Atention-driven Perceptual Classification and Reasoning

Simon Dobnik* **and John D. Kelleher†**

*CLASP
University of Gothenburg, Sweden
simon.dobnik@gu.se

†School of Computing
Dublin Institute of Technology, Ireland
john.d.kelleher@dit.ie

Abstract

One of the central challenges for embodied and situated agents is how to process and react to a large amount of information that they receive at some regular rate through their sensors while perceiving their environment and interacting with others through language given that they have limited processing and memory resources. Secondly, we also need a model of information fusion: an account for how perceptual and linguistic/conceptual information interact and drive their reasoning. We present a computational model based on a cognitive notion of attention for perceptual classification in probabilistic Type Theory with Records (TTR), a formal meaning representation layer for such agents.

1. Type Theory with Records

Probabilistic Type Theory with Records (prob-TTR) (Cooper et al., 2015) has its origins in natural language semantics and views meaning assignment (both sense and reference) being in the domain of an individual agent who can make probabilistic *judgements* about situations (or invariances in the world) of being of types (written as $a : T$ with probability p). The type inventory of an agent is not static but is continuously refined through agent's interaction with its physical environment and with other agents through dialogue interaction which provides instances and feedback on what strategies to adopt for learning. Such view is not novel to mobile robotics (Dissanayake et al., 2001) nor to approaches to semantic and pragmatics of dialogue (Clark, 1996), but it is novel to formal semantics (Dowty et al., 1981; Blackburn and Bos, 2005) which represents important body of work on how meaning is constructed compositionally and reasoned about. The premise that sense is defined in terms of how an agent experiences language over perceptual scenes and interactions makes it a highly suitable knowledge representation framework for situated dialogue systems as it provides information fusion of perceptional and linguistic/conceptual information. As argued in (Dobnik et al., 2013) there are several advantages of such a unifying framework over a standard layered approach (Kruijff et al., 2007).

The rich type system of TTR gives us a lot of flexibility in modelling natural language semantics. For example, (i) most types are not atomic objects but complex record types that contain both real and nominal/conceptual (ii) dependent types; (iii) they are *intensional* which means that a situation may be assigned more than one type; and (iv) there is a structural relation between the types in the form of sub-typing. However, unfortunately, the flexibility with which types are assigned to records of situations and which is also required for modelling natural language and human cognition comes with a computational cost. Each type assignment involves a probabilistic classification (yes/no) which means that an agent with n types must make n judgements/classifications of each situation. Agents have limited processing and memory resources and therefore an optimisation mechanism is required that allows them prioritise what classifications to try first. Given the regularities in the physical world (spaces are associated with other spaces, objects and actions) and dialogue contexts (topics of conversations are related and predictable from other topics), judgements are not equally likely across dialogue and perceptual contexts. On the other hand, agents also access perceptual information that is supplied to them continuously, regardless of the context, as they must be aware of the world around them to preserve their existence. Hence, they have to make certain judgements all the time.

2. Attention-driven classification

This problem has been investigated in psychology as *attention*. For example, the Load Theory (LT) (Lavie et al., 2004) distinguishes between *perceptional selection* or bottom-up attention which is not under conscious control and is task independent and *cognitive control* or top-down attention which is consciously directed and task dependent/primed. The proposal views attention as a shared resource between tasks bound by the available resources and attention policy. Hence, during intensive activity related to cognitive control, activity from perceptional selection is reduced and vice versa. Following this division of attention we can talk about two kinds of attention-driven judgements, each with its own control mechanism. *Pre-attentive judgements* are related to types that are defined by agent's biology and embodiment: they are classifications made by its sensors and actuators. As such they are basic (not learned) and finite in number. Classifications of these types are fundamental to agent's basic operation and therefore are made continuously at the rate determined by LT. On the other hand *task and context primed judgements* are driven by a priming mechanism that predicts what types an agent should expect in the current state based on its knowledge

about the world. The priming is driven by discovery of *thematic relations:* spatial, temporal, causal or functional relations between individuals occuring in the same situations (Lin and Murphy, 2001; Estes et al., 2011). The idea is analogous to the notion of *type resources* (Cooper, 2008), bundles of types that are employed and learned in different situational contexts.

3. Computational model for attention-driven judgements

We propose a computational model for attention-driven judgements which comprises of two parts: (i) creation of thematic relations between task/context types; (ii) priming mechanism for task/context types based on perceptional and cognitive contexts.

