Online Prediction of Activities with Structure: Exploiting Contextual Associations and Sequences

Nicholas H. Kirk[+*], Karinne Ramírez-Amaro[*], Emmanuel Dean-León[*], Matteo Saveriano[+] and Gordon Cheng[*]

Abstract— **Many human activities, given their intrinsic modularity, present structural information which can be exploited by classification algorithms: this enhances the capability of robots to predict activities. We introduce a semantic reasoning paradigm in which, via logical and statistical learning, we discriminate between actions on the basis of contextual associations. An example of this is considering the co-occurrence of scenario objects when predicting an action. We also combine such probabilistic reasoning with traditional sequence likelihood modeling. The system, given partial execution evidence of a task (e.g. assembling a car), first reasons in logical terms over qualitative primitives to constrain the space of possibilities, and then predicts the most sequentially likely action (e.g. 'PickAndPutScrew'). A further claim is also the representation of actions in tractable logic, enabling online-capable recognition. Our evaluation, adopting annotated primitives of motion and tool usage, proves that simple sequence-only prediction methods (i.e. bigram sequence information, 59.80%) are outperformed by the proposed polynomial-time context- and sequence- aware inference (i.e. with 8 primitives, various degrees of partial evidence and bigram sequence information, 78.43%), proving the effectiveness of the combined approach.**

I. INTRODUCTION

Artificial assistants require to effectively classify actions performed by humans in the shortest time frame possible. A first prior question is how to represent such action concepts: today activity classification algorithms which consider actions and objects combined lack context generalizability as they mainly work with trajectory information [1], [2]. It is apparent that while such generalization at a sensorimotor level encapsulates important information for task reproduction, robustness towards context variation (in terms of adopted objects and their parameterization) can only be achieved at a higher level of abstraction, i.e., by making use of semantic notions and reasoning, as currently investigated by some lines of research [3], [4]. However, the prediction of actions exploiting the intrinsic high logical modularity has not been tackled from a combined statistical and logical perspective.

In this work, we make use of action sequence statistics and semantic reasoning over contextual relations and qualitative discrete motions for recognizing executed actions (e.g. Flip, PutScrew). This is useful in human-robot task collaboration or human intention understanding

[+]Department of Electrical and Computer Engineering, Institute of Automatic Control Engineering, Technical University of Munich, Germany {nicholas.kirk, matteo.saveriano}@tum.de

[*]Department of Electrical and Computer Engineering, Institute for Cognitive Systems, Technical University of Munich, Germany {karinne.ramirez, dean, gordon}@tum.de

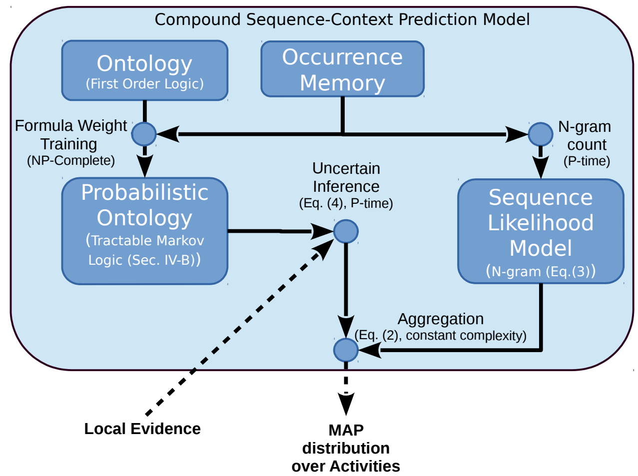


Fig. 1: Block diagram of the polynomial-time context- and sequence-based action recognition model.

[5]. For use in practical settings, semantic representations allow for context generalization as they disregard non-salient details [6]. A representation of the implemented system is illustrated in Fig. 1. In the system, ontological knowledge and sequences are weighted according to seen occurrence frequencies (from Occurrence Memory, creating Probabilistic Ontology and Sequence Likelihood Model respectively). After such training phase, both latter modules, given partial instance knowledge (i.e. Local Evidence), are able to infer independently the most likely action. Such distributions are then aggregated to obtain a final posterior probability (i.e. action MAP distribution).

