

Cite as: G. Sukthankar, R. Goldman, C. Geib, D. Pynadath, and H. Bui, "An Introduction to Plan, Activity, and Intent Recognition", In G. Sukthankar, R. Goldman, C. Geib, D. Pynadath, and H. Bui, editors, *Plan, Activity, and Intent Recognition*, Elsevier, 2014

An Introduction to Plan, Activity, and Intent Recognition

Gita Sukthankar, Robert P. Goldman, Christopher Geib,
David V. Pynadath, and Hung Hai Bui

Overview

The ability to recognize the plans and goals of other agents enables humans to reason about what other people are doing, why they are doing it, and what they will do next. This fundamental cognitive capability is also critical to interpersonal interactions because human communications presuppose an ability to understand the motivations of the participants and subjects of the discussion. As the complexity of human-machine interactions increases and automated systems become more intelligent, we strive to provide them with comparable intent recognition capabilities.

Research addressing this problem is variously referred to as plan recognition, activity recognition, goal recognition, and intent recognition. This synergistic research area combines techniques from user modeling, computer vision, natural language understanding, probabilistic reasoning, and machine learning. Plan recognition algorithms play a crucial role in a wide variety of applications including smart homes, intelligent user interfaces, personal agent assistants, human-robot interaction, and video surveillance.

Plan recognition research in computer science dates back at least thirty-five years, initially defined in a paper by Schmidt, Sridharan, and Goodson [64]. In the last ten years, significant new advances have been made on this problem by researchers in artificial intelligence and related areas. These advances have been driven by three primary factors: first, the pressing need for sophisticated and efficient plan recognition systems in a wide variety of applications; second, the development of new algorithmic techniques in probabilistic modeling, machine learning, and optimization (combined with more powerful computers to use these techniques); and finally, our increased ability to gather data about human activities.

Recent research in the field is often divided into two subareas. "Activity recognition" focuses on the problem of dealing directly with noisy low-level data gathered by physical sensors such as cameras, wearable sensors, and instrumented user interfaces. The primary task in this space is to discover and extract interesting patterns in noisy sensory data that can be interpreted as meaningful activities. For example, an activity recognition system processing

a sequence of video frames might start by extracting a series of motions and then will attempt to verify that they are all part of the activity of filling a tea kettle. “Plan and intent recognition” concentrates on identifying high-level complex goals and intents by exploiting relations between primitive action steps which are elements of the plan. Relations that have been investigated include causality, temporal ordering, coordination among multiple sub-plans (possibly involving multiple actors), and social convention.

A Brief History

The earliest work in plan recognition was rule-based [63, 64, 77, e.g.], following the dominant early paradigm in artificial intelligence. Researchers attempted to create inference rules that would capture the nature of plan recognition. Over time, it became clear that without an underlying theory to give them structure and coherence, such rule sets are difficult to maintain and do not scale well.

In 1986, Kautz and Allen published an article, “Generalized Plan Recognition” [35] that has provided the conceptual framework for much of the later work in plan recognition to date. They defined the problem of plan recognition as identifying a minimal set of *top-level actions* sufficient to explain the set of observed actions. Plans were represented in a plan graph, with top-level actions as root nodes and expansions of these actions as unordered sets of child actions representing plan decomposition. To a first approximation, the problem of plan recognition was then a problem of graph covering. Kautz and Allen formalized this view of plan recognition in terms of McCarthy’s circumscription. Kautz [34] presented an approximate implementation of this approach that recast the problem as one of computing vertex covers of the plan graph. These early techniques are not able to take into account differences in the *a priori* likelihood of different goals. Observing an agent going to the airport, this algorithm views “air travel,” and “terrorist attack” as equally likely explanations, because they explain (cover) the observations equally well.

To the best of our knowledge, Charniak was the first to argue that plan recognition was best understood as a specific case of the general problem of *abduction* [11, e.g.]. Abduction, a term originally defined by the philosopher C.S. Peirce, is reasoning to the best explanation, the general pattern being “if A causes B , and we observe B , we may postulate A as the explanation.” In the case of plan recognition, this pattern is specialized to “if an agent pursuing plan/goal P would perform the sequence of actions S , and we observe S , we may postulate that the agent is executing plan P .” Understanding plan recognition as a form of abductive reasoning was important to the development of the field, because it enabled clear computational formulations, and facilitated connections to areas such as diagnosis and probabilistic inference.

One of the earliest explicitly abductive approaches to plan recognition was that of Hobbs, *et al.*[27]. In this work, they defined a method for abduction as a process of cost-limited theorem-proving [65]. They used this cost-based theorem-proving to find “proofs” for the elements of a narrative, where the assumptions

underlying these proofs constitute the interpretation of the narrative—in much the same way a medical diagnosis system would “prove” the set of symptoms, in the process identifying the underlying disease. Later developments would show that this kind of theorem-proving is equivalent to a form of probabilistic reasoning [12].

Charniak and Goldman [9] argued that, if plan recognition is a problem of abduction, it can best be done as Bayesian (probabilistic) inference. Bayesian inference supports the preference for minimal explanations, in the case of equally likely hypotheses, but also correctly handles explanations of the same complexity but different likelihoods. For example, if a set of observations could be equally well explained by three hypotheses, going to the store to shop *and* to shoplift, being one, and going to the store only to shop or going to the store only to shoplift being the others, simple probability theory (with some minor assumptions), will tell us that the simpler hypotheses are more likely. On the other hand, if as above, the two hypotheses were “air travel” and “terrorist attack,” and each explained the observations equally well, then the prior probabilities will dominate, and air travel will be seen to be the most likely explanation. As one example of the unifying force of the abductive paradigm, Charniak and Shimony showed that Hobbs and Stickel’s cost-based abductive approach could be given a probabilistic semantics [12], and viewed as search for the most likely *a posteriori* explanation for the observed actions. While the Bayesian approach to plan recognition was initially quite controversial, probabilistic inference, in one form or another, has since become the mainstream approach to plan recognition.

