CONSPECT: Monitoring Conceptual Development

Fridolin Wild[1], Debra Haley[1] Katja Bülow[1]

[1] The Open University, Milton Keynes UK MK6 7AA
(F.Wild, D.T.Haley, K.Buelow @open.ac.uk)

Abstract. This paper describes and evaluates CONSPECT, a service that analyses states in a learner's conceptual development. CONSPECT combines two technologies – Latent Semantic Analysis (LSA) and Network Analysis (NA) into a technique called Meaningful Interaction Analysis (MIA). It uses LSA for the language analysis and NA to provide visualisations of the semantic relatedness information calculated by LSA. CONSPECT was designed to help both learners and tutors monitor conceptual development. This paper reports on the verification activities undertaken to show how well LSA matches humans in clustering similar concepts. The verification used the edit distance between card sorts to quantitatively evaluate the service.

Keywords: Latent Semantic Analysis, Network Analysis, card sorting, edit distance, CONSPECT

1 Introduction

The CONSPECT service uses language technologies to support learners and tutors by analysing the semantic content of a learning blog, diary or other form of written exposition. It presents the results in a visualisation called a conceptogram, which highlights concepts and gaps in a learner's text. It was designed to alleviate the heavy workload of tutors responsible for evaluating learners' conceptual development and to provide immediate feedback to learners, thus avoiding the common problem of a lengthy wait for assessment results.

CONSPECT can be used by learners and tutors to analyse the progress of a learner, individually; and to compare the learner with another learner, with the course learning outcomes, with the emerging reference model, and with a group of learners.

CONSPECT was tested with medical students; however, with the proper training material, it transfers easily to other domains. It is suitable for both formal and informal learning. It was tested in a Problem Based Learning situation but can be used in any environment that requires periodic summaries written by the learner.

CONSPECT combines two technologies – Latent Semantic Analysis (LSA) and Network Analysis (NA) into a technique called Meaningful Interaction Analysis (MIA). MIA uses LSA for the language analysis and NA to provide visualisations of the semantic relatedness information calculated by LSA.

The paper is organised as follows. Section 2 gives the theoretical underpinning of the approach to monitoring conceptual development. Section 3 briefly discusses

Latent Semantic Analysis (LSA), Network Analysis (NA), and Meaningful Interaction Analysis (MIA), a novel extension of LSA. Section 4 describes the methodology used to verify CONSPECT. Section 5 discusses the test results. Finally, the paper ends with conclusions in Section 6.

2 Theoretical Underpinnings

Two related Cognitive Linguistics theories that informed the CONSPECT approach are Fauconnier's Mental Spaces Theory and Conceptual Blending Theory [1]. These theories hold that the meaning of a sentence cannot be determined without considering the context. Meaning construction results from the development of mental spaces, or conceptual structures [2] and the mapping between these spaces.

Mental spaces and their relationships are exactly what LSA tries to quantify. LSA uses words in their contexts to calculate semantic closeness and similarity. Because words are used in their contexts, LSA is not just a simple bag-of-words approach. This use of context is consistent with Fauconnier's claim that context is crucial to construct meaning.

Other researchers have posited the existence of mental representations similar in purpose to Fauconnier [3]. Psychologists interested in learning have been studying *associations* since at least the 18th century [4]. According to Deese, "Associations themselves are supposed to arise by contiguity, similarity, etc., but they occur in the strengths and distributions that they do because of their frequency, vividness, and so on." He claims that the associations exist in well and tightly organized networks and that the difference between high and low scorers is the degree to which their networks are well organized. These networks are similar to Fauconnier's mental spaces as well as *mental models* [5] and *situation models* [6]. Zwaan & Radvansky [3] claim that "... many researchers have argued that the construction of a coherent situation model is tantamount to the successful comprehension of a text."

Thus, one can measure a learner's knowledge by measuring his or her situation model. How does one measure a situation model? The following paragraphs give examples from the literature of attempts to measure a situation model.

