Towards a Model for the Selection Task: A Tensor Based Architecture

Eshaa M. Alkhalifa

University of Bahrain, College of Information Technology, P.O. Box 32038, Isa Town Bahrain
ealkhalifa@itc.uob.bh

Abstract **This paper presents a tensor based neural network model for semantics that utilizes co-occurrence data obtained through Latent Semantic Analysis (LSA) and adds to that word order information. This powerful architecture is competition based, so it is not trained to obtain specific results. Yet, it is shown to be capable of isolating groups of rules only given their numerical representation in a way that is justified by subject behaviors reported by respected researchers in the field. Results indicate that this architecture is capable of probing highly detailed representations of meaning in cases that are hard to uncover or achieve using vector based analysis.**

I. INTRODUCTION

THE questions raised by the simple conditional problem proposed by Wason [15] remain to this date mostly unanswered. A card selection task, with a rule that has to be checked, has persistently resulted in extremely low levels of accuracy amongst students. Typically the cards show an "A", a "B", a "4" and a "7" and students are asked to indicate the least number of cards that they would turn over to check a rule. A possible rule is: If A then 4 and the correct answer is to turn over the cards showing "A" and "7". However, only 4% to just over 20% of students select these two cards [12]. In general, almost all the work from different directions united to show that the effect of content is highly influential on subject performance which led to dividing the materials into two groups; thematic and abstract. 'Thematic' materials are characterized by a thematic relationship between the two propositions and were observed to result in a much higher level of accuracy [13], [14] when compared to the classical selection task question shown above which was classified as 'abstract' and 'concrete'. The dividing line between the two was never clear-cut.

Further experiments showed that having a 'rationale' specifically in permission and obligation situations, resulted in higher levels of accuracy [7] even when dealing with abstract material. Another competing theory is that of 'social contracts' introduced by Cosmides [8]. The contract relates *perceived benefits* to *perceived costs.* In a sense, it works as an exchange of cost for benefit and cheating occurs when one fails to pay the cost and accepts the benefit. Gigerenzer and Hug [9] further supported this theory by showing that when subjects are asked to adopt a 'social role perspective' they would perform much better at the same task.

It goes without saying that the effort involved into the investigations made into these "errors" is phenomenal with theories and explanations of behavioral patterns equally diverse yet the division of materials into 'thematic' and 'abstract' ignores the interplay of 'form' and 'content' or structure versus semantics. "The conditional is not a creature of constant hue, but chameleon like, takes on the color of its surroundings; its meaning is determined to some extent by the very propositions it connects." [13]. This indicates that rule meanings may have a large role to play in subject behavior in this task. The influence of order, temporal distancing and color [3],[4],[5] was extensively studied in the past to show an extremely important role of semantics in the task and the influence of the 'directionality' or order of the rule on subject performance which encouraged the work presented here.

II. LATENT SEMANTIC ANALYSIS

Latent Semantic Analysis (LSA) is a powerful high dimensional semantic model that ignores the effects of directionality on semantics. It treats contexts of around 2000 words as bags of words ignoring any word order by studying word use in context. The assumption is that a word is in some way semantically described by the words that appear alongside it in the same context.

Landauer and Dumais [11] created a multidimensional vector representation in "semantic spaces". The process starts out with a co-occurrence matrix that has as entries the number of times each word appears in a particular context. Then these numbers are modified with a special function that takes the "entropy" or importance of the word into consideration and takes its log. Then the adjusted matrix is taken through a Singular Value Decomposition to transform it into three matrices. The first gives a vector representation of the words in terms of the contexts. The second gives a vector representation of the contexts in terms of the words. The third is a scaling matrix that only has non-zero elements along the diagonal. When the three resulting matrices are W, P, S then the original would be obtained when we multiply W*S* P'.

The secret of LSA lies in that before the vector components are put together again, the number of columns with values is reduced to usually 300 dimensions. When this is done, the matrix that results from the multiplication process, redistributes the effects of these numbers over a large span of cells. So altering a single value in the original matrix, results in a change of a large group of cells in the resulting matrix.

Results show a high degree of correlation between the predictions made by the model and human behavior. It also seems to have the ability of extracting a vector representation that is capable of assessing the "semantic" distances between words in a contextual space. The semantic space LSA uses to represent vectors, is a "world of words" where each word has a location based on the distance its meaning is from other words. This distance is estimated without any regard to where the word appeared in the sentence, nor does it accommodate for multiple possible "orders". So, it should not be surprising that it confuses the two sentences "Mark killed the Tiger" and "The tiger killed Mark" as meaning exactly the same, when no human would equate the life of a human with that of an animal.