Another broad area of attack on the problem of plan recognition has been to reformulate it as a parsing problem [74, e.g.], based on the observation that reasoning from actions to plans taken from a plan hierarchy was analogous to reasoning from sentences to parse trees taken from a grammar. Early work on parsing-based approaches to plan recognition promised greater efficiency than other approaches, but at the cost of making strong assumptions about the ordering of plan steps. The major weakness of early work using parsing as a model of plan recognition is that it did not treat partially-ordered plans or interleaved plans well. Recent approaches that use statistical parsing [55, 19, 20, e.g.] combine parsing and Bayesian approaches and are beginning to address the problems of partially-ordered and interleaved plans.

Finally, there has been a large amount of work done using extensions of Hidden Markov Models (HMMs) [6], techniques that came to prominence in signal processing applications, including speech recognition. These approaches offer many of the efficiency advantages of parsing approaches, but with the additional advantages of incorporating likelihood information and of supporting machine learning to automatically acquire their plan models. Standard HMMs are nevertheless not expressive enough to sufficiently capture goal-directed behavior. As a result, a number of researchers have extended them to hierarchical formulations which can capture more complicated hierarchical plans and intentions [6, 5].

Much of this latter work has been done under the rubric of *activity recognition* [15]. The early research in this area very carefully chose the term *activity*

or *behavior recognition* to distinguish it from plan recognition. The distinction to be made between activity recognition and plan recognition is the difference between recognizing a single (possibly complex) activity and recognizing the relationships between a set of such activities that result in a complete plan.

Activity recognition algorithms discretize a sequence of possibly noisy and intermittent low-level sensor readings into coherent actions that could be taken as input by a plan recognition system. The steady decline in sensor costs has made instrumenting smart spaces practical and brought activity recognition to the forefront of research in the computer vision and pervasive computing communities. In activity recognition, researchers have to work directly with sensor data extracted from video, accelerometers, motion capture data, RFID sensors, smart badges, and Bluetooth. Bridging the gap between the noisy, low-level data and high-level activity models is a core challenge of research in this area. As data becomes more readily available, the role of machine learning and data mining to filter out noise and abstract away from the low-level signals rises in importance. As in other machine learning tasks, activity recognition can be viewed as a supervised [57] or an unsupervised [78] learning task, depending on the availability of labeled activity traces. Alternatively, it can be treated as a problem of hidden state estimation and tackled with techniques such as hierarchical hidden (semi)-Markov models [47, 15], dynamic Bayesian networks [39], and conditional random fields [79, 73, 40].

A specialized subfield of “action recognition” is dedicated to the problem of robustly recognizing short spatio-temporally localized actions or events in video with cluttered backgrounds (see [53] for a survey); generally activity recognition carries the connotation that the activity recognized is a more complex sequence of behavior. For example, “throwing a punch” is an example of an action that could be recognized by analyzing the pixels within a small area of an image and a short duration of time. In contrast, “having a fight” is a complex multi-person activity that could only be recognized by analyzing a large set of spatio-temporal volumes over a longer duration.

Several researchers have been interested in extending plan recognition to multiagent settings [62] and using plan recognition to improve team coordination [29, 33]. If agents in a team can recognize what their teammates are doing, then they can better cooperate and coordinate. They may also be able to learn something about their shared environment. For example, a member of a military squad who sees a teammate ducking for cover may infer that there is a threat and hence take precautions.

In domains with explicit teamwork such as military operations or sports, it can be assumed that all of the agents have a joint persistent commitment to executing a goal, share a utility function, and have access to a common plan library grounded in shared training experiences. This facilitates the recognition process such that in the easiest case it is possible to assume that all of the actions are being driven by one centralized system with multiple “actuators.” For simpler formulations of the multi-agent plan recognition problem, multi-agent plan recognition can be performed in polynomial time [4]. In the more complex case of dynamic teams, team membership changes over time and accurate plan

recognition requires identifying groupings among agents, in addition to classifying behaviors [67]. Grouping agents in the unconstrained case becomes a set partition problem, and the number of potential allocations rises rapidly, even for a small number of agents. Prior work on multi-agent plan recognition has looked at both extending single agent formalisms for the multi-agent recognition process [62, 41, 60] and creating specialized models and recognition techniques for agent teams [66, 3].

Thus, we see how far the field has evolved, from the genesis of plan recognition as a subproblem within classical artificial intelligence to a vibrant field of research that stands on its own. Figure 1 illustrates the diversity of concepts, methods, and applications that now drive advances across plan, activity, and intent recognition. This book provides a comprehensive introduction to these fields by offering representative examples across this diversity.

Chapter Map

The collection of chapters in this book is divided into four sections: 1) classical plan and goal recognition approaches, 2) activity discovery from sensory data, 3) modeling human cognitive processes, 4) multi-agent systems, and 5) applications of plan, activity, and intent recognition. We discuss each of these areas and the chapters we have grouped under the headings below.

Classical Plan and Goal Recognition

The book begins with chapters that address modern plan recognition problems through the same abductive perspective that characterized the seminal work in the field. The first chapter in this section addresses two important challenges in modern plan recognition. The first is the question of how much recognition is actually needed in order to perform useful inference? Can we perform a more limited, but still useful, inference problem more efficiently? Blaylock and Allen, in their chapter, “Hierarchical Goal Recognition,” argue that in many cases we can, and propose to solve the simpler problem of *goal recognition*. They also address a second challenge: evaluating plan recognition techniques, proposing to use synthetic corpora of plans to avoid the problems of acquiring human goal-directed action sequences annotated with “ground truth” motivation.

Blaylock and Allen’s chapter provides a definition of goal recognition as a proper subset of plan recognition. In *goal recognition* all we are interested in is the top level goal of the agent while in *plan recognition* we also ask the system to produce a hypothesis about the plan being followed by the agent, and answer questions about the state of plan execution such as “how much of the plan is completed?” and “what roles do particular actions play in the plan?” Blaylock and Allen present an approach to goal recognition based on Cascading Hidden Markov Models.