Shavelson [7] used the association work of Deese [4] to answer the question: "To what extent does the structure in the student's memory, after learning, correspond to the structure in the instructional material?" He gave association tests using key concepts to physics learners and used the results to create directed graphs (digraphs) that showed the relationships between the concepts. From the digraphs, he created adjacency matrices and relatedness-coefficient matrices [7] to represent the learners' cognitive structures. He found that the control group had fairly stable structures while the learners' structures changed over the course of the instruction. He concluded that digraph theory could be used to analyze content structure.

Diekhoff [8] reports that Fenker [9] used association tests (which Diekhoff calls relationship judgments) analyzed through a concept map produced by multidimensional scaling (MDS). Like Shavelson [7], Fenker found that the learners' concept maps changed over the course of instruction and became more similar to the instructor's map and that there was a high correlation between learners' grades and the similarity of their maps to expert maps. Diekhoff used a different method but came up with a similar conclusion: "... relationship judgments testing taps both definitional knowledge of single concepts and knowledge of the structural interrelationships between concepts."

While the researchers mentioned in the previous paragraph use MDS techniques to analyze situation models, others use network analysis (NA) [10], [11] [12]. These researchers use a particular class of network called Pathfinder, which are derived from proximity data [10]. They assume that "concepts and their relationships can be represented by a structure consisting of nodes (concepts) and links (relations)." The strength of the relationships can be measured by the link weights. The networks of novices and experts can be compared to gauge the learning of the novices. Using Pathfinder networks (PFnets), Goldsmith et al [11] found a relationship between conceptual representation and level of performance. Clariana & Wallace [12] found that experts' PFnets were more extensive and better organized than novices' PFnets.

MDS and Pathfinder techniques require the creation of proximity matrices by association, or relationship testing. LSA, on the other hand, requires no such explicit proximity judgments. It uses textual passages to compute automatically a proximity matrix. Thus LSA requires less human effort than these other techniques.

This section presented previous research that attempted to measure a learner's concept structure; the next section describes the technologies used by CONSPECT to do the same thing.

3 Technologies Used

CONSPECT combines two technologies – Latent Semantic Analysis (LSA) and Network Analysis (NA) into a new technique called Meaningful Interaction Analysis (MIA). MIA uses LSA for the language analysis and NA to provide visualisations of the semantic relatedness information calculated by LSA.

3.1 Latent Semantic Analysis (LSA)

LSA is a statistical method for capturing meaning from text. A seminal paper [13] gives a more formal definition: "Latent Semantic Analysis is a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text".

LSA "induces global knowledge indirectly from local co-occurrence data in a large body of representative text" [14]. The LSA technique is essentially a method for solving a huge set of simultaneous equations that represent terms in documents [15]. The basic idea behind LSA is that texts have a semantic structure that is obscured by word usage (e.g. through synonymy or polysemy). LSA unveils this latent semantic structure by using conceptual indices derived statistically from co-occurrences via a truncated singular value decomposition,. Typically, spaces are calculated using a large number of generic and domain-specific *background* documents, thus establishing a dimensional system of human-like complexity. LSA assumes that through the projection of texts into this dimension system, the meanings expressed become salient. The context provided by the bag-of-words activates the correct meaning structures in this multi-dimensional system.

3.2 Network Analysis (NA)

The raw output from an LSA process is difficult to interpret. The fields of Network Analysis and Social Network Analysis (SNA) offer guidance about how to analyse and display the output in a way that can be more easily understood, that is, as a graph. LSA graphs are not, strictly speaking, social networks. However, social networks can be thought of as a metaphor for LSA networks Social networks show the inter-relationships among (usually) people; LSA networks show the inter-relationships among words and documents.

Social network data has been displayed with the help of sociograms, which are visualisations that use an optimised lay-out algorithm to project a complex graph structure onto a 2D display in a way that it most closely resembles the actual structure (see [16] for force-directed lay-out algorithms). This projection becomes harder to interpret and less precise and reliable, when the underlying graph structure is large and complex, which is often the case with social networks and is certainly the case with LSA semantic data.

