Building and Learning from a Contextual Knowledge Base for a Personalized Physical Therapy Coach

Robert Fisher RWFISHER@CS.CMU.EDU
Thomas Kollar TKOLLAR@CS.CMU.EDU
Reid Simmons REIDS@CS.CMU.EDU
Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213

Abstract

Robotic systems that interact with humans must modify and personalize their behavior based on contextual information if we wish for them to be effective. In this paper, we present a novel learning framework for collecting and utilizing information in a *Contextual Knowledge Base*. This knowledge base can be used to represent information that is unique to a given user and environment, and the contents of the knowledge base can be used as input to a learning algorithm to personalize the behavior of the system to the current context. We present preliminary results when using this framework as part of a physical therapy exercise coach.

1. Introduction

Contextual and personal information is a critical element in human interactions but is often ignored by applications in the field of human-robot interaction. In particular, if a robot is tasked with coaching a user through physical therapy exercises, a stern approach may be most appropriate for some users in some contexts, while other situations may call for a gentler style of coaching. To accomplish this, the robot must first build a model of the current user and environment, and then learn the most appropriate action for the given context.

Contextual information can be collected using sensor processing algorithms applied to the output of sensors such as infrared range cameras, GPS, RGB cameras, or accelerometers. Some contextual information can also be acquired directly through language. For instance, if a dialogue agent wished to learn the gender or age of a conversation partner, it could conduct inference on audio and video data—or it could simply ask the user.

In this work, we propose methods for building and utilizing a Contextual Knowledge Base (CKB) that contains all the information an interactive agent has acquired about its user and the environment. Unlike large scale knowledge bases like NELL (Carlson, 2010) and Freebase (Bollacker et al., 2008), the information in a Contextual Knowledge Base must be collected in an online setting for each individual user of the system. The contextual knowledge base consists of a set of expressions from a machine-readable Meaning Representation Language (MRL). The expressions can be acquired through conversation, analysis of sensor data, or some combination of the two. Once the CKB has been assembled, it can be used as input to a learning algorithm which will predict the best action for the agent to take in the given setting.

As a testbed application for this context aware language system, we are investigating a virtual coaching system that will help people who have suffered from a stroke to conduct physical therapy exercises in their homes. This coaching system is be equipped with sensing capabilities provided by a Microsoft Kinect, which allows the system to visually perceive the environment, conduct activity recognition, and interact with users through a verbal dialogue. Table 1 shows some examples of expressions that may occur in the Contextual Knowledge Base. We see from these examples that time plays a more important role in a Contextual Knowledge Base than other knowledge bases. Some entries in the CKB, such as which activity the user is currently performing, will be true only for a very short time period. Therefore, it may be beneficial to index the knowledge base by time, though we do not address the temporal dimension of the CKB in this paper.

Proceedings of the 30^{th} International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28.

CKB Entry	MRL
User is 42 years old	user_age(42)
User is currently conducting exercise 4	current_activity(exercise4)
User reported pain level of 9 on the Likert scale during exercise 2	likert_pain(9, exercise_2)
User's arm was too high during exercise 1	incorrect(exercise1, high_arm)

Table 1. Contextual Knowledge Base Examples

There are many potential applications for a Contextual Knowledge Base. In this paper, we will discuss how to use the CKB to learn a policy for personalized interactions that the exercise coach takes with a given user. For example, the coach could request the user to conduct a variable number of reps of a given exercise depending the user's proclivity towards that exercise, the user's current physical or mental state, or the exercise goals set forth by a clinician. A clinician supervising the user's physical therapy could also query the knowledge base and concerning trends observed by the system during physical therapy could be brought to the clinician's attention.

In this paper, we propose using natural language as a means of acquiring information to be added to the knowledge base. Some information that we would like to track in the CKB, such as a record of the user's dietary habits, would be difficult or impossible to acquire using sensor processing. Because the CKB contains information represented in the same language as the semantic parser output, it is straight-forward to add this information directly to the database. Additionally, there are instances in which the correct parse for a given sentence will depend on the current contents of the CKB. For example, if the user says "I want to do this exercise", it is impossible to understand the referent of the statement without additional information. However, the activity recognition software can add an item to the CKB which tells us that the user has begun performing a leg exercise, which would allow us to disambiguate the meaning of the input. If we were given this sentence and the CKB shown in Table 1, the target MRL statement would be $user\_request(begin(exercise4))$.