The remainder of this paper first describes relations among current and past work (Sec. II) and provides an introduction to the terminology and adopted learning formalisms (Sec. III). We then provide an in-depth description of our novel compound prediction model, as well as its representation and execution complexity (Sec. IV-A and IV-B respectively), together with implementation and evaluation details (Sec. V) to prove the effectiveness of the logical reasoning and of the combined model. We conclude by summarizing our claims and discussing future opportunities (Sec. VI).

II. RELATED WORK

Many learning approaches concerning the construction and recognition of plans adopts generative (e.g. Hidden Markov Models [7], [8], Markov Random Fields [9]) or discriminative (e.g. Conditional Random Fields [10]) trajectory model learning. Such work does not consider context, focus of this work. Conversely, Object-Action Complexes (OAC) [1] consider the impact of executed sensorimotor instances on context in terms of causality, but do not use the contextual associations with objects during action prediction [11]. Furthermore, trajectory level representations are not able to properly handle multiple objects involved in an action, especially when interactions between them often differ in type. However, the presented work introduces semantic reasoning over discrete, qualitative motion primitives as features, which are seldom adopted [4]. Closer work identifies the need to predict sequences of actions, and partly understands the importance of context [12], however it does not employ discrete features which enable sensorimotor abstraction and consequent semantic reasoning. The closest work known to date [13] exploits statistical relational learning for identifying similar actions which present partial order variability, but does not consider associations with contextual objects for such recognition. To our knowledge no work provides training and inference means of structured logic sequences representing actions, given both context and partial motion evidence.

III. ADOPTED CONVENTIONS AND FORMALISMS

We now describe the adopted conventions and learning features which constitute action plans (Sec. III-A), to then present two pre-existing, widely adopted probabilistic modeling approaches (Sec. III-B) which are used as basis for the compounded model presented thereafter (Sec. IV-A).

A. Motion Primitives

This work makes use of discrete, sequential primitives as basic constituents of actions, which intend to describe a qualitative variation *with respect to the previous state of the instance*. These have been successfully segmented from movement for human action description, and some of the ones adopted within this scope are presented in more detail in [6]. More specifically, we define the following:

`Motion related primitives`

- $\mathrm{Move}(o)$: the object 'o' has begun a motion
- $\mathrm{NotMove}(o)$: the object 'o' has stopped its motion

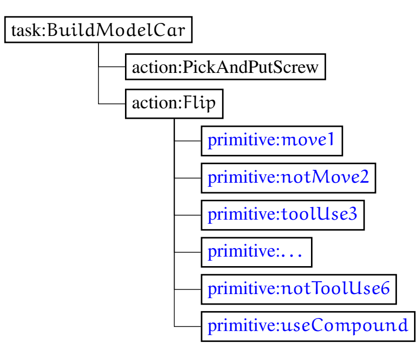


Fig. 2: Exemplification of the scope's terminology.

- $\mathrm{ToolUse}(o,i)$: an object 'o' has engaged in interacting with an object 'i'
- $\mathrm{NotToolUse}(o,i)$: an object 'o' has stopped an interaction with an object 'i'

`Object use related primitives`
(assembly scenario examples, as in use case of Sec. V-B)

- $useScrew(o)$: the tool in use is a Screw instance
- $useChassisBar(o)$: the tool in use is a Chassis instance
- $useCompound(o)$: the tool in use is a Compound instance
- $useWheelCompound(o)$: the tool in use is a Screw-Wheel compound instance

These primitives are generated by a perception segmentation module, and are basic constituents of actions (e.g. $PutScrew$), which in turn compose tasks (e.g. $BuildModelCar$, full hierarchy example in Fig. 2). For example, from an initial state ($notMove1$) any motion would be perceived as a sequential variation of state ($move2$), where the numbering at the end of the primitive is a convention for defining the discrete time instant. Any fully stationary instant after such movement would again be considered a variation ($notMove3$). Likewise reasoning applies to tool usage. It is noteworthy that the here presented system can elaborate an arbitrary number of primitives, and the framework is agnostic towards the nature and meaning such primitives have. Such priorly designed relations could describe contextual aspects or any arbitrary information available from the sensor array. The only constraint is that such properties have to be in a symbolic form. The segmentation of such labels from sensory arrays has to be catered by a prior module external to this scope (such as [14]). This work considers the informativeness analysis of primitives, as well as segmentation performance details, as out of scope. The chosen type and amount of primitives here used is a minimal demonstrative example.