In order to further understand the problem we should look a little deeper into the makings of LSA. It attempts to capture meaning based on the structural making or word co-occurrence within contexts without paying any attention to the location of the word in the sentence or the syntactical category it belongs to. This led Wiemer-Hastings [16] to further investigate the effects of the neglected syntax and attempt to incorporate it into the LSA framework. He separated the sentences into atomic clauses or propositions and then segmented them by hand to break them into strings composed of subject noun phrase, verb and object noun phrase. Antecedents were used to resolve pronouns and conjunctions were dealt with by distributing the arguments. Then he attempted to evaluate the similarity of this presentation using a variety of measures. Results showed that the best approach to combine the similarities of the sentence parts is non-linear and even that was not as close to human judgment as LSA. Wiemer-Hastings and Zipitria [17] then went on to a further test, incorporating syntax through two methods. The first was to tag the words used for the training corpus at 100, 200, 300, and 400 dimensions and this did not produce any favorable results. Then they tried a structured LSA or SLSA where they broke up sentences into parts as was shown above and used that as training material to find results that correlate slightly better than LSA.

However, this approach is not the only one followed to incorporate syntax into word meaning. The Hyperspace Analogue to Language (HAL) is similar to LSA in that it is based on co-occurrences of words, except that word order information is retained [6]. For each window of 5 or 10 words a table is formed to indicate the words that precede it and those that follow it. This information is then concatenated into a vector that can be up to 100 dimensions long. The main difference between HAL and LSA is that the latter uses around 2000 word passages to infer co-occurrence and uses singular vector decomposition to re-distribute the data utilizing only the highest 300 influential factors. So word meaning is extracted from a larger set of words than in HAL, yet it does not give any importance to word order while HAL does. Since LSA utilizes a wider space of 'context' words, it seems wise to attempt to add 'order' to LSA and this only has one problem which is simply that the model must be raised one rank to become a tensor-based representation.

III. Introducing Word Order Through a Tensor Architecture

A way to regard this view of sentences should not be so complex when we regard them as terrains of land that have peaks and valleys. This implies that a matrix is required to describe any tensor to give complete information about the position and orientation of every point. In order to achieve this representation, the LSA results are taken as is to represent words as vectors while sentences are represented through tensors. If we wish to represent the word "FLAG" in 6-dimensional space the result may be as follow:

FLAG 0.29 0.17 0.31 0.27 0.32 0.19

However when one wishes to represent a sentence in LSA then the sentence too will be represented in the same number of dimensions as determined by the operations followed to obtain a numerical representation for it. If on the other hand, we assume that a sentence requires by definition a higher level of representation, then we can keep each of the individual word vectors as is and represent it as a whole through a matrix.

TABLE 1
THE MATRIX DESCRIBES THE SENTENCE MAINTAINING THE ORDER OF THE WORDS

THE FLAG FLIES HIGH					
0.99	0.78	1.00	0.80	0.96	0.85
0.29	0.17	0.31	0.27	0.32	0.19
0.26	0.23	0.24	0.24	0.24	0.23
0.45	0.33	0.46	0.33	0.45	0.40

The architecture utilized here [1], [2] stems from the architecture of a self-organizing neural network algorithm, designed by Tuevo Kohonen that allows the inputs to organize themselves in a high dimensional space as per a competition [10]. Input patterns are presented through a training phase and each is classified by the units it activates based on its similarities to other members of that group. These similarities are mapped into 'closeness' measure that is dynamically estimated by the competitive layer. The networks are unsupervised in the sense that they are only given the inputs and allowed to search for 'closeness' measures between the different inputs. The decision of how close any input is, must be assessed through a special function that can take various different forms or ways to measure the differences between the points. In order to estimate tensor distances, this formula is used for 12 word sentences where each word is represented in 218 dimensional space in order to estimate how "close" a weight tensor is to the sentence.

$$\underline{\underline{T}} - \underline{\underline{U}} = \sum_{i=1}^{12} \sum_{j=1}^{218} T_{ij} - U_{ji}$$

Results of that are in tensor form so the following operation is used to estimate the magnitude of the tensors, in order to select the closest one.

$$\left| \underline{\underline{T}} \right| = \sqrt{\frac{1}{2} \sum_{i=1}^{12} \sum_{j=1}^{218} T_{ij}^{2}}$$

Once the winning weight tensor is selected, it can be adjusted to bring it closer to the input tensor. For the evaluation phase the dot product operation is required and it is as follows.

$$\underline{\underline{T}} : \underline{\underline{U}} = \sum_{i=1}^{12} \sum_{j=1}^{218} T_{ij} U_{ji}$$

IV. A Semantic Model of the Selection Task

In order to do that, the vector operations required by the process must be replaced with tensor operations. Weights have to be three-dimensional as well in order to enable them to partition the tensor space. The words used in this model were collected from rules used in 68 runs in the Wason selection task and then the values of their representations in terms of each other were found through LSA's online site. This resulted in a matrix of 218x218 words, such that each word has a descriptive vector or length 218. Then ten rules were selected from the group such that their lengths do not exceed 24 words in order to ensure comparable sentence length. A part of the rules are presented as follows:

R12=A card has an 'A' on one side, only if it has a '4' on the other side.

