leftarrow$ Compute graph potential for all subgraphs formed with $\tilde{N}$ and $\tilde{C}$
  if $\delta_r \geq \delta_p$ **then**
    DECOMPOSITION
    JOIN
    Get the candidate graph g_r
    $\tilde{G} \leftarrow g_r$
  end if
end for
SELECTION (of the candidate graph(s) ($\leq \delta_p$) from $\tilde{G}$)

These two techniques are motivated to extract the skeleton mind-graph and manage the mind-graph complexity for complex connections. Also there exists some other techniques or can be formalized these as per some other specific need of the mind-graphs.

4 CONCLUSIONS

In this work we used graphs (mind-graphs) to represent the coupling between the knowledge in the course of textual conversation. The similarity measures between the mind-graphs have been considered for information representation. Also, the algorithms associated to the mind-graphs extraction and normalization have been formalized. Initial experimental framework has been established. It works with test sentences, where extracted word cells and their associated neighbor cells (form the mind-graphs) explicitly defined. Currently, we continue the test with a larger corpus.

ACKNOWLEDGEMENTS

The current work has been performed at the University of Luxembourg within the project EAMM. We thank all project members for their support and engagement.

REFERENCES

Bunke, H. and Shearer, K. (1998). A graph distance metric based on the maximal common subgraph. *Pattern Recogn. Lett.*, 19:255–259.

Clark, A. and Chalmers, D. (1998). The extended mind. In *Analysis*, volume 58, pages 7–19.

Haghighi, A. D., Ng, A. Y., and Manning, C. D. (2005). Robust textual inference via graph matching. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 387–394. Association for Computational Linguistics.

Hensman, S. (2004). Construction of conceptual graph representation of texts. In *Proceedings of the Student Research Workshop at HLT-NAACL 2004*, HLT-SRWS '04, pages 49–54. Association for Computational Linguistics.

Ikonomakis, M., Kotsiantis, S., and Tampakas, V. (2005). Text classification: a recent overview. In *Proceedings of the 9th WSEAS International Conference on Computers*, pages 125:1–125:6.

Jin, W. and Srihari, R. K. (2007). Graph-based text representation and knowledge discovery. In *Proceedings of the 2007 ACM symposium on Applied computing*, SAC '07, pages 807–811. ACM.

Poray, J. and Schommer, C. (2009). A cognitive mind-map framework to foster trust. In *Proceedings of the 2009 Fifth International Conference on Natural Computation - Volume 05*, ICNC '09. IEEE Computer Society.

Poray, J. and Schommer, C. (2010). Managing conversational streams by explorative mind-maps. In *Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications - AICCSA 2010*. IEEE Computer Society.

Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22.

Schenker, A., Last, M., Bunke, H., and Kandel, A. (2003). Classification of web documents using a graph model. In *Seventh International Conference on Document Analysis and Recognition*, pages 240–244.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Comput. Surv.*, 34:1–47.

Tuulos, V. and Tirri, H. (2004). Combining topic models and social networks for chat data mining. In *Web Intelligence, 2004. WI 2004. Proceedings. IEEE/WIC/ACM International Conference*, pages 206 – 213.

Ziegler, C.-N. and Golbeck, J. (2006). Investigating interactions of trust and interest similarity. *Decision Support Systems*, 43(2):460 – 475.