B. Logico-Statistical Learning

In action recognition we need to predict sequences of actions, i.e. sequences of logic predicates, therefore we look at combined statistical and logical formalisms for estimating the confidence of an individual action (i.e. via statistical relational learning, by providing context understanding), to then combine it with information regarding sequentiality.

a) Statistical Relational Learning: Markov Logic Networks (MLN) [15] is a knowledge representation formalism which enables probabilistic learning and inference via the combined use of first order logic and probabilistic undirected graphical models (i.e. *Markov Random Fields*). More formally, we can define a probability over the world x as a log-linear model in which we have an exponentiated sum of weights w_j of a binary feature f_j, and the partition factor Z:

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_j w_j f_j(x)\right) \tag{1}$$

In our case, we consider the binary formula $f_j(x)$ as an evaluation of a logic relational formula which comprises information regarding our motion and tool usage relations, and we substitute such term with $n_j(x)$, where the latter is the number of true groundings of such formula f_j in x_j. By defining informative context-aware relations (which can be devised on the basis of the domain of application), we exploit MLN to infer the likelihood of occurrence of our actions, without taking into account sequence information, given contextual information (see likelihood term in Eq. 4, e.g. we compute the probability of an action occurring, given the objects in use). A variant of such relational model is used in this work to represent all the action definitions and their likelihoods in a probabilistic fashion (Sec. IV-B).

b) the n*-gram language Model:* is an efficient probabilistic modeling of sequences of symbols of arbitrary nature. It first appeared for natural language modeling [16], to then be used in DNA and protein sequencing, and can be applied to statistically model sequential, discrete processes. The present work uses such model for describing the *sequentiality of actions which compose a task* (see likelihood term Eq. 3).

IV. PREDICTION MODEL

We now introduce our model for sequence prediction given the context, obtained by compounding probability distributions over both the *sequence likelihood* (Eq. 3, via n-gram modeling) and *context-based likelihood* (Eq. 4, inferred via Markov Logic Networks) (Sec. IV-A), to then discuss representational and inference complexity (Sec. IV-B).

A. Compound Prediction

The presented model, during execution, exploits relations of objects which usually co-occur with the executed actions. It first discards all candidates which are inferred as impossible due to contextual conditions, to then predict, given such sparse array of remaining hypotheses, the most sequentially likely action. For example, within a vehicle assembly task, if no wheels are present in context, it follows that we cannot be building a 4-wheeled vehicle model, even if the latter might be more statistically likely in terms of sequence occurrence. This is specifically designed in view of computing inference on a high number of action candidates, as often is the case in real-world ontologies (evaluation of such feature is in Sec. V-C (a)). A further potential of the model is that if the logical associations are not unique, and therefore the maximum a posteriori estimate does not comprise only one candidate, we will have a re-partition of the posterior probability density over the candidates which have not yet been estimated as impossible (evaluation of such feature is in Sec. V-C (b) and illustrated in Fig. 6): at that stage, the sequence likelihood will further discriminate on the basis of past occurrence frequencies. Let Task_i be composed by actions $\mathrm{Action}_{0_i} \dots \mathrm{Action}_{m_i}$, for $i \in T$, where T is the set of all known task concepts. Also let $n > 0$, where $n \in \mathbb{N}$ is the order of approximation which can be varied on the basis of how many past elements should be considered in the likelihood computations (i.e. by relaxing the Markov property). Then the constructed independent probability of the task occurring is:

$$P(\mathrm{Task}_i) = P(\mathrm{Action}_{(0_i)}, \dots, \mathrm{Action}_{(m_i)}) \quad (2)$$

$$\approx \prod_{j=0}^{m} \Big\{ \underbrace{P\left(\mathrm{Action}_{(j_i)} \middle| \, \mathrm{Action}_{(j_i-n)}, \dots, \mathrm{Action}_{(j_i-1)}\right)}_{\text{sequence likelihood (via } n\text{-gram modeling)}} \quad (3)$$

$$\cdot \underbrace{P\left(\mathrm{Action}_{(j_i)} \middle| \, \mathrm{Context}, \mathrm{Ontology}\right)}_{\text{context-based likelihood (via statistical relational learning)}}. \quad (4)$$

which is presented as the product of our n-gram sequence modeling (3) by our MLN probabilistic logic reasoning (4), where *context* is the instance's partial motion evidence and observed tool usage (described as primitives, as in Sec. III-A), and *ontology* is the set of previously learned actions.