Due to the limited amount of labeled textual and contextual data in the physical therapy coaching domain, a semantic parser for this application will be reliant on human annotators to provide training data for the system. Because annotation is a bottleneck in the training of the system, we also introduce a decision theoretic, active learning scheme to make efficient use of a small number of annotations applied to a large set of unlabeled data.

2. Related work

Work in building knowledge bases through information extraction often focuses on examining large corpuses of text data (Carlson et al., 2010; Bollacker et al., 2008). The items in the knowledge base are often thought of as the output of a semantic parser, but there has also been some work to use the contents of a knowledge base in the training of a semantic parser (Krishnamurthy & Mitchell, 2012).

There has been some recent work using sensing technology, particularly smartphones, to establish models of context (Fisher & Simmons, 2011; Lane et al., 2010). Most of these approaches do not utilize natural language understanding to acquire contextual information.

A great deal of work has focused on representing conversational context for dialogue systems (Morbini & Sagae, 2011), but there is a significant opportunity to couple work in sensory, context awareness with natural language dialogue systems.

Active learning has been a popular topic in machine learning for quite some time, but it has only recently begun to be popularized in natural language processing applications (Settles, 2010).

Other social coaching robots have demonstrated that personalization is an important factor in user acceptance(Fasola & Mataric, 2012; Kidd & Breazeal, 2008), but the personalization elements of these systems generally react only to a few, limited pieces of information such as the user's name. One study that directly compared a robot with personalized social interactions with generic social interactions demonstrated that the personalized interactions resulted in more active user engagement, better rapport, and increased levels of user cooperation(Lee et al., 2012).

3. Method

In this paper, we will expand on a previously established semantic parsing framework(Clarke et al., 2010) by incorporating a contextual knowledge base, CKB, as well as active learning to expedite training of the parser. In this section, we will give a basic overview of the semantic parser and describe how we have integrated the parser with the CKB. We will then discuss a method for building personalized models, which would allow an agent to select the most appropriate action for a given context according to a cost function.

3.1. Semantic Parsing with the CKB

In this semantic parsing task, we begin with an ontology, O, which consists of a set of concepts and functions operating over those concepts (such as the function $user\_age()$, or the concept $exercise2$). Dur-

Example 1	Example 2
Sentence	**Sentence**
How am I doing?	I would rate my pain as a 5 during exercise 2.
CKB	**CKB**
current_activity(exercise3)	current_activity(idle)
user_age(42)	user_gender(male)
likert_pain(exercise1, 9)	last_question(query(user_pain))
Target MRL	**Target MRL**
request(assess(exercise3))	likert_pain(exercise2, 5)

Table 2. Input Data Examples

ing parsing, we are given a natural language sentence, S, as well as the Contextual Knowledge Base—which consists of a set of statement created by composing elements of the ontology together. The goal is to construct a semantically equivalent MRL statement using the items in the ontology. Table 2 gives two concrete examples of sample input and desired output.

To accomplish the semantic parsing task, we will define two smaller two subtasks 1) determine which elements of the ontology are present in the target MRL statement 2) compose those elements together in a directed, acyclic arrangement. The first summation in the objective function shown in Table 3 relates to task 1, while the second summation relates to task 2.