As plan recognition is maturing, it is moving away from exploratory engineering of proof-of-concept plan recognition algorithms. However, it is difficult

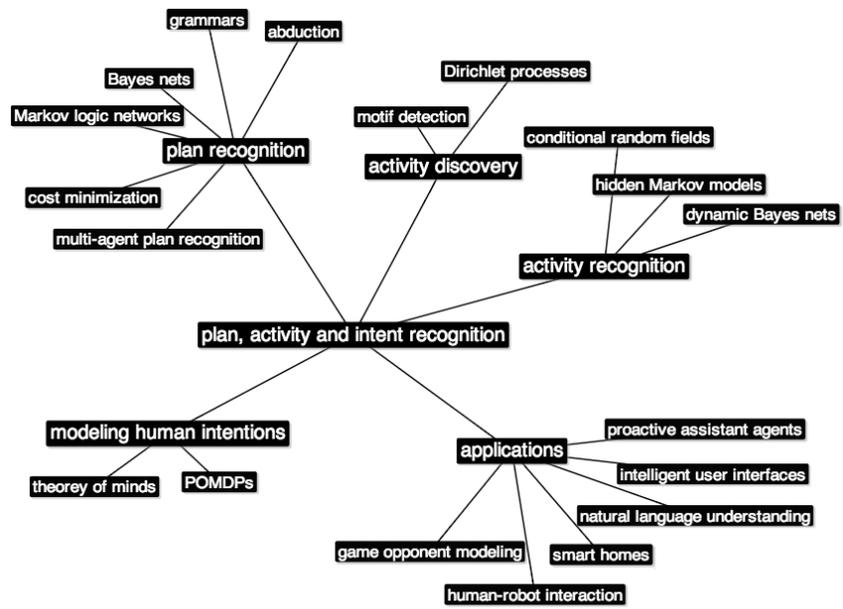


Figure 1: A mind map of research directions, methods, and applications in plan, activity, and intent recognition.

to do “apples to apples” comparisons of different techniques without shared datasets. The Monroe Corpus of plans and observation traces created by Blaylock for his Ph.D. dissertation was one of the first publicly available corpora for training and testing plan recognition systems. It has been a significant resource for the plan recognition community as it attempts to move from an exploratory to a more empirical foundation. This chapter introduces the Monroe Corpus, describes the synthetic generation approach for creating the corpus, and then uses it to evaluate the accuracy and performance of Blaylock and Allen’s goal recognition system.

The next chapter “Weighted Abduction for Discourse Processing Based on Integer Linear Programming,” by Inoue, *et al.* represents two important threads in the history of plan recognition: the use of plan recognition in service of language understanding and the theoretical development of plan recognition in terms of abduction. Some of the earliest work in plan recognition was done in the service of understanding natural language, both in comprehending the motivations and actions of characters in stories [63, 10, 11] and in order to identify the interests of participants in discourse [13, 52, 77]. Work by both Charniak’s group at Brown and Hobbs’s group (originally at SRI) went further, and integrated language processing and deeper interpretation in ways that fed backward and forward, such that information about plans could be used to resolve semantic ambiguity in text interpretation. Inoue, *et al.* describe an application to discourse processing, evaluating their work by measuring accuracy in recognizing textual entailment (RTE). RTE is the problem of determining whether particular hypotheses are entailed by the combination of explicit and implicit content of text. In RTE identifying the implicit content of text requires combining explicit content with commonsense background knowledge, including plan recognition.

Inoue, *et al.* further develop Hobbs and Stickel’s cost-based approach to abduction. They review the concepts of weighted abduction, and describe an enhancement of these methods that uses integer linear programming (ILP) as a method for the cost-based reasoning. They show that this method can speed up the interpretation process by allowing them to exploit both highly optimized ILP solvers, and machine learning methods for automatically tuning the cost parameters. They experimentally compare the technique with other methods for plan recognition and show that their wholly-automated approach has a higher accuracy than manually-tuned plan recognition methods.

The next chapter “Plan Recognition using Statistical Relational Models” by Raghavan, *et al.* is also heavily influenced by an abductive view of the problem of plan recognition. Here, abductive reasoning is formulated in the framework of *statistical relational learning* (SRL) [22]. This framework unifies logical and probabilistic representation and provides expressive relational models that support efficient probabilistic reasoning and statistical parameter estimation from data.

Structured models have long been a challenge for plan recognition techniques, especially those using probabilistic methods. Traditionally probabilistic models have had very simple structures. Handling more complex structures, including

nesting (for subplans), inheritance, and coreference constraints (when shopping, the thing purchased is typically the same as the thing taken from the shelf), was a primary challenge to the development of the first Bayesian methods for plan recognition [9, 24]. This early work combined logical and probabilistic *inference* techniques, but had no means to perform efficient approximate inference, or to *learn* the required parameters of the models.

In their chapter, Raghavan, *et al.* apply *Markov Logic Networks* (MLNs) [59] and *Bayesian Logic Programs* (BLPs) [36] to the problem of plan recognition. To do so, they extend both of these modeling frameworks. MLNs are a very general modeling framework. For MLNs, they provide a number of alternate encodings of abductive reasoning problems. BLPs are theoretically less general, but can exploit directionality in the underlying probabilistic graphical model to encode causal relationships. Raghavan, *et al.* develop an extension of BLPs called Bayesian *Abductive* Logic Programs (BALPs). They compare the performance of these techniques on plan recognition benchmarks, showing that BALPs combine efficient inference with good quality results, outperforming the more general MLNs.

This section of the book then pivots to consider the particularly challenging problem of adversarial plan recognition. In the case of adversarial agents, we cannot expect the observed agents to obligingly provide us with their plan libraries, and they may attempt to evade our observational apparatus, or mislead our plan recognition through stealth or feints. In the chapter “Keyhole Adversarial Plan Recognition for Recognition of Suspicious and Anomalous Behavior”, Avrahami and Kaminka describe a hybrid plan recognition system that employs both standard plan recognition and *anomaly detection* to improve recognition in adversarial scenarios. The anomaly detection subsystem complements recognition of known suspicious behavior by detecting behaviors that are not known to be benign. This chapter also investigates the use of utility reasoning in conjunction with likelihood reasoning in plan recognition. Instead of simply identifying the most likely plan for a set of actions, their system also identifies hypotheses that might be less likely but have a larger impact on the system’s utility function; in this context these are more threatening hypotheses.

Activity Discovery/Recognition

An important precursor to the task of activity recognition is the discovery phase—identifying and modeling important and frequently repeated event patterns [43]. Two chapters in the book focus on this emerging research area. Rashidi’s chapter on “Scaling Activity Recognition,” and “Learning Latent Activities from Social Signals with Hierarchical Dirichlet Processes” by Phung, *et al.* Rashidi’s chapter, “Scaling Activity Recognition”, discusses the problem of analyzing activity sequences in smart homes. Smart homes are dwellings equipped with an array of sensors and actuators that monitor and adjust the home control system settings to improve the safety and comfort of the inhabitants. Key advances in this area have been driven by several research groups who have made activities of daily living (ADL) datasets publicly available [48, 71, 70].