As a way around this misinterpretation pitfall, network analysis provides a variety of measures that are both visual and non-visual in nature to more accurately investigate the nature of the graph structure. For example, see Figure 1, which displays the degree centrality of nodes – the number of in/out connections to other nodes. See Brandes & Erlebach [17] for more information about this, and other, measures.

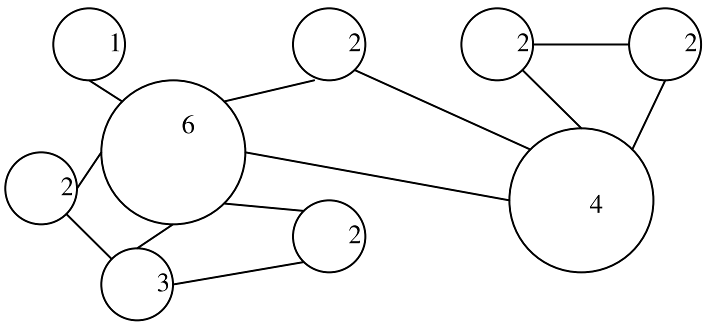

Fig. 1. Network diagram showing degree centrality with labels

3.3 Meaningful Interaction Analysis (MIA)

The graphs created from latent semantic analysis form a complex network structure expressing a manifold of relations between the core actors (terms and documents). In other words, a multitude of differently weighted edges between the nodes is what the resulting spaces and matrices express. This network can be investigated with network analysis. This latent semantic network analysis, a contribution to the field of language technology, is called *meaningful interaction analysis* (MIA).

Terms and documents (or anything else represented with column vectors or row vectors) are mapped into the same space by LSA. Semantic proximity can be measured between them: how close is a term to a document? The resulting graph structure can be filtered and with the help of (S)NA further analysed using e.g. cluster or component analysis, identifying e.g. central descriptors for clusters, etc. Analogous to sociograms, these latent semantic networks can be visualised with the help of force-directed lay-out algorithms [16]. These visualisations are called *conceptograms*.

3.4 Interpreting Conceptograms

This section offers a coherent example of how CONSPECT can be used from the perspective of learners in the more reflective parts of their learning cycles. Evidence for creating a representation of what conceptual knowledge a learner may possess can be gathered from, for example, a learning diary or a forum discussion. This example uses authentic student blog postings by medical students.

The raw text of a blog entry or learning diary is processed, using MIA, into a graph-like representation that reflects the student’s conceptual knowledge in a geometrical, latent semantic interaction model: the student's text is folded into a latent