B. Representation for Real-Time Recognition

We now describe how we meet timing and inference exactness requirements, as well as how we implement class concepts and their inference. As per intuition the recognition of structures within sequences of segmented primitives is computationally complex. We look at the complexity of our compound model (2), and notice that the complexity of the n-gram modeling (3) is computable in polynomial time for any arbitrary n, in terms of both training and inference [16]. However, context-based likelihood (4) has non-efficient complexity ($NP-\mathrm{complete}$ for both training and inference, i.e. no polynomial algorithm is yet known [17], [18]). In the domain of cognitive robotics and machine learning, this is often partly remedied by computing approximated or lifted inference [19]–[21]. However these provide only average case efficiency increase, but not the upper-bound constraints on execution times necessary for real-time recognition.

One of the advances here presented is the online capability brought by polynomial inference time ($P - \mathrm{time}$), and the syntactic conditions for polynomial training time [22].

a) Syntactic Restriction: Our timing claims are enabled by a non-restrictive consideration: our semantic representation as well as many knowledge intensive applications require only predicate relations between *actions* (taxonomy), *actions and their constituting objects* (meronymy), and between *motions and their formal object arguments*, i.e. objects which are subject or instrumental to the motion. It is noticeable that the logic expressiveness needed for these is confined to an efficient subset of probabilistic first order logic, i.e. expression of hierarchy relations and relations among same-level constituents (e.g. respectively, the action $\mathrm{PickAndPutScrew}$ is child of $\mathrm{PickAndPlace}$, and Obj relates to $\mathrm{RightArm}$, which are both child nodes of $\mathrm{PickAndPutScrew}$, via the motion relationship $\mathrm{toolUse(object, part)}$, as shown in Fig. 3). We use an existing tractable subset of MLN which suits all specified requirements, called *Tractable Markov Logic* (TML) [22]. The main syntactic restriction imposed by TML (i.e. that relations have to be present only among entities which are

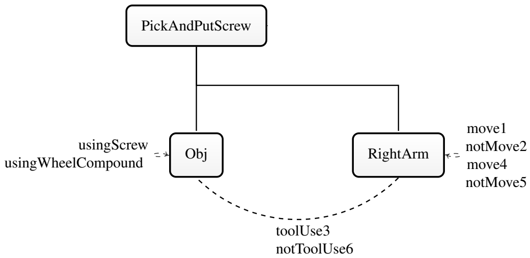


Fig. 3: Structure of an action class concept modeled in Tractable Markov Logic (TML).

children of the same father node) does not constrain the expressiveness of our solution. We implemented two modeling approaches in the TML formalism: the first approach (henceforth $\mathsf{CModel1}$, example in Fig. 3) describes action concepts as containers of parts of the robot and of contextual objects (as *class objects*, e.g. $\mathsf{RightArm}$ and Obj respectively), while any qualitative primitive is a *relation* among such objects (e.g. $\mathsf{toolUse3}(\mathsf{object}, \mathsf{part})$ as relation between Obj and $\mathsf{RightArm}$), or characteristic of a single object (e.g. $\mathsf{usingScrew}$ as label of Obj). Object label primitives have been introduced to overcome subclass grounding problems in the current TML implementation. Conversely the second model (henceforth $\mathsf{CModel2}$), in addition to what has been just stated, also explicitly defines which relations *do not* occur (e.g. $\mathsf{move1}(\mathsf{part})$, $!\mathsf{toolUse1}(\mathsf{object}, \mathsf{part})$, $!\mathsf{notMove1}(\mathsf{part})$, $!\mathsf{notToolUse1}(\mathsf{object}, \mathsf{part})$). We derive the context-based likelihoods (Eq. 4) via Maximum A Posteriori estimation over such relations.

V. EVALUATION

We now describe our implementation details (Sec. V-A), to then show our evaluation results for context understanding (Sec. V-C) and for the latter and sequence prediction combined (Sec. V-D), which makes use of the annotated data deriving from our assembly task use case (Sec. V-B).