3.1 Creation of thematic relations

An agent experiences the world through perception, embodiment and linguistic interaction. Experiencing perceptual and linguistic contexts an agent forms associations between types co-occuring in its memory or information state (IS). Type associations can be seen as cognitive states. A particular type may be associated with more than one state. There are several computational mechanisms that could be used to automatically create states (clusters of types) with the above properties from data, for example *Latent Dirichlet Allocation (LDA)* (Blei et al., 2003): associating words in documents with topics is analogous to associating types in agent's memory with cognitive states. Our model requires (i) that an agent has discovered a finite set of states, e.g. $\mathcal{S}1$, $\mathcal{S}2$ and $\mathcal{S}3$ with known prior probabilities; (ii) a stochastic matrix defining a probabilistic relationship between each of the pre-attentive types, e.g. F1, F2 and F3 (low-level perceptual features whose occurrence is biased to particular physical environments) and a state $P(type|\mathcal{S}i)$; (iii) a stochastic matrix defining a probabilistic relationship between each of the task/context types, e.g. WEATHER, MACH.-LEARN. and WELL-BEING ("topics" discussed in particular discourse and physical contexts) and a state $P(type|\mathcal{S}i)$; and (iv) a stochastic state transition matrix which defines a probability of transition from state *i* to state *j* in each time step.

3.2 Priming of task/context judgements: update mechanism for state-space

Priming of task/context judgements works as follows. As stated earlier, pre-attentive judgements (*Pre*) are made continuously at each time step t, only task/context judgements (*Task* and *Cont*) are primed. Hence, we require an update mechanism for the posterior distribution over states of task/context types $P(s_t|Pre_t, Task_{t-1}, Cont_{t-1}, AS_{t-1})$ after encountering new perceptual (Pre_t) and dialogue contexts ($Task_{t-1}$ and $Cont_{t-1}$), successful classifications of those types. We also need a mechanism for reduction of the state-space of task/context types at a rate determined by the LT and a selection of types to be primed for.

The current state-space is conditionally dependent on the (active) state space at $t-1$ (AS_{t-1}) and the probabilities that condition observed types ($Task_{t-1}$, $Cont_{t-1}$ and Pre_t) with states. This way, the more judgements an agent makes, the more it reduces its ambiguity of being in several states. $Task_{t-1}$ and $Cont_{t-1}$ are classifications that an agent has successfully made following the priming in the previous time step, while Pre_t is new perceptual evidence, as the agent might have also changed its location. The probability distribution over states at t is:

$$\begin{aligned} P(s_t|Pre_t, Task_{t-1}, Cont_{t-1}, AS_{t-1}) = \\ \eta \times P(Pre_t|s_t) \times P(Task_{t-1}|s_t) \\ \times P(Cont_{t-1}|s_t) \times P(AS_{t-1}|s_t) \\ \times P(s_t) \end{aligned} \quad (1)$$

The equation is assuming independence of conditioned events. $P(Pre_t|s_t)$, $P(TaskCont_{t-1}|s_t)$, $P(AS_{t-1}|\mathcal{S}_t)$ and $P(s_t)$ are obtained from our model of thematic relations between types described in Section 3.1 (for the matrix $P(AS_{t-1}|\mathcal{S}_t)$ we have to apply Bayes' rule to revert it from $P(\mathcal{S}_t|AS_{t-1})$). Each factor on the RHS of Equation 1, $P(Pre_t|s_t)$, $P(TaskCont_{t-1}|s_t)$, $P(AS_{t-1}|s_t)$ and $P(\mathcal{AS}_{t-1}|\mathcal{S}i_t)$ expands into several conditionally independent factors depending on the evidence observed ($Task_{t-1}$, $Cont_{t-1}$ and Pre_t) and the current state of active states. For example, if there are three active states:

$$\begin{aligned} P(\mathcal{AS}_{t-1}|\mathcal{S}i_t) = P(\mathcal{AS}1_{t-1}|\mathcal{S}i_t) \\ \times P(\mathcal{AS}2_{t-1}|\mathcal{S}i_t) \times P(\mathcal{AS}3_{t-1}|\mathcal{S}i_t) \end{aligned}$$

η denotes a normalisation process that ensures that the total probability mass of the posterior distribution sums to 1.

After determining poster probabilities over states we need to select types to be primed. The most straightforward solution is to select a state $s_t \in S$ with the maximum a posteriori probability and load the types from s_t into short-term memory. However, there are two disadvantages to this approach: (i) the agent assumes it is only in 1 state and (ii) it may end up switching between two states that have similarly high probability. A more sophisticated solution is to (i) rank the states by posterior probability; (ii) prune the state-space according to $P(s_t)$ by a threshold (θ_{AS}) determined by available resources and LT, where high cognitive load corresponds to high threshold. The states that are left after pruning are our active states for which we (iv) renormalised probabilities so that they sum to 1. Using the set of active states $\mathcal{AS}_t$ we (v) compute a posterior probability over the set of types in $\mathcal{AS}_t$ using a *Bayes Optimal Classifier*. Here, several states may be maximising a probability of a particular type and hence the system is more stable in making decisions than the one using *argmax*. However, the approach is computationally more demanding as we are calculating posterior probabilities over types in active states rather than probabilities of states. However, similarly as before (vi) pruning that is sensitive to a threshold determined by the LT and available resources (θ_T) can be applied to types to determine a set of active types ($\mathcal{AT}$). Finally, we (vii) normalise the posterior probabilities of active types and load them into working memory where they can be applied to classify new tasks/contexts. In step (v) the posterior distribution for *type* where $P(type|\mathcal{AS}i) > 0$ is calculated as follows:

