Title: Context Tracking for Natural Language Processing

Authors Names and Affiliation:

W.H. Wilson  
Flinders University of South Australia  
School of Mathematical Sciences  
BEDFORD PARK  S.A.  5042

Abstract (100 words or less)

The aims of this paper are to outline what linguistic context is, in general, and to describe an algorithm for tracking changing context. The algorithm is designed to simulate the effect of activation, inhibition, and decay of activation, in nerve nets, using a list of items with their activation levels.

Computing Reviews Keywords

Language models, Language understanding

Categories  I.2.7, I.2.4
CONTEXT TRACKING FOR NATURAL LANGUAGE PROCESSING

Abstract: The aims of this paper are to outline what linguistic context is, in general, and to describe an algorithm for tracking changing context. The algorithm is designed to simulate the effect of activation, inhibition, and decay of activation, in nerve nets, using a list of items with their activation levels.

1. THE NATURE OF CONTEXT

Introduction

Natural language processing systems have, for the most part, avoided the general problem of context. Some systems have dealt with specific context-related problems. For example, Schank & Abelson (1977) concentrate on contextually relevant knowledge in a specific situation such as visiting a restaurant. De Jong (1982) concentrates on the problem of how to select such a knowledge module given a text whose topic is, initially, unknown. Wilks (1978) looks at word-sense ambiguity in context - for instance, deciding which sense of pitcher is intended in the sentence The batter drank a pitcher, on the basis that drank prefers a liquid object, and a pitcher may figuratively be identified with the liquid it might contain.

However, context is a wider concept than might be inferred from the sum of these three examples. Of course, none of these theories claim to solve the context problem in its entirety. De Jong's system, for example, while it handles texts on a range of topics, would most likely not do well with a text in which the topic changed from time to time.

There are also nerve-net style models of context (e.g. Waltz and Pollack, 1985). These operate by associating nodes in the net, or groups of nodes, with concepts: nodes for semantic alternatives for a word, subnets corresponding to alternative syntactic structures for an utterance. Once a node is "active", it sends activation to other nodes with which it is positively associated, and inhibition to nodes with which it is negatively associated, and the whole system is allowed to iterate until a stable activation pattern emerges. Activation may decay away if not reinforced by other nodes.

1.1 Facets of Context

What makes up the context in which linguistic items - or indeed non-linguistic items such as sights, smells, non-linguistic sounds, etc. - are interpreted? One answer is: everything. That is, everything that we know or believe, or have experienced or are experiencing. Unfortunately, this is not a very useful answer: "everything" is not machine-
representable in any useful way. So we refine our question. What aspects of "everything" are likely to be useful parts of the context in a given situation?

We will now describe a context model which tries to answer this question. We make the assumption that it is possible to decompose context information into context facets. Facets may overlap: the information in the facets forms a cover of the complete context.

(a) Relevant knowledge. Knowledge about a myriad of subject areas would be necessary only in suitably specialised discussions. All discussions are specialised, in the sense that they have a theme, so suitable knowledge must be available.

(b) Theme. Sometimes the theme of a passage is announced as a heading or in the first sentence. The understander thus acquires referents, with respect to which the inexplicit terms used in the discourse can be grounded. Themes can be thought of as structures containing pointers to the relevant knowledge.

(c) Physical surroundings. The current physical surroundings will sometimes, but not always, be tacitly included in the likely topics of discourse. A part of the physical surroundings usually only becomes relevant when some discourse participant draws attention to it. This might be done with a deictic reference: This room is too hot.

(d) Things in common between discourse participants. The past shared experiences of the discourse participants will automatically be called to mind by both participants. If the shared experiences are extensive, or part of some standard cultural pattern, then they will certainly have been digested and organized, perhaps along the lines suggested by Schank (1982).

(e) Things already said in current discourse. This facet differs from the theme facet in that it provides a different digested version of the previous discourse. Theme may not mention items considered to be unimportant parts of the discourse, but it is still necessary to keep track of such items, in case they are mentioned again. This facet is likely to have sub-structure: for example lists of items ordered according to their relative importance, their recency of mention, and their textual prominence.

(f) Things related to the current context, though not explicitly part of it. Such items implicitly form part of the context, but would be used in interpreting a particular part of the incoming discourse only if no more immediate item allows an interpretation. In some cases there will be cues indicating that the context is moving on to related issues or related concepts.
(g) Plans, goals, and knowledge of the discourse participants, and perceptions of the plans, goals, and knowledge of a participant by the other participant(s). Schank and Abelson (1977) have developed a model of the goals and plans of human actors in isolation. Wilks and Bien (1984) have investigated the way in which actors model the goals, plans, and knowledge of other actors. The plans and goals of actors in discourses, both as perceived by the actors and those observing them, impinge on the importance of the concepts being discussed and hence affect the context.