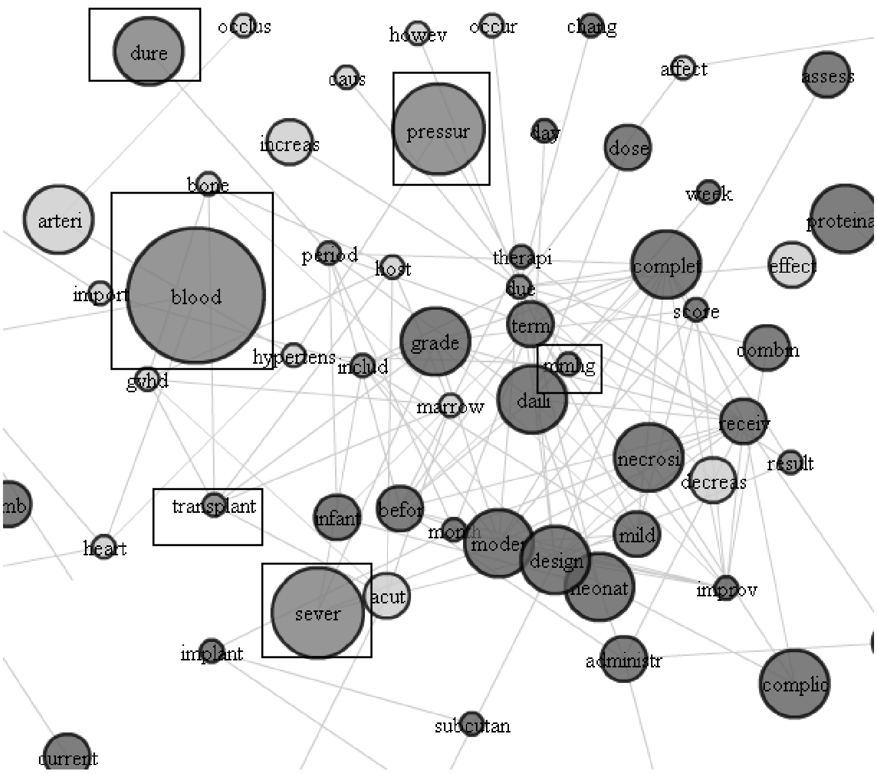

Figure 2 Comparing two learners. Concepts in nodes surrounded by rectangles are mentioned by both learners, light grey nodes are mentioned only by one learner, dark grey nodes are mentioned only by the other learner.

semantic space and within that space all words exert closeness or distance to each other according to their relatedness in meaning. The student's text uses a specific vocabulary and thus stimulates only certain concepts in this latent semantic network. The resulting representation, a network-like structure, is called a *conceptogram*.

CONSPECT offers two types of conceptograms - representations of a single source and combination conceptograms, where two sources are compared. For either type of conceptogram, the labels on the circles (known as nodes) are concepts written about in the source text. The size of the node indicates the relative frequency with which a concept was mentioned in the text. The nodes are connected together by lines, also known as edges. The length of the edge between two nodes is a visual representation of how semantically similar are the two concepts represented by the nodes. (This semantic similarity is calculated from the LSA semantic space and is based on the background, or training, corpus.) In single source conceptograms, the concepts are separated into component networks by clustering; each cluster is depicted with a separate colour. Combination conceptograms are discussed in the next paragraph. Each component shows concepts that are related semantically. Overlapping nodes are not significant; they are a result of the lay-out algorithm's attempt to depict a great deal of information in a limited space.

Figure 2 is a partial screen shot (converted to grey scale for publication requirements) comparing the conceptograms of two learners. Green nodes (surrounded by a rectangle in Figure 2) show the concepts mentioned by both learners. Blue nodes (dark grey) show which concepts one learner blogged about that did not appear in the other learner's blog. Yellow nodes (light grey) show the reverse. The latter two types of nodes are most important for learners as they highlight possible gaps and irrelevancies in their learning blogs.

4 Methodology

The aim of the verification experiment was to see how closely learners' mental models of a domain compare with that of CONSPECT. In other words, does CONSPECT cluster concepts the same way that humans do? Note that this paper reports on just one type of verification activity. A text annotation experiment is described in [18]. The accuracy of CONSPECT was verified using card sorting [19], a technique often used by web designers but used here in a more unusual way. Card sorts allow a researcher to view a participant's mental model of the words on the cards, which is exactly what was wanted.

Eighteen medical students were recruited from the University of Manchester Medical School. By chance, there were nine females and nine males. These students were in their first year of a medical degree program, thus their ages ranged from about 19 to 21. The students were given a £10 Amazon voucher for two hours of their time.

CONSPECT was trained with a background corpus comprising PubMed abstracts (21,091 abstracts, 24,346 words). Using posts from four authentic learning blogs about safe prescribing, CONSPECT generated a list of about 50 concepts for each post. The number varied slightly based on the content of the posts. These approximately 50 words represented concepts activated by the posts and using the background corpus. These words were printed and cut out.

Each student was given a set of cards and asked to arrange them into at least two piles, grouping each pile according to the semantics of the words. This procedure was followed four times – once for each set of blog postings.

Edit distance was used to analyse the card sorts. Edit distance is the number of cards from one card sort that must be moved from one pile to another in order to match another card sort [20].