R18=If a bird has a purple spot underneath each wing, then it must build nests on the ground.

R27=If a bolt of cloth has any red threads in it, then it must be stamped with a triangle.

R7=If a card has an 'A' on one side, then it must have an 'A' on the other side.

R78 or 79=If a customer is drinking an alcoholic beverage then he or she must be over 18 years old.

R15=If a customer is to drink an alcoholic beverage, then she is at least 18.

R72or 73=If a home owner gets a subsidy, then that person must have installed a modern heating system.

R34 or 35 or 36=If a house was built before 1979, then it has a fireplace.

R66 or 67=If a small-time drug dealer confesses, then he will have to be released.

R37=If a steel support is intended for the roof, then it must be rustproof.

R61=If a student is intended to be assigned to Grover High School, then that student must live in Grover City.

R19=If a washing label has silk on one side, then it has "dry clean only" on the other side.

R68 or 69=If an employee works on the weekend, then that person gets a day off during the week.

R74 or 75=If an envelope is sealed, then it must have a 1-mark stamp.

R51=If I go to Leeds, I travel by car.

R52=If I go to Leeds, I travel by train.

R49=If I go to Manchester, I travel by car.

R50=If I go to Manchester, I travel by train.

R5=If one is to take action 'A' then one must satisfy precondition 'P'.

R44=If one works for the Armed Forces, then must vote in the elections.

R60=If someone stays overnight in the cabin, then that person must bring along a bundle of wood from the valley.

R54=If the letter is N, then the number is 3.

R9=If the tablecloth is brown, then the wall is white.

R20=if two objects carry like electrical charges, then they will repel each other.

R59=If you eat duiker meat, then you have found an ostrich eggshell.

R1=If a letter is sealed, then it must carry a 20-cent stamp.

R10=The tablecloth is brown only if the wall is white.

The rule numbers are shown as taken from the larger database so they are not consecutive and duplicates have been removed as well as extra long rules leaving a total of 37 unique rules to test. The problem that remains is that some of the rules were surrounded by a context that is not included in the study. Consequently, one must expect some overlap between the formed groups, yet it does seem to go quiet a bit of the way in order to properly isolate similarities. This is itself a large step forward in the path of realising a model of this task.

V. Results and Discussion

The model was capable of isolating several groups of rules that seemed semantically similar. The first that was easily identifiable is the group of rules; R49, R50, R51, R52 all of which have a 62.5 percentage accuracy. In order to estimate expected subject behavior for the social contract rules that altered perspectives a mean was calculated for the two rules selected. A second group was formed from R20,R1, R74 or 75, R34, 35 or 36, R66 or 67,R54 and R5 with respective accuracies of 34%, 86%, 54%, 40%, 57%, 6.25%, 61%. Another group was composed of rules R60, R27, R68 or 69, R61 with subject accuracies of 89%, 40%, 75%, 77%. Clearly a full data analysis of the results has to be carried out and various statistical comparisons with the basic information obtained from LSA. However, a point worth noting here is that, isn't it odd that the model fumbles every time it meets an "abstract" rule in a way that is not different from what happens to people?

To show the power of this model ten rules were extracted from the above data and shown in the figures below; namely R34,35 or 36, R74 or 75, R51,R52, R49, R50, R54, R9, R1, R10. It should be clear from the figure the power of the network in grouping similar rules to each other in Figure 1 while Figure 2 shows LSA results.

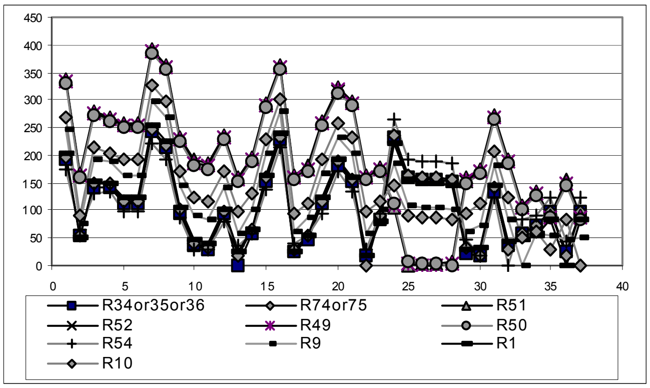

Fig. 1. Model groups for ten rules. Notice that the rules are clearly grouped into three groups.