Rashidi’s work was conducted using data from the CASAS testbed at Washington State [56]; examples of other smart environment projects include Georgia Tech’s Aware Home [1] and MIT’s House_n [68].

Smart environments pose a challenging data analysis problem because they output non-stationary streams of data; new elements are continuously generated and patterns can change over time. Many activity discovery approaches (e.g., [43, 72]), use time-series *motif detection*, the unsupervised identification of frequently repeated subsequences, as an element in the discovery process. The term “motif” originated from the bioinformatics community where it is used to describe recurring patterns in DNA and proteins. Even though these techniques are unsupervised, they make the implicit assumption that it possible to characterize the user’s activity with one dataset sampled from a fixed period of time. Problems arise when the action distribution describing the user’s past activity differs from the distribution used to generate future activity, due to changes in the user’s habits. Hence it can be beneficial to continue updating the library of activity models, both to add emerging patterns and discard obsolete ones.

Rashidi proposes that activity discovery can be modeled as a *data stream processing* problem in which patterns are constantly added, modified, and deleted as new data arrives. Patterns are difficult to discover when they are discontinuous, due to interruptions by other events, and also when they appear in varied order. Rashidi’s approach, STREAMCom, combines a tilted time window data representation with pruning strategies to discover discontinuous patterns that occur across multiple time scales and sensors. In a fixed time window, older data is forgotten once it falls outside the window of interest, but with the tilted time representation the older data is retained at a coarser level. During the pruning phase, infrequent or highly discontinuous patterns are periodically discarded based on a compression objective that accounts for the pattern’s ability to compress the dataset. The chapter presents an evaluation of STREAMCom’s performance on discovering patterns from several months of data generated by sensors within two smart apartments.

The second chapter in this section, “Learning Latent Activities from Social Signals with Hierarchical Dirichlet Processes” by Phung, *et al.* describes a method for analyzing data generated from personal devices, such as mobile phones [37], sociometric badges [49], and wearable RFID readers [18]. Wearable RFID readers, such as Intel’s iBracelet and iGlove, are well suited for reliably detecting the user’s interactions with objects in the environment, which can be strongly predictive of many activities of daily living [51]. Sociometric badges are wearable electronic devices designed to measure body movement and physical proximity to nearby badge wearers. These badges can be used to collect data on interpersonal interactions and study community dynamics in the workplace. Two datasets of particular importance, the Reality Mining [16] and Badge [49] datasets, were released by the MIT Human Dynamics lab to facilitate the study of social signal processing [50].

Phung *et al.* describe how a Bayesian nonparametric method, the hierarchical Dirichlet process [69], can be used to infer latent activities such as driving, playing games, and working on the computer. The strength of this type of ap-

proach is twofold: (1) the set of activity patterns (including its cardinality) can be inferred directly from the data, and (2) statistical signals from personal data generated by different individuals can be combined for more robust estimation via a principled hierarchical Bayesian framework. The authors also show how their method can be used to extract social patterns such as community membership from the Bluetooth data capturing colocation of users in the Reality Mining dataset. The activity discovery techniques described in these two chapters will be of interest to readers working with large quantities of data who are seeking to model unconstrained human activities using both personal and environmental sensors.

Modeling Human Cognition

Much of this book presents computational mechanisms that try to recognize a human being’s plans, activities, or intentions. This section, in contrast, examines the human brain’s own mechanisms for performing such recognition in everyday social interaction. These mechanisms include a *Theory of Mind* [75, 76] that allows people to attribute others with the same kind of mental states and processes that they possess themselves.

Empirical studies have shown that people typically ascribe goals and beliefs to the observed behavior of others using a causal model informed by their own decision-making [26]. This causal model often includes the observed agent’s own theory of mind, leading to recursive levels of recognition [7]. Researchers have sought to build computational models that capture this naturally occurring Theory of Mind by combining models of rational decision-making, with reasoning from observed behavior to underlying beliefs and utilities. Such quantitative representations of uncertainty and preferences have provided a rich language for capturing human decision-making, and the chapters in this section are emblematic of a growing number of human-inspired approaches to plan recognition [54, 58].

This section’s first chapter, “Modeling Human Plan Recognition using Bayesian Theory of Mind”, presents a framework for Theory of Mind that, like many computational approaches to plan recognition, starts with a generative model of decision making and then uses that model for abductive reasoning. Baker and Tenenbaum frame a person’s decision as a *partially observable Markov decision problem* (POMDP), representing uncertain beliefs as a probability distribution and preferences as a reward function. The POMDP also captures the effects of the person’s action choices, supporting domain-independent algorithms that compute a value function over those action choices. These algorithms operate on the assumption that the choices that generate the highest expected reward will have the highest value to the decision maker. By inverting this value function, an observer can perform Bayesian inference to reconstruct the observed person’s belief state and reward function, conditional on the observed behavior. The chapter presents empirical evidence showing that this Bayesian Theory of Mind is an accurate predictor of human judgments when performing plan recognition in experimental settings.

This section’s second chapter, “Decision-Theoretic Planning in Multiagent Settings with Application to Behavioral Modeling”, similarly uses POMDPs as a basis for abductive reasoning about human behavior. However, just as human Theory of Mind operates within the context of social interaction, Doshi, *et al.* place POMDP models of others within the context of the observing agent’s own decision-making. In particular, their *interactive* POMDPs (I-POMDPs) use nested POMDPs to model an observing agent’s decision-making while also ascribing Theory of Mind in a recursive fashion to the observed agent. Thus, the I-POMDP framework supports plan recognition when observing the behavior of people, who may also be performing plan recognition of people, who may also be performing plan recognition, etc. While this recursion may be arbitrarily deep in theory, the chapter also presents a technique by which I-POMDPs of fixed nesting depth can fit data gathered from human behavior when reasoning about others.

Multi-Agent Systems

Plan and activity recognition formalisms generally assume that there is only one person or agent of interest, but in many real-world deployment scenarios, multiple people are simultaneously performing actions in the same area or cooperating to perform a group task. The presence of multiple agents can lead to action interdependencies that need to be accounted for in order to perform accurate recognition.