We will first address how to determine which items from the ontology are present in the target MRL. We begin with a set of *constituents*, which consist of single words from the input sentence, phrases from the input sentence, and items from the contextual knowledge base. We will denote $X = (s, ckb)$ are the pair containing the input sentence and the current contextual knowledge base. X can be thought of as a structured set, which contains all the constituents, but also the syntax of the input sentence. Each constituent taken from X will have a similarity score for each item in the ontology. For constituents taken from the sentence, a lexical similarity function (Clarke et al., 2010) is used to compute this similarity. For example, if the word "car" appears in the sentence, it would have much higher similarity to an item in the ontology called "vehicle" than an item in the ontology called "house." Constituents taken from the contextual knowledge base will have high similarity to items in the ontology if the given item from the ontology is present in the CKB element. For instance, if the CKB tells us that the user is currently performing exercise 4, the likelihood the the concept $exercise4$ appears in the MRL increases. For each constituent c and ontology item s we denote the similarity score between these items as $\Phi_1(X, c, s)$. Additionally, there will be a weight $w_{c,s}$ which is learned with annotated training data using a structured support vector machine.

Now we consider composing the items from the ontology together to form an expression in the MRL. We will leverage a structured ontology with type constraints. For instance, we would know *a priori* that the item *user_age* in the ontology is a function that takes a number as an argument. Therefore, a pair of items from the ontology, s and t, can only be composed together as $s(t)$ if s is a function that can take t's type as an argument. Given constituents, c and d, and ontology item s and t, we denote the composition score of $s(t)$ as $\Phi_2(X, c, d, s, t)$. This composition score is always 0 if $s(t)$ is not a valid composition. If $s(t)$ is valid, and the constituents c and d are taken from the sentence, $\Phi_2(X, c, d, s, t)$ will be the distance in the dependency tree between c and d.

We must also account for the case when one or both of the constituents are taken from the CKB, and syntactic information is therefore unavailable. We operate under the assumption that context should only be used in the creation of the semantic representation if the language of the input sentence is ambiguous. For example, if the input sentence is "how am I doing?", a CKB item that indicates that the user is currently performing a certain exercise could be used to disambiguate the referent of the question. However, if the sentence instead is "how am I doing on exercise 3?", we do not need to consult the CKB to disambiguate the input, and the response should be the same regardless of which activity the user is currently conducting. Therefore, the composition scores for CKB items are smaller than the corresponding composition scores for constituents taken from the input sentence. Once again, there are also a set of weights $w_{c,d,s,t}$ that are trained using labeled data.

Given the scoring functions Φ_1 and Φ_2, as well as the parameter weights $\mathbf{w}$, we use an integer linear program to find the best possible MRL given and input and the current parameterization. The objective function and constraints of the integer linear program are given in Table 3. In this formulation, we have a binary variable α_{cs} for each constituent c and ontology item s. If this variable is set to 1 by the ILP, this indicates that this constituent maps to this ontology item, and the ontology item s is present in the final MRL. The binary variable $\beta_{cs,dt}$ indicates that ontology items s and t are composed as $s(t)$. The variable $\beta_{cs,dt}$ can only be set to 1 if $\alpha_{cs} = 1$ and $\alpha_{dt} = 1$. All α and β variables will have values assigned by the integer linear program. More detail about this framework can be found in (Clarke et al., 2010).

3.2. Active Learning

To train the semantic parser, we utilize a decision theoretic approach that allows the parser to select from two different types of questions. In the first type of ques-

Maximize: $f(X) = \sum_{c \in X} \sum_{s \in O} \alpha_{cs} \cdot \mathbf{w^T} \Phi_1(X, c, s) + \sum_{c,d \in X} \sum_{s,t \in O} \beta_{cs,dt} \cdot \mathbf{w^T} \Phi_2(X, c, d, s, t)$ **Objective Function**
Subject to: $\forall (c, s) \;\; \alpha_{c,s} \in \{0, 1\}$ $\forall (c, d, s, t) \;\; \beta cs, dt \in \{0, 1\}$ (All variables are binary) **Constraint I** $\forall c \sum_{s \in D} \alpha_{c,s} = 1$ (Every constituent mapped to exactly one item in Ontology) **Constraint II** $\forall (c, d, s, t) \;\; \frac{\alpha_{c,s}}{2} + \frac{\alpha_{d,t}}{2} \geq \beta_{cs,dt}$ (An ontological combination can only be active if both its constituents are active) **Constraint III**

Table 3. Integer Linear Program Formulation

tion, the annotator is given a sentence, an instance of a CKB, and an MRL, and the annotator must determine if the given MRL is correct for the given input. The annotator then provides a binary label indicating if the semantic representation is correct. In the second type of question, the annotator is given only the sentence and the CKB, and is asked to provide the correct MRL. Clearly the second type of the question will be more time consuming for the annotator, so these questions have an elevated cost associated with them in the active learning algorithm.