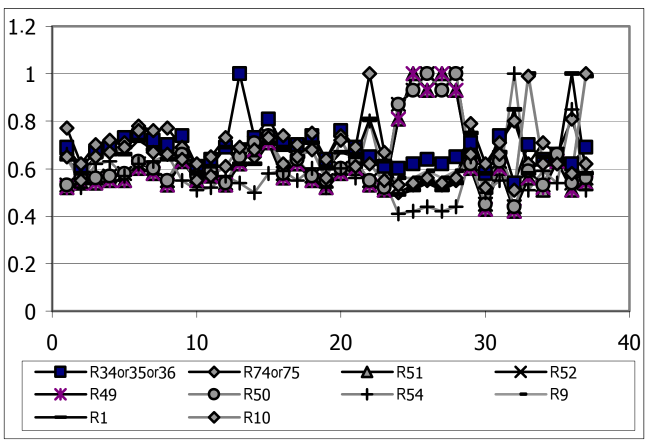

Fig. 2. LSA graph for the same ten rules. The source data for the semantic representation does not show any clear differences between rules when it has no sensitivity to word order.

The results show that although the LSA model is extremely powerful, it has neglected word order and its influence on meaning. The tensor model proposed here is better able to distinguish groups of rules that are similar as is shown figure 1 where the rules, R49, R50,R51 and R52 are grouped together and the rules R54, R34, R74 and R1 are grouped together leaving rules R9 and R10 alone. This corresponds to subject behavior results of 62.5% for the first four rules and the corresponding averages of 6.25%, 40%, 54% and 86%. This implies that an error is only made with R54 which is the rule "If the letter is an N, then the number is a 3". Note that this rule is confusing because of the different sense of 'letter' used here and the sense of number used. As previously said, the model is 'confused' when it sees the abstract rules and has trouble classifying them while having less trouble classifying the thematic tasks which supports the current assumptions that subject behavior is influenced by semantics. The model presented here shows an opportunity to further analyze this aspect through providing a tensor based representation for rules.

REFERENCES

[1] Alkhalifa, E. M.: A Tensor-Competition Based Architecture: To Capture the Infleunce of Word Sense. *9th International Conference on Neural Information Processing*, ICONIP'02, Singapore (2002b)

[2] Alkhalifa, E. M.: The "Meaning" of Structure in Digits, *9th International Conference on Neural Information Processing*, ICONIP'02, Singapore (2002a)

[3] Alkhalifa, E.M.: Directional thought effect in the Selection Task. *Proceedings of the Third International Conference on Cognitive Science*, Beijing, China, (2001b) 171-176.

[4] Alkhalifa, E.M.: Egoistic reasoning 'In Time' in the Selection Task. *Proceedings of the Third International Conference on Cognitive Science*, Beijing, China, (2001a) 109-113.

[5] Alkhalifa, E.M.: The effects of meaning in a white/black setting on reasoning. *The 6th Australasian Cognitive Science Society, OzCogSci'02*, Fremantle, Western Australia, (2002).

[6] Burgess, C., Lund, K.: Modeling parsing constraints with high-dimensional semantic space. Language and Cognitive processes, Vol. 12, (1997) 1-34.

[7] Cheng. P. W. & Holyoak, K. J.: Pragmatic Reasoning Schemas, *Cognitive Psychology*, Vol. 17, (1985) 391-416.

[8] Cosmides, L.: The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, Vol. 31, (1989)187-276.

[9] Gigerenzer, G. & Hug, K.: Domain-specific reasoning: social contracts, cheating and perspective change. *Cognition*, Vol. 43, (1992) 127-171.

[10] Kohonen, T: *Self-Organizing Maps*, Springer, Berlin, (1995).

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

[12] Manktelow, K. I., Evans, J. St. B. T.: Facilitation of reasoning by realism: effect or non-effect? *British Journal of Psychology*, Vol. 71, (1979) 227-231.

[13] Wason, P. C. & Johnson-Laird, P. N.: Psychology of Reasoning: Structure and content, *Harvard University Press*, Boston (1972).

[14] Wason, P. C. & Shapiro, D.: Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, Vol. 23, (1971) 63-71.

[15] Wason, P.C: Reasoning, In: Foss, B.M. (eds.): New horizons in Psychology, Harmondsworth, UK: Penguin, (1966).

[16] Wiemer-Hastings, P. Adding syntactic information to LSA, *Proceedings of the Twenty Second Annual Conference of the Cognitive Science Society*, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 989-993, (2000).

[17] Wiemer-Hastings, P., Zipitria, I.: Rules for syntax, vectors for semantics. *Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society*, Mahwah, NJ: Lawrence Erlbaum Associates, (2001).