The first chapter in this section, “Multi-Agent Plan Recognition from Partially Observed Team Traces,” frames the multi-agent plan recognition process as a weighted maximum satisfiability (MAX-SAT) problem, rather than treating it as abduction or inference, as was presented in the early chapters. In a weighted MAX-SAT problem, the aim is to determine the maximum number of clauses in a Boolean formula that can be satisfied by a variable assignment. Zhuo outlines two representation options: 1) team plans expressed as a set of matrices or 2) a set of action models and goals in the STRIPS planning language. Assuming the existence of a plan library, Zhuo’s MARS (Multi Agent Recognition System) finds candidate occurrences of team plans in the observed trace and generates constraints based on this candidate set that are used by the solver. In the case where no plan library exists, Zhuo’s alternate framework DARE (Domain-based multi Agent Recognition) identifies plans constructed using the pre-defined action models that explain all observed activities and have the highest likelihood of achieving the goal, as measured by a combination of coordination costs and plan length. Both frameworks are reasonably robust to increases in the number of agents and the number of missing observations.

The second chapter in this section, “Role-based Ad Hoc Teamwork,” moves from plan recognition in STRIPS domains to examining movement-oriented team tasks such as foraging and capture the flag. Motivated by pickup soccer games, Genter *et al.*’s objective is to develop agents that are capable of participating in *ad hoc* teams. In order to be an effective participant, these agents adaptively decide on future actions after assessing their teammates’ cur-

rent play. In Genter *et al.*'s approach, team activities are expressed as sets of roles filled by the different players. Assuming that it is possible to accurately recognize the roles of the other players, the agent joining the *ad hoc* team performs marginal utility calculations to select the best role to fill gaps in the current team's strategy. Analyzing multi-agent activities is an area of ongoing research, and the two chapters in these section show the breadth of work in this area.

Applications

This section of the book presents work on the practical application of plan and activity recognition techniques. The core plan recognition algorithms are both versatile and broadly applicable to any application that involves human interaction. However specialized customization, or “secret sauce”, is often required to make systems with different types of input data—video [28], natural language [8], or user interface events [38]—perform well, and to adapt general-purpose heuristics to specific situations. These chapters discuss how the recognition process should interface with other system components, rather than focusing on algorithmic improvements to activity and plan recognition.

The first chapter in this section, “Probabilistic Plan Recognition for Proactive Assistant Agents,” by Oh *et al.*, illustrates one of the most common applications of plan and activity recognition techniques: automated systems that assist human users. To be able to choose the best assistance to provide, such systems must be able to infer the users' current tasks and goals, as well as anticipate their future needs. Oh *et al.* pay special attention to the need to be *proactive* in providing assistance when the users are under heavy cognitive load, as in emergency response domains. They apply probabilistic plan recognition algorithms that use a generative *Markov decision problem* (MDP) model of the domain as the basis for the agent's inference of the users' goals. The agent can then use that inference to generate predictions of the users' chosen course of action and to inform its own planning process in assisting that course of action. The chapter illustrates the successful application of this general approach within the specific domains of military peacekeeping operations and emergency response.

Another application area of particular interest is the use of plan/activity recognition as a tool for modeling players in computer games and virtual worlds. Player modeling differs from other types of user modeling problems because much of the user experience is driven by the players' interpretation of virtual world events, rather than being limited to their interactions with menus, mouse, and keyboard. The human user simultaneously occupies multiple roles: software customer, inhabitant of the virtual world, and, in serious games, student seeking to perfect skills. Yet people's activities in virtual worlds are more structured than their real-world behavior due to the limited vocabulary of actions and the presence of artificial winning conditions. Also, data collection is easier in virtual environments due to the lack of sensor noise. Thus human behavior recognition in computer games offers more complexity than other user modeling problems

with fewer deployment issues than analyzing data from smart environments.

A popular game format is to provide players with quests that can be completed for character advancement; this style of game supports a non-linear narrative structure, offers limited freedom to the players to select quest options, and easy extensibility for the game designers. Researchers modeling player behavior in games can assume that all of the players' actions are performed in the service of completing quests and formalize the problem as one of goal recognition. Albrecht *et al.* implemented the earliest demonstration of online goal recognition for text-based computer adventure games using dynamic Bayesian networks to recognize quest goals and to predict future player actions [2]. Adding additional game context information to the model has been shown to be helpful for identifying transitions between goals. For example, Gold's system [23] employs low-level inputs in conjunction with input-output hidden Markov models.

The chapter "Recognizing Player Goals in Open-Ended Digital Games with Markov Logic Networks" describes research done on one of the most famous testbeds, Crystal Island, a game-based learning environment in which the students solve a science mystery [42]. Crystal Island has been used as a testbed for both pedagogical research and earlier studies on performing goal recognition using Bayes nets and scalable n-gram models [45]. In this chapter, Ha *et al.* describe how Markov logic networks (discussed in Chapter 3 by Raghavan *et al.*) improve upon the previous n-gram model. The authors show that a major advantage of their factored MLN model is that it can leverage associations between successive goals, rather than treating the goals individually.

Good player modeling is an important stepping stone toward the creation of player-adaptive games that automatically adjust gameplay to enhance user enjoyment. For instance, dynamic difficulty adjustment games modify the challenge level of the scenario by changing the number, health, and firing accuracy of the opposing forces [30]. Previous work in this area has concentrated on simple numeric attribute adjustments or scenario modifications [32] rather than changing the action choices of the automated opponents.

The third chapter, "Using Opponent Modeling to Adapt Team Play in American Football," tackles the problem of creating player-adaptive sports games that learn new strategies for countering the player's actions. Football plays are similar to conditional plans and generate consistent spatio-temporal patterns; the authors demonstrate that it is possible to recognize plays at an early execution stage using a set of supervised classifiers. This differs from prior work on camera-based football recognition systems in which the emphasis has been on recognizing completed plays rather than partial ones (e.g., [31]). Play recognition is used in multiple ways: 1) to learn an offline play book designed to be challenging for a specific player and 2) to make online repairs to currently executing plays. With the rapid expansion of game telemetry systems that collect massive amounts of data about players' online experience, it is likely that future game systems will increase their usage of machine learning for player modeling [17].