There will be one active learning metric for each of the two question types. These metrics are computed for each instance in a large unlabeled corpus.

We will use the value of the objective function defined in Table 3 as a proxy for the confidence of the semantic parser for a given sentence, CKB pair. We will denote this confidence as $f(X)$. It is worth noting here that when asking for binary feedback from an annotator, correct parses are much more informative because the correct MRL is implicitly included with the label. However, when asking for a full annotation, inputs for which the parser is currently highly uncertain are most informative.

$$density(X) = \sum_{c \in X} \sum_{Y_i} \frac{1}{|Y_i|} \sum_{d \in Y_i} D(c, d)$$

Density

In addition to confidence, we will also consider how representative a given input is when compared to the whole unlabeled corpus. For given input X, we compare the similarity of X to all other inputs in the corpus, Y. We use normalized lexical similarity between words to compute the similarity of these sentences. The lexical similarity of words c and d is denoted $D(c, d)$. We assume that X and Y_i have already been syntactically parsed, the constituents have been chunked, and determiners and conjunctions have been removed.

$$WD_1(X) = f(X) \cdot density(X)$$
$$WD_2(X) = \delta \cdot \frac{1}{f(X)} \cdot density(X)$$

Active Learning Metrics

WD_1 and WD_2 represent the two active metrics—WD_1 is used to score a binary annotation, while WD_2 is used to score a complete annotation. The parameter δ denotes the cost of asking for a complete annotation. This cost has been estimated to be around 20, based on the time required to produce a whole MRL compared to simply validated a given MRL.

3.3. Personalization

Now we turn to the problem of personalizing the behavior of the system given a contextual knowledge base instance. The task here is to select an action a from a finite set of actions A to minimize a cost function $C(a, CKB)$. The state-space of the contextual knowledge base is very large and sparse. Furthermore, the cost of taking an action with a given CKB may differ between users. For the physical therapy coach, the cost of taking an action will be a function of the user's response to the action. If the action taken by the coach is to sternly tell a user to do one more repetition of an exercise, the cost of the action will be 0 if the user does the exercise and does it correctly, while the cost will be much greater if the exercise is performed incorrectly or not at all. The set of actions includes giving feedback and instruction to a user, allowing the system to issue a given instruction in different ways, *e.g.* aggressively or supportively.

To personalize the coaching system, we train a conditional random field (CRF) with features extracted from the contextual knowledge base. Each feature will be a logical statement consisting of elements of the CKB joined with conjunctions, disjunctions, and negations, for example $[user\_gender(male) \wedge current\_activity(exercise4)]$. A feature will have a value of 1 for given knowledge base if and only if the assertion encoded by the feature evaluates to true in the current knowledge base. We can think of the space of all such logical assertions formed using CKB entries as a lattice, wherein statements containing only a single CKB statement lie at the bottom level of the lattice. During training, all features in the lowest k levels of the lattice (all first-order logical statements with at most k terms) are considered, and L1 regularization is used to constrain the feature space.

The system will initially use a baseline model trained on many users to select actions, and this model will slowly be replaced by a model trained post-deployment based on a given user's interactions with the system. To capture this, we will have two weight vectors: $\mathbf{W_B}$ contains the baseline weights, and $\mathbf{W_P}$ contains the personalized weights. Only those elements of the feature lattice with non-zero $\mathbf{W_B}$ weights are used when training $\mathbf{W_P}$. The personalized weight vector is further scaled by a parameter γ, which increases as more data becomes available for the current user, thus allowing the personalized model to slowly replace the baseline model.

4. Results

In this section, we present preliminary experimental results when using the semantic parser and personalization algorithms.