A. Implementation Details

In order to compute the n-gram based likelihoods (Eq. 3 of the compound model) we consider action frequency counts, utilizing bigram instances based on sequential maximum likelihood estimation. For our problem, we view the entirety of possible actions as a formal grammar with closed world assumption, which is representative of our transitions between known actions [23]. The latter can be implemented as weighted finite state transducers, and for this we specifically use *OpenGrm* [24]. Our implementation is therefore an overlay of the latter for sequence MAP estimation and of Alchemy Tractable Markov Logic [22] for the context-based MAP estimation. The latter however only provides implemented tractable inference, while weight training algorithms are not yet efficient and are the same of Alchemy MLN.

B. Use Case and Adopted Data

Within the context of Programming by Demonstration (PbD), in order to teach an action concept, humans teach multiple sensorimotor instances to robots [25]. Some approaches incorporate active learning into the paradigm, so that the human can perform a critic selection of the most informative trajectories [26]. We hypothesized a use case in which a human demonstrator, via kinesthetic teaching, demonstrates a series of trajectories which are segmented by the robot (into $\{\mathsf{move}, \mathsf{notMove}, \mathsf{toolUse}, \mathsf{notToolUse}\}$). In the envisioned system the robot then predicts, at every new piece of primitive evidence, the most likely action which is executed (as in Fig. 4). As this is thought for long sequences, the system can be used for reducing training times, if the robot intervenes when highly confident to declare that an action concept is already known.

In order to retrieve a training and testing dataset of primitives, we partly constructed the use case scenario now described and then annotated by hand video-only data with primitive information (therefore performing manually the segmentation of Fig. 4). Four video samplings ($\sim$ 6 minutes of length each) were performed on a Willow Garage Personal Robot 2 (PR2) in gravity compensation mode, whose arms were moved by a human to construct car models made out of wooden components (see Fig. 5). Such scenario has been constructed only for video annotation and not robot execution purposes.

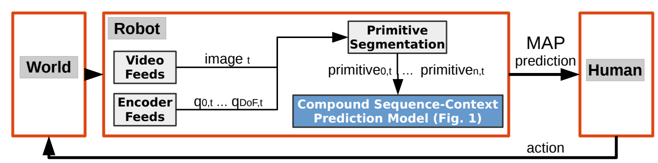


Fig. 4: Block diagram of the envisioned use case of 'discrete active learning by demonstration' scenario.

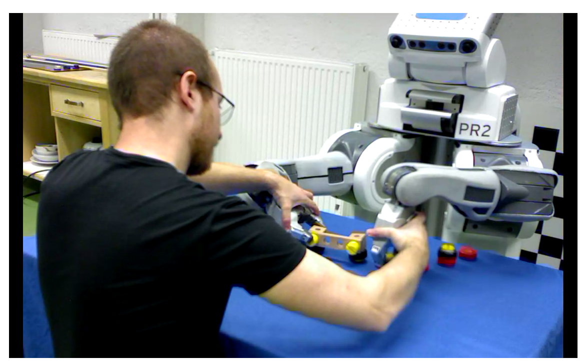


Fig. 5: Image captured from an assembly domain training scenario, comprising a PR2 robot, a human demonstrator, and a series of wooden objects which will be compounded to obtain a model car, focus of the assembly task.

During the manual annotation of the video samples it was observed that many motions performed by the human do not contribute to the population of the ontology, and were therefore discarded (such as backtracking due to erroneous movements). We consider all four sampling sessions as one single, sequential training session, in which 8 distinct actions were learned and repeated (i.e., numbered for reference in the text and ordered by frequency of occurrence in all tests: $\mathsf{ActionDataSet} = \{0\ \mathsf{PickAndPutScrew},\ 1\ \mathsf{PickAndHold},\ 2\ \mathsf{Release},\ 3\ \mathsf{PickAndPlace},\ 4\ \mathsf{Flip},\ 5\ \mathsf{PutScrew},\ 6\ \mathsf{FlipAndRelease},\ 7\ \mathsf{FlipAndHold}\}$).