The edit distance data in Table 1 were calculated as follows. For each test, the edit distance was calculated between Participant 1 and each of the other 17 participants. Next, Participant 2 is compared with participants 3 through 18, and so on until all possible pairs are calculated resulting in 153 comparisons. For the 153 comparisons of Sort A1, the minimum distance was 21, the maximum distance was 34 and the averages of the quartiles are 26.6, 28.7, and 29.8.

5 Discussion

Table 1. Card Sort Results

Card Sorting Results in Terms of Edit Distance							
		Min	Quartile 1	Quartile 2	Quartile 3	Max	#comparisons
Test A1	Average	21	26.6	28.7	29.8	34	153
CONSPECT		30	31.25	33	34	35	18
Test M1	Average	18	22.6	24.7	26.4	31	153
CONSPECT		24	27	28	29	31	18
Test M2	Average	21	30.9	32.7	34.5	40	153
CONSPECT		31	33.3	35.5	36	37	18
Test M3	Average	20	28.7	31.0	33.1	38	153
CONSPECT		30	31	32.5	35	37	18

Table 1 shows the results of four card sorts, each sort conducted by 18 participants. The table reports on the minimum, maximum, and 1st, 2nd, and 3rd quartile edit distances found by the UW Card Sort Analyzer [21] for the participants. The lines labelled *CONSPECT* show the same information when it was compared with the 18 participants. (CONSPECT's sorting data are the clusters calculated by LSA.)

By looking at the edit distance information, one can compare how CONSPECT performs in relation to the human participants. For the min, max, and average quartile edit distances, the CONSPECT figures are larger in each case. For example, CONSPECT was just as good as the humans in Test M1 when the maximum edit distances are compared. The comparison that shows CONSPECT performing least well is in Test M2 and Test M3 when the minimum edit distances are compared.

Table 2 shows the results of an attempt to further understand the card sorting data from Table 1. An interesting question was whether or not the edit distances were dependent on a particular variable. The first column, Sort, is the number of the sort. The second column, *difference*, was calculated by subtracting the average of the means. The third column, *%diff*, is the difference divided by the average of the means. The fourth column, *#cards*, is the number of cards sorted by the participants. The fifth column, *%of cards* is the #cards divided by the average of the means, so this indicates, for example, that out of a total of 52 cards for Sort 3, 67% of them had to be moved (i.e., the edit distance) to achieve identical sort piles. The higher the percent, the worse the sort is. Finally, the last column, *words in posts*, is the number of words captured in the blog and used to extract the concepts to be sorted. There is no clear relationship among these variables. Therefore, one cannot say that shorter posts result in larger edit distances, for example.

Table 2. Distance Information.

Sort	**Difference**	%diff	#cards	%of cards	words in posts
3	**2.26**	6.5%	52	67%	1117
4	**2.33**	7.0%	51	65%	533
2	**3.2**	11.5%	43	65%	557
1	**4.5**	13.6%	48	68%	228
	average	9.7%			

The third column indicates how much larger (as a percentage) the edit distances were for CONSPECT than for the human participants. These figures range from 6.5% to 13.6% with a mean of 9.7%. This analysis suggests that CONSPECT has an edit distance of about 10% larger than the human participants.

6 Conclusions

The results show that in the four sorting experiments, the agreement of the 18 human participants was about 10% better than CONSPECT. Thus, the conclusion is that CONSPECT *approaches* agreement with humans in the clustering of related concepts.

Of course, the experiments tested just 18 learners, however, they were exactly the target audience of this version of CONSPECT.

It is planned to repeat these experiments in another language. In principle, nothing about the way LSA works is dependent on the language used. In addition, the experiments will be repeated for other domains. Specifically, we will test CONSPECT with Dutch psychology students.