1.2 Relevant Knowledge

In this section, we expand on the concept of relevant knowledge. To avoid some problems of ambiguity, and also for the sake of efficiency in searching the space of knowledge, a language understander should have access only to knowledge that is relevant, when this can be predicted. So we must keep track of what knowledge is relevant, as a discourse progresses.

The term "relevant knowledge" finds two natural uses. It is used to refer to the knowledge itself, but it is also used to refer to a structure which describes or lists what knowledge is relevant. These two might have the same representation, if the knowledge came in such small chunks that pointers to the chunks imposed more overhead than they were worth. But if the knowledge is packaged in larger modules, then it may be worthwhile to maintain a list of the names of these modules, particularly if some of them are only potentially relevant.

1.3 Context Change

As a discourse progresses, its context may change. Participants may leave or join; physical surroundings may change; plans and knowledge of the participants may change; theme may change. A context representation needs to keep track of the changes in whatever context facets it handles. The context can change in at least the following ways:

(a) Elaboration. The discourse context is expanded to include related domains at the same level of generality. For example, a discussion of one person's job might expand to a discussion of the jobs of several persons' dissimilar jobs.

(b) Particularisation. The context moves from a more general set of domains to a subset, or to a set of domains subordinate to some of the original domains. For example, a discussion of a person's job might contract to a discussion of a particular aspect of it, say the hours of work.

(c) Abstraction. The context changes to a set of domains higher up an abstraction (ako/isa) hierarchy from the
current set. For example, a discussion of one person's job might be transformed into a discussion of the labour market in general.

(d) Association. The context moves to a set of domains different from but related to and suggested by some of the current domains. For example, a discussion of one person's job might drift into a discussion of that person's leisure interests, educational qualifications, or social life.

(e) Reversion to Earlier Topic. The context moves to include a set of domains referred to earlier in the current discourse.

(f) Logical Progression. In a discourse with an overall sequential structure, a particular part of the sequence completes, and the context moves on to the domains relating to the next part of the sequence. This mechanism could mimic some of the other forms of change listed above. For example, in a discussion of a person's general well-being, there might well be a sequential structure, moving from the person's job to that person's social life, and subsequently to the person's interpersonal relations.

(g) Abrupt Change. Caused by an obscure reminding process, or by a new external stimulus. Since the change is seen as abrupt, the reminding process must have been obscure, but the change must have been prompted by some external stimulus or internal process.

(h) Change in Physical Surroundings. If the discourse participants are moving, joining or leaving, or if the surroundings are changing independent of what the discourse participants are doing, a particular phrase uttered at two different points in time could have two different meanings. For example, the phrase this cat could have different referents at different times. Similarly, there can be changes in non-physical aspects of the environment such as ownership.

(i) Change in Knowledge and Plans of Participants. As a discourse progresses, participants may learn things, and the things they learn may change both the way other participants communicate with them, and the plans and attitudes of the learner.

1.4 Context Tracking

A context tracking method should ideally be able to cope with any of the types of context change mentioned in the previous section. Systems specialising in some of them include that of Grosz (1978), which dealt with structured dialogues relating to machine assembly tasks, and that of Reichman (1985), which dealt with the use of discourse cues.
The algorithm to be described later in this paper does not rely on the detailed semantics of the text, as do the algorithms of Grosz and Reichman, but rather on concept activation and semantic association to track what is happening in a text. Unless the association network was set up with unnatural constraints, this algorithm would not, when "understanding" a recipe, for example, know which of slicing or frying should follow the appearance of an onion in the text. Grosz-type schemes would do better at this.

2. CONTEXT IN ARTIFICIAL INTELLIGENCE

2.1 Partitioned Semantic Nets and Focus

Hendrix (1978) describes a method in which knowledge, represented in an underlying semantic network, has an additional structuring into spaces (subsets of the network), and vistas (sets of spaces). In another part of the same project, Grosz (1978) used spaces which represented sets of nodes which might be the current context at some time during a dialogue. This cover is referred to by Grosz as the focus partitioning.