The last chapter in this section, "Intent Recognition for Human-Robot Interaction," by Kelley *et al.*, addresses human-robot interaction. Many people

dream of the day when a “robot butler” will be able to do all of the boring repetitive tasks that we wish we didn’t have to do. However, creating a robot that can perform the tasks is only half the battle; it is equally important that the user’s interactions with the system be effortless and pleasant. Ultimately, we want household assistants that can anticipate our intentions and plans and act in accordance with them. As discussed in many chapters of this volume, building proactive assistant systems like a robot butler requires plan recognition.

General-purpose autonomous, physically-embodied systems like the robot butler rely on the successful integration of a large number of technologies. The system described in this chapter provides a good example, involving the integration of research in vision, planning, plan recognition, robotics, and natural language processing.

Highly integrated systems like this one provide us many opportunities to test our research systems. First and most obviously they provide us with an opportunity to explore the limits of algorithms when they are taken out of the controlled conditions of the lab. Since plan recognition must share a limited pool of computational resources with other tasks, the real time requirements in such “embodied” systems are often more demanding than in other application domains. For example, given how much time it takes to plan and execute a response, how much time can we spend on plan recognition?

Second, integration into whole real world systems can give us much needed perspective on the challenging parts of our respective research questions when applied to actual problems rather than the theoretical cases. For example, what quality and detail can we expect from action observations that come from actual vision or sensor systems?

Finally, highly integrated applications provide us with opportunities to learn from the solutions of others. It exposes us to approaches that researchers in other sub-areas have employed to address problems that may be similar to ours. For example, can we use knowledge from language to form context structures to help disambiguate plans?

This final chapter illustrates all of these issues, and shows us some of the first steps that plan recognition algorithms are taking to help create applications that will be indispensable in the future.

Future Directions

The immediate future holds many exciting opportunities as well as challenges for the field. The new wave of user-centric and context-aware applications including personal assistants; customized recommendation and content delivery; personalized health and elder care; smart and interactive spaces; and human-robot interaction all share one essential requirement: to accurately capture and track the current user’s activities. The continued growth of such applications ensures that plan and activity recognition will receive increased attention from academia and industry. Thanks to the efforts of many research groups, there has been a democratization of recognition techniques in which more software

developers are creating and deploying systems that use limited forms of plan and intent recognition. Software toolkits such as the Google Activity Recognition API [25] have made common algorithms freely accessible for mobile phone platforms.

Yet important unsolved research questions remain and new challenges abound. Interestingly, prominent application areas of today have conflicting requirements for plan recognition. Big data and cloud computing drive the demand for large scale analysis of data using vast computing power. On the other hand, personal assistants and robotic systems typically contend with resource-constrained mobile platforms. In many cases we wish to handle more complex activities by larger groups of users over extended time periods.

Since the conception of the field, researchers have grappled with a fundamental question: what are the suitable computational models and structures for representing plans, activities and intents that also facilitate efficient and robust learning and recognition algorithms? Early work in the field mostly employed representations based on predicate or first-order logic, which are convenient for representing the kind of structures that often encountered in a plan (pre-conditions, post-conditions, subgoals). As the need to work with sensor data and events becomes more acute, recent work has made heavy use of probabilistic and statistical models. In doing so, researchers trade expressivity for robustness against noise, a necessity in real world applications. As the type of activities and intents we consider increases in complexity, the question about suitable representation structure becomes the focus once more. For more ambitious applications, we require (1) a rich structure for representing high-level plans and intentions, and (2) a mechanism for scoring and ranking the different structures beyond what logical inference can offer. Recent statistical relational learning frameworks that combine first-order logic, probabilistic reasoning and statistical estimation promise to satisfy both of these requirements. New work [46, 61, 44] including the chapter by Raghavan, *et al.* in this book, has already started to explore this important direction, but much more needs to be done.

More efficient and scalable recognition methods will continue to be a key research theme in theoretical and algorithmic development in the field. We expect that the development of new techniques in machine learning, probabilistic inference and optimization will serve as the foundation for advancing plan and activity recognition algorithms. A relatively unexplored direction is parallelization [21], and leveraging cloud computing resources. On a local scale these could make our plan recognition algorithms more responsive, and more globally, they could yield algorithms that can operate at the big data scale.

Automatic creation of models and plan libraries through activity discovery remains a significant problem. The exploding use of personal mobile devices, and location-based social media apps has led to an increase in the amount of data available on human activities that can be leveraged by activity discovery algorithms. Wearable devices and sensors represent another potentially very large source of data of a different modality. It is clear that in many domains, to author a large enough set of plans and activities to describe all of the possibilities is impractical; therefore discovering and estimating personalized models from

this mostly unlabeled and multimodal data is important. It is also desirable for developers to be able to contribute their domain knowledge by authoring plan models, while not inhibiting the model discovery process.

An issue of growing importance will be protecting the individual's need for privacy in environments where sensors are more common and potentially intrusive. Creating opt-in policies allowing users greater freedom to protect their data is the first start in addressing this issue. Proactive assistant agents and intelligent user interfaces have the potential to provide valuable services to users who choose to expose data to the application. For instance, many users opt to let their mobile phone apps utilize their GPS data because of the growing number of context-sensitive apps. Ultimately, we hope to see privacy-aware techniques for learning personalized models while allowing trade-offs between maintaining privacy and statistical estimation efficiency [14].

The combination of exciting research problems driven by high-impact real-world applications and the exponential growth of user data will ensure that the field of plan, activity, and intent recognition continues to be an important and fruitful area for researchers and developers in the future.

References

- [1] G. Abowd and E. Mynatt. Designing for the human experience in smart environments. In D. Cook and S. Das, editors, *Smart Environments: Technology, Protocols, and Applications*, pages 153–174. Wiley, 2004.
- [2] A. Albrecht, I. Zukerman, and A. Nicholson. Bayesian models for keyhole plan recognition in an adventure game. *Journal of User Modeling and User-Adapted Interaction*, 9:5–47, 1998.
- [3] D. Avrahami-Zilberbrand and G. Kaminka. Towards dynamic tracking of multi-agents teams: An initial report. In *Proceedings of Workshop on Plan, Activity, and Intent Recognition (PAIR 2007)*, 2007.
- [4] B. Banerjee, L. Kraemer, and J. Lyle. Multi-agent plan recognition: Formalization and algorithms. In *Proceedings of National Conference on Artificial Intelligence*, pages 1059–1064, 2010.
- [5] H.H. Bui, D.Q. Phung, and S. Venkatesh. Hierarchical hidden Markov models with general state hierarchy. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*, pages 324–329. AAAI Press, 2004.
- [6] Hung H. Bui, Svetha Venkatesh, and Geoff West. Policy recognition in the Abstract Hidden Markov Model. *Journal of Artificial Intelligence Research*, 17:451–499, 2002.
- [7] C. Camerer, T.H. Ho, and J.K. Chong. A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119:861–898, 2004.