Further experiments will investigate the effect of certain changes. One, a better clustering algorithm could be located. Also, the kmeans clustering was run with a small number of iterations. Using a much larger number should create better clusters. Two, the LSA threshold was varied to obtain groups of about 50 concepts. Better results might be obtained by using a higher threshold. Third, the participants could add a description of each pile. This could provide more insight into the reasoning behind the sorting. Finally, it would be instructive to repeat the experiments using a think-aloud protocol or summary interviews. Of particular interest is whether or not a participant had difficulty in deciding in which pile a concept should be placed. This would provide an indicator of the certainty of the sorting choices.

Acknowledgements

CONSPECT was developed and evaluated as part the LTfLL (http://www.ltfll-project.org/) project, which is funded by the European Union under the ICT programme of the 7th Framework Programme (Contract number: 212578).

We would like to thank the University of Manchester, specifically Alisdair Smithies, for help and support in this investigation.

References

1. Evans, V., Green, M.: Cognitive Linguistics, An Introduction. Edinburgh University Press, Edinburgh (2006)
2. Saeed, J.I.: Semantics. Wiley-Blackwell (2009)
3. Zwaan, R.A., Radvansky, G., A.: Situation Models in Language Comprehension and Memory. Psychological Bulletin **123** (1998) 162-185
4. Deese, J.: On the Structure of Associative Meaning. Psychological Review **69** (1962) 161-175
5. Johnson-Laird, P.N.: Mental Models: Towards a cognitive science of language, inference, and consciousness. Harvard University Press, Cambridge, MA (1983)
6. van Dijk, T.A., Kintsch, W.: Strategies in discourse comprehension. Academic Press, New York (1983)
7. Shavelson, R.J.: Some Aspects of the Correspondence Between Content Structure and Cognitive Structure in Physics Instruction. Journal of Educational Psychology **63** (1972) 225-234
8. Diekhoff, G.M.: Testing Through Relationship Judgments. Journal of Educational Psychology **75** (1983) 227-233

9. Fenker, R.M.: The organization of conceptual materials: A methodology for measuring ideal and actual cognitive structures. Instructional Science **4** (1975) 33-57
10. Schvaneveldt, R.W., Durso, F.T., Dearholt, Donald, W.: Network Structures in Proximity Data. The Psychology of Learning and Motivation **25** (1989) 249-284
11. Goldsmith, T.E., Johnson, P.J., Acton, W.H.: Assessing Structural Knowledge. Journal of Educational Psychology **83** (1991) 88-96
12. Clariana, R.B., Wallace, P.: A Computer-based Approach for Deriving and Measuring Individual and Team Knowledge Structure from Essay Questions. Journal of Education Computing Research **37** (2007) 211-227
13. Landauer, T.K., Foltz, P.W., Laham, D.: An Introduction to Latent Semantic Analysis. Discourse Processes **25** (1998) 259-284
14. Landauer, T.K., Dumais, S.T.: A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. Psychological Review **104** (1997) 211-240
15. Landauer, T.K.: LSA as a Theory of Meaning. In: Landauer, T.K., McNamara, D., Dennis, S., Kintsch, W. (eds.): Handbook of Latent Semantic Analysis. Lawrence Erlbaum Associates, New Jersey, US (2007) 3-35
16. Fruchterman, T., Reingold, E.: Graph Drawing by Force-directed Placement. Software - Practice and Experience **21** (1991) 1129-1164
17. Brandes, U., Erlebach, T.: Network Analysis: Methodological Foundations. Springer-Verlag, Berlin, Heidelberg (2005)
18. Wild, F., Haley, D., Bulow, K.: Monitoring Conceptual Development with Text Mining Technologies: CONSPECT. EChallenges, Warsaw (2010)
19. Rugg, G., McGeorge, P.: The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. Expert Systems **14** (1997) 80-93
20. Deibel, K., Anderson, R., Anderson, R.: Using Edit Distance to Analyze Card Sorts. Expert Systems **22** (2005) 129-138
21. Deibel, K.: UW Card Sort Analyzer: Version 1.1. Vol. 2010. University of Washington (2005)