For the semantic parser, we collected 98 natural language sentences, with each sentence accompanied by its own CKB. The CKB inputs for these tests were artificially created for the purposes of testing and training the semantic parser. Each of the 98 inputs also has a manually created target MRL statement accompanying it.

Input sentences included statements about pain levels for an exercise, user requests to start or stop an exercise, responses to queries for personal information (like age, gender, and weight), and requests for feedback from the coaching system for a given exercise. The CKB inputs included statements such as current and previous user activity, MRL statement of last utterance by the system and user, reported pain levels for various exercises, personal information, and current date and time.

The weights for the semantic parser were trained using a structured SVM algorithm. For evaluation, we used 50 rounds of cross-validation, with 75 randomly selected questions held out for testing in each round. Figure 1(a) shows the training curve with and without active learning enabled, with the X axis indicating the number of questions the system was allowed to ask, and the Y axis indicates the average parsing accuracy. With active learning enabled, the metrics described in section 3.2 were used to determine which input in the training data the system would request, and whether it would request a full MRL or binary feedback on it's current prediction. Without active learning enabled, the system was randomly given labels from the training data, with a 20% chance of being given a full MRL, and 80% of being given binary feedback.

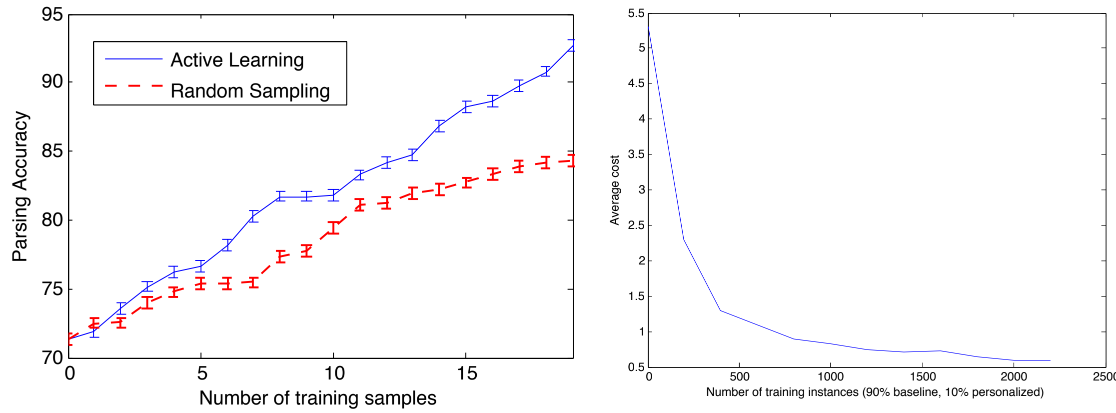
For the personalization CRF, we simulated a dataset with each row consisting of an action, a CKB, and a cost. Ten different distributions of $C(a, CKB)$ were defined, with each distribution representing a single, simulated user of the system. Random samples were taken from each distribution to construct the dataset. The evaluation of the algorithm consisted of 10 stages. In each stage, the algorithm was given all of the data from 9 of the distributions with which to train $\mathbf{W_B}$. $\mathbf{W_P}$ was trained on 10% of the samples from the remaining distribution, and then asked to predict the cost of the remaining 90% of the samples from this distribution. Figure 1(b) shows the average cost of the actions chosen by the system as a function of the number of training samples provided.

4.1. Analysis

We begin by considering the semantic parsing results. Many of the sentences in the testing data would have been impossible to correctly parse without the corresponding CKB entries, and we see from Figure 1(a) that the active learning algorithm generally improved the parsing accuracy compared to random sampling. We see that when $n \geq 4$, the increase in accuracy when using active learning is always statistically significant.

We observed consistent trends in the learned values of the parsing weights across many iterations in the cross validation. In particular, the weights for $current\_activity()$ and $last\_question()$ expressions in the CKB directly affected the target MRL statement in many training examples, and thus, these CKB entries received rather large weights. CKB entries related to personal information, such as Likert pain levels and user age, nearly always ended training with weights close to 0. This is unsurprising, because within the scope of these experiments, these features do not directly inform the content of the target MRLs, but many of these CKB entries appeared in many features selected by the personalization algorithm.