- [8] S. Carberry. *Plan Recognition in Natural Language Dialogue*. MIT Press, 1990.
- [9] E. Charniak and R. P. Goldman. A Bayesian model of plan recognition. *Artificial Intelligence*, 64(1):53–79, November 1993.
- [10] Eugene Charniak. A common representation for problem solving and language comprehension information. *Artificial Intelligence*, 16:225–255, 1981.
- [11] Eugene Charniak and Drew McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley, Reading, MA, 1985.
- [12] Eugene Charniak and Solomon Eyal Shimony. Cost-based abduction and MAP explanation. *Artificial Intelligence*, 66(2):345–374, 1994.
- [13] Philip R. Cohen and C. Raymond Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3:177–212, 1979.
- [14] John C. Duchi, Michael I. Jordan, and Martin Wainwright. Privacy aware learning. In *Neural Information Processing Systems (NIPS)*, 2012.
- [15] T.V. Duong, H.H. Bui, D.Q. Phung, and S. Venkatesh. Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 838–845, 2005.
- [16] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Journal of Personal and Ubiquitous Computing*, 10(4):255–268, 2006.
- [17] M. El-Nasr, A. Drachen, and A. Canossa, editors. *Game Analytics: Maximizing the Value of Player Data*. Springer, 2013.
- [18] K. Fishkin, M. Philipose, and A. Rea. Hands-On RFID: Wireless Wearables for Detecting Use of Objects. In *Proceedings of the IEEE International Symposium on Wearable Computers*, pages 38–41, 2005.
- [19] Christopher W. Geib. Delaying commitment in probabilistic plan recognition using combinatory categorial grammars. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1702–1707, 2009.
- [20] Christopher W. Geib and Robert P. Goldman. A probabilistic plan recognition algorithm based on plan tree grammars. *Artificial Intelligence*, 173(11):1101–1132, 2009.
- [21] Christopher W Geib and Christopher E Swetenham. Parallelizing plan recognition. In *Workshop on Plan, Activity and Intent Recognition at the Twenty-Seventh AAAI Conference on Artificial Intelligence (PAIR)*, 2013.
- [22] Lise Getoor and Ben Taskar. *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2007.

- [23] Kevin Gold. Training goal recognition online from low-level inputs in an action-adventure game. In *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, pages 21–26, 2010.
- [24] R. P. Goldman and E. Charniak. Probabilistic text understanding. *Statistics and Computing*, 2:105–114, 1992.
- [25] Google. Activity Recognition API. <http://developer.android.com/reference/com/google/android/gms/location/ActivityRecognitionClient.html>.
- [26] Alison Gopnik and Andrew N. Meltzoff. *Words, thoughts, and theories*. MIT Press, Cambridge, MA, 1997.
- [27] Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. Interpretation as abduction. *Artificial Intelligence*, 63:69–142, 1993.
- [28] A. Hoogs and A. Perera. Video activity recognition in the real world. In *Proceedings of National Conference on Artificial Intelligence*, pages 1551–1554, 2008.
- [29] M. Huber, E. Durfee, and M. Wellman. The automated mapping of plans for plan recognition. In *Proceedings of Uncertainty in Artificial Intelligence (UAI)*, 1994.
- [30] R. Hunicke and V. Chapman. AI for dynamic difficulty adjustment in games. In *Proceedings of the AAAI Workshop on Challenges in Game Artificial Intelligence*, 2004.
- [31] S. Intille and A. Bobick. A framework for recognizing multi-agent action from visual evidence. In *Proceedings of National Conference on Artificial Intelligence*, pages 518–525, 1999.
- [32] M. Jennings-Teats, G. Smith, and N. Wardrip-Fruin. Polymorph: Dynamic difficulty adjustment through level generation. In *Proceedings of the FDG Workshop on Procedural Content Generation in Games*, 2010.
- [33] G. Kaminka and M. Bowling. Towards robust teams with many agents. In *Proceedings of International Conference on Autonomous Agents and Multi-agent Systems*, 2002.
- [34] H. Kautz. *A Formal Theory of Plan Recognition*. PhD thesis, University of Rochester, 1987.
- [35] Henry Kautz and James F. Allen. Generalized plan recognition. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 32–38, 1986.
- [36] Kristian Kersting and Luc De Raedt. Bayesian logic programming: Theory and tool. In Lise Getoor and Ben Taskar, editors, *Statistical Relational Learning*. MIT Press, 2007.

- [37] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell. A survey of mobile phone sensing. *IEEE Communications*, 49(9):140–150, 2010.
- [38] N. Lesh, C. Rich, and C. Sidner. Using plan recognition in human-computer collaboration. Technical Report TR98-23, Mitsubishi Electric Research Laboratories, 1998.
- [39] L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. In *Proceedings of National Conference on Artificial Intelligence*, pages 348–353, 2004.
- [40] Lin Liao, Dieter Fox, and Henry Kautz. Extracting places and activities from gps traces using hierarchical conditional random fields. *The International Journal of Robotics Research*, 26(1):119–134, 2007.
- [41] Daniele Masato, Timothy Norman, Wamberto Vasconcelos, and Katia Sycara. Agent-oriented incremental team and activity recognition. In *International Joint Conference on Artificial Intelligence*, pages 1402–1407, 2011.
- [42] S. McQuiggan, J. Rowe, S. lee, and J. Lester. Story-based learning: The impact of narrative on learning experiences and outcomes. In *Proceedings of the International Conference on Intelligent Tutoring Systems*, pages 530–539, 2008.
- [43] D. Minnen, T. Starner, I. Essa, and C. Isbell. Discovering characteristic actions from on-body sensor data. In *IEEE International Symposium on Wearable Computers*, 2006.
- [44] Vlad I. Morariu and Larry S. Davis. Multi-agent event recognition in structured scenarios. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [45] B. Mott, S. Lee, and J. Lester. Probabilistic goal recognition in interactive narrative environments. In *Proceedings of National Conference on Artificial Intelligence*, pages 187–192, 2006.
- [46] Sriraam Natarajan, Hung Bui, Prasad Tadepalli, Kristian Kersting, and Weng-Keen Wong. Logical hierarchical hidden markov models for modeling user activities. In *International Conference on Inductive Logic Programming (ILP)*, 2008.
- [47] N. Nguyen, D. Phun, S. Venkatesh, and H. Bui. Learning and detecting activities from movement trajectories using Hierarchical Hidden Markov Models. In *Proceedings of Computer Vision and Pattern Recognition*, pages 955–960, 2005.
- [48] CMU Quality of Life Technology Center. Grand Challenge Data Collection. <http://kitchen.cs.cmu.edu/>.