C. Reasoning Effectiveness

a) Unique Associations: We first prove the concept of the impact of logic by providing as evidence relations which occur only once in the ontology. By performing Maximum A Posteriori (MAP) estimation on the grounded Markov random field network, we verified if MAP identified the only action which contains such relation (e.g., in our dataset, relation $\mathsf{notToolUse1}(\mathsf{object}, \mathsf{part})$ only occurs in action $\mathsf{Release}$). The MAP test for $\mathsf{CModel2}$ yielded 100% prediction effectiveness for each action concept containing a unique relation, while $\mathsf{CModel1}$, due to the unbalanced amount of relations for each action concept impacting on the computation of the partition function Z, lacked scalability and did not yield meaningful results for the present test.

b) Density Partitioning: The majority of actions could not be discriminated by unique relation occurrence (e.g. Flip and $\mathsf{PickAndPutScrew}$ action definitions contain 6 identical non-grounded primitives). We added contextual object use relations (e.g. $\mathsf{usingScrew}$) to implement $\mathsf{CModel2}$. The tests that now follow were performed by giving maximum 3 sequential primitive relations (taken from the set $\{\mathsf{move},\ \mathsf{notMove},\ \mathsf{toolUse},\ \mathsf{notToolUse}\}$) and further object use relations (taken from the set $\{\mathsf{usingScrew},\ \mathsf{usingChassisBar},\ \mathsf{usingWheelCompound},\ \mathsf{usingCompound}\}$) as local evidence (the average number of primitives for each action concept of the dataset is 4.31). In particular, we tested context understanding over every individual action, imposing equal occurrence probability. The information of object use enabled $\mathsf{CModel2}$ to correctly discriminate 64.70% of annotated actions (confusion matrix in Fig. 7a). We tested further, this time exploiting independent

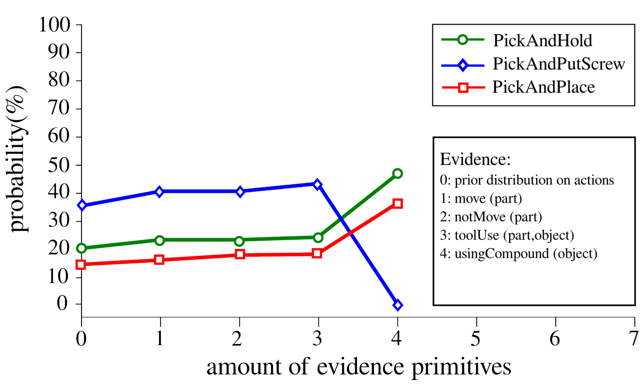
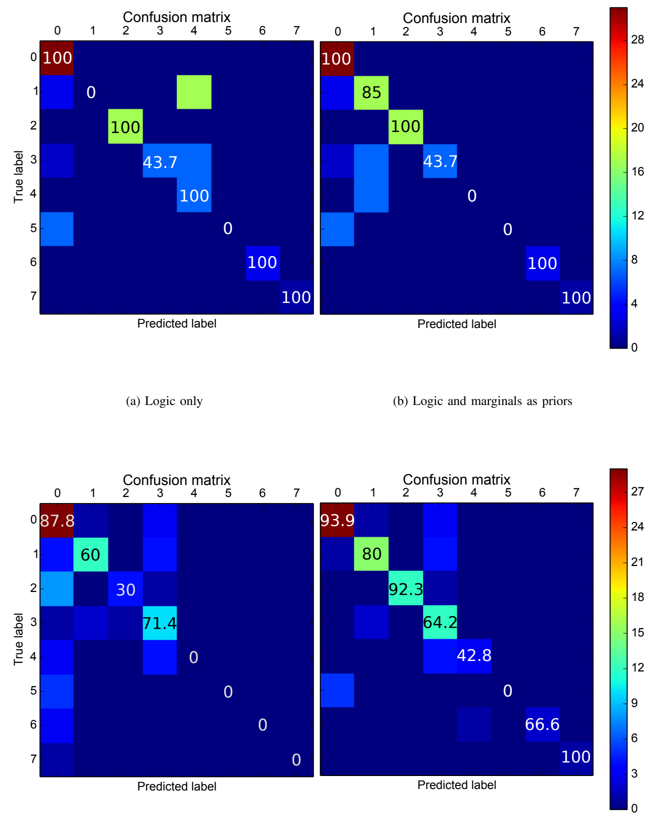


Fig. 6: Tree graph representing the variation of probability density for action candidates upon any new introduction of primitive evidence in a MAP instance, when testing with prior action existence information and with the use of logic.