With the personalization experiments, we used a set of 60 possible CKB entries. When searching the feature lattice, we only considered logical statements with $k = 10$ or fewer terms. This still resulted in more than 1 million possible features, but the vast majority of these features never appeared in any instance in the dataset. On average, between 100 and 300 features received non-zero weights in $\mathbf{W_B}$ when L1 regularization



(a) Semantic Parsing

(b) Personalization

Figure 1. Results

was applied.

Randomly picking an action would result in an expected cost $C(a, CKB) = 5$. We see from figure 1(b) that the knee of the cost curve occurs at around 400 data points, with 40 of those being taken from the user with which the system is interacting. We could realistically expect to see this many interactions with a user of a physical therapy coaching system over the course of a few weeks, so we feel that these results indicate that creating personalized interactions will be viable in this domain.

5. Future Work

The results we have seen within this paper regarding the building and usage of a Contextual Knowledge Base are promising, but there is a great deal of system integration remaining to be done before a complete system can be deployed for evaluation.

While the results we have seen using active learning in this work are promising, human annotation continues to be a bottleneck in the training of a context aware semantic parser. Future work may explore new methods of acquiring feedback information from human annotators, such as allowing human annotators presented with a candidate MRL statement to pinpoint to a learning algorithm exactly which part of the MRL requires adjustment. Alternatively, the weaker forms of training that do not require an expert annotator (such as giving binary feedback on a input/output pair) could be crowdsourced.

Going forward, we feel that Contextual Knowledge Bases represent an exciting framework for incorporating language with context awareness and personalization, but there are many open research questions that need to be addressed before this technology can prove its full potential.

References

Bollacker, Kurt, Evans, Colin, Paritosh, Praveen, Sturge, Tim, and Taylor, Jamie. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pp. 1247–1250. ACM, 2008.

Carlson, Andrew, Betteridge, Justin, Kisiel, Bryan, Settles, Burr, Hruschka Jr, Estevam R, and Mitchell, Tom M. Toward an architecture for never-ending language learning. In *Proceedings of the Twenty-Fourth Conference on Artificial Intelligence (AAAI 2010)*, volume 2, pp. 3–3, 2010.

Clarke, James, Goldwasser, Dan, Chang, Ming-Wei, and Roth, Dan. Driving semantic parsing from the world's response. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pp. 18–27. Association for Computational Linguistics, 2010.

Fasola, Juan and Mataric, Maja J. Using socially assistive human–robot interaction to motivate physical exercise for older adults. *Proceedings of the IEEE*, 100(8):2512–2526, 2012.

Fisher, Robert and Simmons, Reid. Smartphone interruptibility using density-weighted uncertainty sampling with reinforcement learning. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, volume 1, pp. 436–441. IEEE, 2011.

Kidd, Cory D and Breazeal, Cynthia. Robots at home:

Understanding long-term human-robot interaction. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pp. 3230–3235. IEEE, 2008.

Krishnamurthy, Jayant and Mitchell, Tom M. Weakly supervised training of semantic parsers. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 754–765. Association for Computational Linguistics, 2012.

Lane, Nicholas D, Miluzzo, Emiliano, Lu, Hong, Peebles, Daniel, Choudhury, Tanzeem, and Campbell, Andrew T. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

Lee, Min Kyung, Forlizzi, Jodi, Kiesler, Sara, Rybski, Paul, Antanitis, John, and Savetsila, Sarun. Personalization in hri: A longitudinal field experiment. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pp. 319–326. IEEE, 2012.

Morbini, Fabrizio and Sagae, Kenji. Joint identification and segmentation of domain-specific dialogue acts for conversational dialogue systems. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pp. 95–100. Association for Computational Linguistics, 2011.

Settles, Burr. Active learning literature survey. *University of Wisconsin, Madison*, 2010.