2.2 Spreading Activation and Connectionism

Waltz and Pollack (1985) model contextual and other factors in language comprehension using an association network with spreading activation and lateral inhibition. That is, as well as activation of a node resulting in activation of related nodes, it results in inhibition of nodes whose meaning conflicts with that associated with the first node. This scheme has the advantage that it avoids the possibility of having conflicting nodes activated at the same time. It is reasonable to suppose that the Waltz/Pollack model, with time decay added, would resemble the method used by organic brains. In this sense, their model is attractive. Disadvantages of such models include difficulty in answering questions like What is this passage of text about?, as there is no explicit notion of theme, and that there is no obvious way to handle discourse cues.

2.3 Context Factors

Another activation-based model is that of Alshawi (1983). This approach models contextual influences by having them generate context factors roughly corresponding to the context facets discussed earlier. Each context factor has an activation level, which may or may not decay as processing progresses. The correct sense for a particular word is chosen by summing the activation levels of the context factors associated with the different senses of the word, and choosing the sense with the highest activation level.
3. CONTEXT TRACKING

3.1 Contextual Knowledge Structures (CKSs)

Because a context tracking method may not specify all the details of the representation of the knowledge it keeps track of, we shall use the neutral term contextual knowledge structure (normally abbreviated to CKS) to refer to any modular knowledge structure, such as frames, scripts, MOPs, and mini-expert systems, which the method may manipulate. It is assumed that knowledge is modularised in such a way that any particular word will have a unique meaning in a particular CKS, once the word's syntactic role is known. In a hypothetical NAUTICAL CKS, the word port, as a noun, would mean harbour, while, as an adjective, it would mean left hand side of ship, looking forward.

Thus, in order to help deduce the currently relevant CKSs from the text being processed, with each word there will be associated the set of CKSs in which the word has a meaning. We shall call this set of names the Candidate CKSs for the word. Whenever the word is used in a way which humans would regard as unambiguous (excluding puns, for example) we have in effect to decide which member of the candidate CKSs is the correct one in which to determine the word's meaning.

We can make this decision in more than one way. We might somehow arrive at an initial collection of active CKSs, and thereafter only change this only when it proves to be impossible to understand something with the currently available knowledge. Here "understand" might mean "have access to a meaning for each word" or it might embody a more complete notion of meaning. We might assume, alternatively, that incoming phrases should actively determine context changes, and update the CKS information as part of processing each phrase.

3.2 ACKS

Because context may veer away from a topic but later revert to it, it is sensible to retain the names of old CKSs. When an understanding fault (that is, a failure to find a meaning for some word by searching the current CKSs) occurs, we might attempt to understand by inspecting old CKSs. From the viewpoint of the spreading activation model, these old CKSs, or rather the "nodes" within them, are still activated, but less so than the CKSs currently in use. We call this structure the Active CKSs, or ACKS: it can be viewed as a list ordered by activation level.

3.2.1 ACKS Management: Size and Decay.

We have to consider how CKSs would enter the ACKS (activation), how they progress down the list (decay), and whether the list should grow to an unlimited size. In this section we consider decay and list size; we discuss activation in the next section.
If unlimited growth occurred, there would be problems. First, the system would not behave like human language understanders, who lose track of material not mentioned for some time. Second, the amount of searching which would become necessary before one could conclude that a section of text was incomprehensible might become too large. In other words, the system must exhibit the human trait of forgetfulness in order to avoid having large parts of the knowledge base becoming partly activated.

So we must limit the size of the ACKS in some way. We choose to achieve this by having CKSs migrate down the ACKS as their activation level decays, as further utterances or text is processed, unless they acquire extra activation from the new utterances or text as they are processed. When a suitable cut-off level of activation is reached, a CKS is removed from the ACKS. CKSs are thus discarded on a least recently used basis. There would be a de facto bound on list size caused by the inability of the stream of incoming linguistic items to keep more than a certain number of CKSs active. However, this bound might be unreasonably large, and a more complete model might need to incorporate an explicit bound on the number of CKSs which could simultaneously be at all active. The bounded size of short-term memory in humans is a similar (but distinct) phenomenon.

Introspectively, the human method seems to use two levels, current and previous, rather than as a collection of themes or conceptualisations with a wide range of activation values.