(a) Logic only

(b) Logic and marginals as priors

(c) Bigram sequence likelihoods

(d) Full bigram sequence + grounded TML model

Fig. 7: Confusion matrices from Maximum A Posteriori estimation over logic only ontology (a), and over logic and marginals ontology (b), over bigram sequence likelihoods (c), and over the full bigram sequence and grounded TML model (d), given local evidence of maximum three motion primitives and object use.

probability of actions learned from the dataset (example of MAP variation with different degrees of evidence in Fig. 6). This allowed $\mathsf{CModel2}$ to correctly discriminate 74.50% of annotated actions (confusion matrix in Fig. 7b).

The inference time with the presented TML system was on average 0.2s on a 4GB Core i5 system, versus the 14.5s required for the same computation in MLN. Given the computational complexity analysis presented in Sec. IV-B, this difference holds in qualitative terms on all test datasets and grows exponentially with the size of the ontology [17].

D. Sequence And Full Model Forecasting

A 2-order ngram model generated from the dataset yielded a table of 29 bigram entries, each showing the likelihood of every action $a \in \mathsf{ActionDataSet}$, conditioned over all possible predecessors. Use of bigrams in sequence modeling is frequent in practical settings as higher order sequences are more computationally expensive (still polynomial in complexity terms, but present a higher value). MAP estimation over such conditional probabilities for all annotated instances of the dataset yielded 59.80% correct recognition (confusion matrix in Fig. 7c). We then tested the full model, i.e. the stated sequence implementation (as in Eq. 3) combined with the context-based action confidence estimation (as in Eq. 4, and implemented as $\mathsf{CModel2}$ explained in Sec. IV-B, with partial evidence as in Sec. V-C), which yielded 78.43% correct recognition (confusion matrix in Fig. 7d). We have therefore highlighted the benefits of the compounded model estimation over the individual component contributions.

VI. CONCLUSIONS AND FUTURE WORK

In view of efficiently recognizing actions by exploiting contextual structure, we have introduced i) *a novel method to exploit semantic context-related associations for activity recognition*, as well as ii) a *real-time context-sequence compound prediction model*, which performs action inference in polynomial time. We have proven the effectiveness of such association-based discrimination, for uniquely-occurring relations and for density ri-partitioning. Our numerical evaluation shows the effectiveness of the combined sequence-context approach, which within the assembly domain and experiment conditions specified, sequence (i.e. 59.8%, with 8 actions and a bigram assumption) and context reasoning (74.5%, with 8 discriminating qualitative primitives) combined yield higher prediction capability (i.e. 78.4%). Such numerical result per se is data dependent, but proves that the aggregation scheme is able to meaningfully merge context-based and sequence information.

An important focus of future investigations will be to integrate deep learning likelihood prediction to further the current neuro-symbolic integration approach, and make use of very large datasets to verify the recognition potential of non-linear associations. Other research possibilities are:

a) bias the impact of logic over sequence: this can be achieved by reducing the variance of the relative weights describing independent action occurrence, or by performing smoothing of sequence likelihoods. This enhances recognition of tasks which present highly structured contexts.

b) sensorimotor information: this scope envisions coupling with OAC [1] to enable the presented generalization and timing claims, as well as embodied reproduction.

c) implementation of Human-Robot Interaction use cases: will be done starting with the explained example in Sec. V-B, making use of segmentation modules [6], and robot doubt verbalization [27], to verify the impact of inference complexity on turn-taking times.

d) recognition of action variation in quantity and order: will be integrated, as partly described in [13], with the present context- and object-related claims.

REFERENCES

[1] N. Krüger, C. Geib, J. Piater, R. Petrick, M. Steedman, F. Wörgötter, A. Ude, T. Asfour, D. Kraft, D. Omrčen, *et al.*, "Object–action complexes: Grounded abstractions of sensory–motor processes," *Robotics and Autonomous Systems*, vol. 59, no. 10, pp. 740–757, 2011.