$$P(type_t|Pre_t, Task_{t-1}, Cont_{t-1}, AS_{t-1}) = \sum_{s \in AS_t} P(type|s) \times P(s|Pre_t, Task_{t-1}, Cont_{t-1}, AS_{t-1}) \quad (2)$$

where $type_t$ denotes a type at time t, AS_t denotes the set of unpruned (active) states at time t, $P(s|Pre, Task, Cont, AS_{t-1})$ denotes the probability of an active state s after the state set has been pruned and the posterior probability over the active states has been renormalised.

4. Conclusions and future work

We have presented attention-driven type judgements in an agent interacting through perception and dialogue based on discovery of thematic relations and sharing of cognitive resources. The agent maintains (i) a distribution over set of cognitive states and (ii) a distribution over set of types in the active states. The number of active states (and types) is controlled by available cognitive resources that are shared between perceptual selection and cognitive control. The more task/context judgements an agent makes the more it reduces ambiguity of being in several states.

The proposal is not exclusive to the TTR framework we are using. It is a general solution for agents making classifications, an area of research in robotics that is sometimes called *visual search* (Sjöö, 2011; Kunze et al., 2014). The approach could also be applied in situated dialogue systems for disambiguation of speakers utterances/topic priming or for generating new utterances/topic modelling.

In our forthcoming work we will focus on studying the effects of the parameters that the model introduces in a computational simulation: the number of states, the number of types and the size of memory/processing resources.

References

Patrick Blackburn and Johan Bos. 2005. *Representation and inference for natural language. A first course in computational semantics*. CSLI Publications.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning research (JMLR)*, 3:993–1022.

Herbert H. Clark. 1996. *Using language*. Cambridge University Press, Cambridge.

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. 2015. Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology — LiLT*, 10(4):1–43, November.

Robin Cooper. 2008. Type theory with records and unification-based grammar. In Fritz Hamm and Stephan Kepser, editors, *Logics for Linguistic Structures. Festschrift for Uwe Mönnich*, volume 201, page 9. Walter de Gruyter.

M. W. M. G Dissanayake, P. M. Newman, H. F. Durrant-Whyte, S. Clark, and M. Csorba. 2001. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotic and Automation*, 17(3):229–241.

Simon Dobnik, Robin Cooper, and Staffan Larsson. 2013. Modelling language, action, and perception in Type Theory with Records. In Denys Duchier and Yannick Parmentier, editors, *Constraint Solving and Language Processing: 7th International Workshop, CSLP 2012, Orléans, France, September 13–14, 2012, Revised Selected Papers*, volume 8114 of *Lecture Notes in Computer Science*, pages 70–91. Springer Berlin Heidelberg.

David R Dowty, Robert Eugene Wall, and Stanley Peters. 1981. *Introduction to Montague semantics*. D. Reidel Pub. Co., Dordrecht, Holland.

Zachary Estes, Sabrina Golonka, and Lara L Jones. 2011. Thematic thinking: The apprehension and consequences of thematic relations. In Brian Ross, editor, *The Psychology of Learning and Motivation*, volume 54, pages 249–294. Burlington: Academic Press.

Geert-Jan M. Kruijff, Hendrik Zender, Patric Jensfelt, and Henrik I. Christensen. 2007. Situated dialogue and spatial organization: what, where... and why? *International Journal of Advanced Robotic Systems*, 4(1):125–138. Special issue on human and robot interactive communication.

Lars Kunze, Chris Burbridge, and Nick Hawes. 2014. Bootstrapping probabilistic models of qualitative spatial relations for active visual object search. In *AAAI Spring Symposium 2014 on Qualitative Representations for Robots*, Stanford University in Palo Alto, California, US, March, 24–26.

Nilli Lavie, Aleksandra Hirst, Jan W de Fockert, and Essi Viding. 2004. Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, 133(3):339–354.

Emilie L. Lin and Gregory L. Murphy. 2001. Thematic relations in adults' concepts. *Journal of experimental psychology: General*, 130(1):3–28.

Kristoffer Sjöö. 2011. *Functional understanding of space: Representing spatial knowledge using concepts grounded in an agent's purpose*. Ph.D. thesis, KTH, Computer Vision and Active Perception (CVAP), Centre for Autonomous Systems (CAS), Stockholm, Sweden.