3.2.2 ACKS Management: Activation

To extract CKS information from the incoming stream of words, we accumulate the words into groups which can be assumed to correspond to objects, individuals, or actions: for example, a noun group such as the furry purple cube would give rise to a single item contributing activation to whatever CKS or CKSs furry purple cubes were thought to belong to. Other sources of activation exist, corresponding to the context facets listed in section 1.1, and potentially to the types of context change listed in section 1.3. Thus, for example:

- a system which was building semantic structures in an incremental way as the text is read or utterance is heard, could use them, or summary information from them, to keep certain items considered important, (and the corresponding CKSs), at a high level of activation;

- a system which maintains a model of the real or simulated physical surroundings could ensure that all or some of the physical objects present maintained a level of activation for their associated CKSs, even in the absence of explicit reference in the text or utterances;
* a system which has models of the discourse participants and of their knowledge, plans, views of each others, etc., could use this information, for example, to preload activation at the start of a discussion, or to modify activations when it becomes clear that the participants' knowledge or plans are changing;

* a system which tracks discourse cues as used by Reichman (1985) could use a more elaborate context model which, for example, could switch or modify contexts in response to appropriate cues. It would be able to restore those contexts when a subsequent cue indicates reversion to an earlier context;

* certain CKSs might include knowledge of expectations of where the discussion is likely to go next (by elaboration, particularisation, abstraction, association, or logical progression). Such a system could arrange to lightly activate the appropriate CKSs, in anticipation of their being used.

It would be difficult to construct a program which took all of these factors into account, but potentially, they are all available and useful.

3.2.3 Structures and Algorithms for ACKS Management

Data structures required for a base-level implementation of an ACKS-based system would include a nodes corresponding to the names of all the CKSs, and containing their current activation levels. For each word in the lexicon, there would need to be information about which CKSs should be sent activation when that word is encountered as a, or the, principal word in a word group, and what activatory and inhibitory links there should be between those CKSs in the context of that word. Prepositions, conjunctions, and so forth, would not send activation to CKSs directly, though they might help to propagate constraints on word meanings by determining semantic case.

Since activation might flow indirectly via the abstraction hierarchy, the node for each CKS should also indicate which other (usually more abstract) CKSs should be sent activation when the node in question reaches its firing threshold.

The basic algorithms will consist of a word-group finder, a method of maintaining the activation levels of the CKSs, and a way of interpreting the ACKS in order to allocate to a word group the correct, or most likely, CKS(s).

3.3 Deixis, Consistency, and Inhibition

The given/new distinction and other pragmatic factors may give a system which takes them into account a reason to vary the activation given to the associated CKSs, so as to reflect the semantic importance of the item giving rise to the activation.
As noted by Alshawi (1983), certain surface structures such as demonstrative adjectives (This furry purple cube) and proper nouns (The Force™) give one reason to give appropriate CKSs an extra "jolt" of activation: the first because the textual deixis may be taken to emphasise the object it refers to, and the second because more specific objects seem more "important".

When a phrase is first encountered in a text or utterance, it will have to be decided that this phrase should be interpreted in a particular CKS (or group of CKSs). Other CKSs in the Candidate CKSs of the words of which the phrase is composed, corresponding to other senses of the words in the phrase, are not relevant to this particular use. Those other CKSs should receive no further activation when the object or action described by the phrase is mentioned again, whether explicitly or anaphorically. Indeed, work of Waltz and Pollack (1985), and prior simulations performed by cognitive modellers, suggests that winner-take-all mutual inhibition between alternative senses of words is a suitable disambiguation scheme in a neural net style architecture. Even in the sequential architecture for which the current context tracking algorithm is designed, it may be appropriate to inhibit CKSs which have been found inappropriate when disambiguating some phrase. If activation levels are allowed to take on negative values, it is necessary to retain in the ACKS all CKSs whose activation level is non-zero. These principles would have less force in natural language passages which discuss meanings of words.

3.4 Special CKSs

It may be appropriate to devise special CKSs to deal with modules of knowledge which have special roles in relation to language understanding, such as the CKS which knows about TALK. One thing which TALK might know about would be the fact that passages describing conversation may continually remind you that talking is going on, using words like reply, shout, plead, whereas in fact the text is less about talk per se, than about the topic the talkers are discussing.

A system built on these principles would most likely also have one or more permanently active CKSs to deal with closed category words such as prepositions and conjunctions.

4. IMPLEMENTATION

Many of the ideas in section 3 have had two trial implementations in programs which used a small number of domains and a small vocabulary, and only tried to decide in which CKS a particular use of a word should be interpreted. Those implementations worked reasonably well, (though the more recent one tended to find unique interpretations for genuinely ambiguous sentences). Work is under way on a more extensive implementation with a larger vocabulary and making
use of activation levels directly, rather than simulating them with stacks and lists, as was done in the earlier version.

ACKNOWLEDGEMENTS

Part of this work was done while the author was visiting the Cognitive Studies Centre at the University of Essex. The Centre's hospitality is gratefully acknowledged.

REFERENCES