[2] M. Wächter, S. Schulz, T. Asfour, E. Aksoy, F. Wörgötter, and R. Dillmann, "Action sequence reproduction based on automatic segmentation and object-action complexes."

[3] M. Tenorth and M. Beetz, "Knowrob: A knowledge processing infrastructure for cognition-enabled robots," *The International Journal of Robotics Research*, vol. 32, no. 5, pp. 566–590, 2013.

[4] K. Ramirez-Amaro, E.-S. Kim, J. Kim, B.-T. Zhang, M. Beetz, and G. Cheng, "Enhancing human action recognition through spatio-temporal feature learning and semantic rules," in *Humanoid Robots, 2013, 13th IEEE-RAS International Conference*, 2013.

[5] H. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," 2013.

[6] K. Ramirez-Amaro, M. Beetz, and G. Cheng, "Understanding the intention of human activities through semantic perception: Observation, understanding and execution on a humanoid robot," *Advanced Robotics*, vol. 29, no. 5, 2015.

[7] N. Landwehr, "Modeling interleaved hidden processes," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 520–527.

[8] I. Thon, B. Gutmann, and G. Van den Broeck, "Probabilistic programming for planning problems." in *Statistical Relational Artificial Intelligence*, 2010.

[9] S. D. Tran and L. S. Davis, "Event modeling and recognition using markov logic networks," in *Computer Vision–ECCV 2008*. Springer, 2008, pp. 610–623.

[10] B. Gutmann and K. Kersting, "Tildecrf: Conditional random fields for logical sequences," in *Machine learning: ECML 2006*. Springer, 2006, pp. 174–185.

[11] F. Wörgötter, A. Agostini, N. Krüger, N. Shylo, and B. Porr, "Cognitive agents-a procedural perspective relying on the predictability of object-action-complexes (oacs)," *Robotics and Autonomous Systems*, vol. 57, no. 4, pp. 420–432, 2009.

[12] V. Magnanimo, M. Saveriano, S. Rossi, and D. Lee, "A bayesian approach for task recognition and future human activity prediction," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*. IEEE, 2014, pp. 726–731.

[13] M. Tenorth, F. De la Torre, and M. Beetz, "Learning probability distributions over partially-ordered human everyday activities," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 4539–4544.

[14] K. Ramirez-Amaro, M. Beetz, and G. Cheng, "Automatic segmentation and recognition of human activities from observation based on semantic reasoning," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014.

[15] M. Richardson and P. Domingos, "Markov logic networks," *Machine learning*, vol. 62, no. 1-2, pp. 107–136, 2006.

[16] P. F. Brown, P. V. Desouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai, "Class-based n-gram models of natural language," *Computational linguistics*, vol. 18, no. 4, pp. 467–479, 1992.

[17] M. Jaeger, "On the complexity of inference about probabilistic relational models," *Artificial Intelligence*, vol. 117, no. 2, pp. 297–308, 2000.

[18] D. Roth, "On the hardness of approximate reasoning," *Artificial Intelligence*, vol. 82, no. 1, pp. 273–302, 1996.

[19] G. Van den Broeck, N. Taghipour, W. Meert, J. Davis, and L. De Raedt, *Lifted probabilistic inference by first-order knowledge compilation*, 2011.

[20] P. Singla and P. Domingos, "Memory-efficient inference in relational domains," in *AAAI*, vol. 6, 2006, pp. 488–493.

[21] ——, "Lifted first-order belief propagation," in *AAAI*, vol. 8, 2008, pp. 1094–1099.

[22] P. Domingos and W. A. Webb, "A tractable first-order probabilistic logic." in *AAAI*, 2012.

[23] K. Pastra and Y. Aloimonos, "The minimalist grammar of action," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1585, pp. 103–117, 2012.

[24] B. Roark, R. Sproat, C. Allauzen, M. Riley, J. Sorensen, and T. Tai, "The opengrm open-source finite-state grammar software libraries," in *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 2012, pp. 61–66.

[25] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.

[26] S. Chernova and M. Veloso, "Multi-thresholded approach to demonstration selection for interactive robot learning," in *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*. IEEE, 2008, pp. 225–232.

[27] N. H. Kirk, D. Nyga, and M. Beetz, "Controlled natural languages for language generation in artificial cognition," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE RAS, 2014, pp. 6667–6672.