- [49] D. Olguin, B. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics: Part B*, 39(1):43–55, 2009.
- [50] A. Pentland. *Honest Signals: How They Shape Our World*. MIT Press, 2008.
- [51] W. Pentney, A. Popescu, S. Wang, H. Kautz, and M. Philipose. Sensor-based understanding of daily life via large-scale use of common sense. In *Proceedings of National Conference on Artificial Intelligence*, pages 906–912, 2006.
- [52] C. Raymond Perrault and James F. Allen. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3–4):167–182, 1980.
- [53] Ronald Poppe. A survey on vision-based human action recognition. *Journal of Image Vision Computing*, 28(6):976–990, June 2010.
- [54] David V. Pynadath and Stacy C. Marsella. PsychSim: Modeling theory of mind with decision-theoretic agents. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1181–1186, 2005.
- [55] David V. Pynadath and Michael P. Wellman. Accounting for context in plan recognition with application to traffic monitoring. In *Proceedings of Uncertainty in Artificial Intelligence*, pages 472–481, 1995.
- [56] P. Rashidi and D. Cook. The resident in the loop: Adapting the smart home to the user. *IEEE Transactions on Systems, Man, and Cybernetics (Part A)*, 39(5):949–959, 2009.
- [57] N. Ravi, N. Dandekar, P. Mysore, and M. Littman. Activity recognition from accelerometer data. In *Proceedings of the National Conference on Innovative Applications of Artificial Intelligence*, pages 1541–1546, 2005.
- [58] D. Ray, B. King-Casas, P.R. Montague, and P. Dayan. Bayesian model of behavior in economic games. In *Advances in Neural Information Processing Systems*, volume 21, pages 1345–1352, 2008.
- [59] Matt Richardson and Pedro Domingos. Markov logic networks. *Machine Learning*, 62:107–136, 2006.
- [60] Adam Sadilek and Henry Kautz. Location-based reasoning about complex multi-agent behavior. *Journal of Artificial Intelligence Research*, 43:87–133, 2012.
- [61] Adam Sadilek and Henry A Kautz. Recognizing multi-agent activities from GPS data. In *Proceedings of National Conference on Artificial Intelligence*, 2010.

- [62] S. Saria and S. Mahadevan. Probabilistic plan recognition in multiagent systems. In *Proceedings of International Conference on Automated Planning and Scheduling*, pages 287–296, 2004.
- [63] Roger C. Schank and Robert P. Abelson. *Scripts, Plans, Goals, and Understanding*. Lawrence Erlbaum Associates, Mahwah, NJ, 1977.
- [64] C.F. Schmidt, N.S. Sridharan, and J.L. Goodson. The plan recognition problem: an intersection of psychology and artificial intelligence. *Artificial Intelligence*, 11:45–83, 1978.
- [65] Mark E. Stickel. A Prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. Technical report, Artificial Intelligence Center, SRI International, 1988.
- [66] Gita Sukthankar and Katia Sycara. Simultaneous team assignment and behavior recognition from spatio-temporal agent traces. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 716–721, Boston, MA, July 2006.
- [67] Gita Sukthankar and Katia Sycara. Activity recognition for dynamic multi-agent teams. *ACM Transactions in Intelligent Systems Technology*, 3(1):18:1–18:24, October 2011.
- [68] E.M. Tapia, S. Intille, and K. Larson. Activity recognition in the home using simple and ubiquitous sensors. In *Proceedings of the International Conference on Pervasive Computing*, pages 158–175, 2004.
- [69] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476), 2006.
- [70] University of Rochester Activities of Daily Living Dataset. <http://www.cs.rochester.edu/~rmessing/uradl/>.
- [71] UW Lahar Project. RFID Data. <http://lahar.cs.washington.edu/displayPage.php?path=./content/Download/RFIDData/rfidData.html>.
- [72] A. Vahdatpour, N. Amini, and M. Sarrafzadeh. Toward Unsupervised Activity Discovery using Multi-Dimensional Motif Detection in Time Series. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1261–1266, 2009.
- [73] D. Vail, J. Lafferty, and M. Veloso. Conditional random fields for activity recognition. In *Proceedings of International Conference on Autonomous Agents and Multi-agent Systems*, pages 1326–1333, 2007.
- [74] M. Vilain. Getting serious about parsing plans: A grammatical analysis of plan recognition. In *Proceedings of National Conference on Artificial Intelligence*, 1990.

- [75] H.M. Wellman. *The child's theory of mind*. MIT Press, Cambridge, MA, 1990.
- [76] Andrew Whiten, editor. *Natural Theories of Mind*. Basil Blackwell, Oxford, UK, 1991.
- [77] Robert Wilensky. Why John married Mary: Understanding stories involving recurring goals. *Cognitive Science*, 2(3):235–266, 1978.
- [78] D. Wyatt, M. Philipose, and T. Choudhury. Unsupervised activity recognition using automatically mined common sense. In *Proceedings of National Conference on Artificial Intelligence*, pages 21–27, 2005.
- [79] Liyue Zhao, Xi Wang, Gita Sukthankar, and Rahul Sukthankar. Motif discovery and feature selection for CRF-Based activity recognition. In *Proceedings of the IAPR/IEEE International Conference on Pattern Recognition*, pages 3826–3829, Istanbul, Turkey, Aug 2010.