

Workflow Complexity for Collaborative Interactions: *Where are the Metrics? – A Challenge*

Kartik Talamadupula and **Biplav Srivastava** and **Jeffrey O. Kephart**

Human-Agent Collaboration Group

IBM T. J. Watson Research Center

Yorktown Heights, NY 10598

{krtalamad, biplavs, kephart} @ us.ibm.com

Abstract

In this paper, we introduce the problem of denoting and deriving the complexity of workflows (plans, schedules) in collaborative, planner-assisted settings where humans and agents are trying to jointly solve a task. The interactions – and hence the workflows that connect the human and the agents – may differ according to the domain and the kind of agents. We adapt insights from prior work in human-agent teaming and workflow analysis to suggest metrics for workflow complexity. The main motivation behind this work is to highlight metrics for human comprehensibility of plans and schedules. The planning community has seen its fair share of work on the synthesis of plans that take diversity into account – what value do such plans hold if their generation is not guided at least in part by metrics that reflect the ease of engaging with and using those plans?

1 Introduction

The emergence of the Internet and application (app) centric service-oriented platforms for various kinds of consumer tasks have resulted in an explosion in the interactions between humans and automated agents that assist them in tasks. Given their large number, a formal measurement of the inherent complexity of these interactions is desirable to assist in the design of useful and efficient decision-making algorithms and systems.

We present a usecase that illustrates the kinds of interactions we are discussing – consider a person living in New York who wants to book a travel itinerary for a short personal trip to Seattle. The workflow for planning this trip will consist of a flight reservation, hotel reservation, and reservation for local travel in the source and destination cities. These bookings could each be made via websites, over a dialog interface, via IoT interfaces, or manually over the phone with a travel agency. Each communication modality introduces its own constraints and complexity. We highlight the workflow complexity in this specific example using Figure 1. Here, three action instances are shown for booking a flight, a hotel, and local travel. The data artifacts are the booking confirmations, whose variables constrain the other actions in the workflow. In the workflow fragment that is shown, local travel at the destination is most constrained as it depends on the flight’s arrival time, as well as the location

of the hotel at the destination. In general, the flight booking will result in dates (and times) which create a dependency for the hotel reservation. Finally, flight and hotel reservations give the date and locations for which local travel needs to be booked. The overall complexity of booking this short leisure trip may differ from a business trip, where meeting schedules have to be taken into account; and may further differ from an international trip where the processing of travel documents has to be taken into account.

In such scenarios, automation faces two main challenges. The first is the problem of knowledge acquisition and engineering pertaining to the domain of interest – in the travel scenario above, this knowledge would constitute the various actions available to the agent to create a successful workflow, and the dependencies between those actions. This kind of problem is the purview of the flourishing Knowledge Engineering for Planning & Scheduling (KEPS) community. The second major problem is that of explaining the plan and the interactions underlying it to the (human) user/consumer of the plan. An important sub-problem in this is *measuring* the complexity of the said interaction – without such measures, an automated system that is trying to aid in such interactions will be unable to distinguish between and rank workflows of vastly differing complexities that all achieve the same goal. Complexity measures provide the ability to rank the planner’s mediation in such scenarios, and allow the planner to produce directed help that will enable easier achievement of the user’s goals. We highlight this second problem in this paper.

2 Prior Work

There is a rich body of work on workflow representation, composition, and execution (van der Aalst and van Hee 2004). Over the past decade, there have been approaches for semi or fully automated composition of workflows using planning that look at control and data driven issues (Srivastava and Koehler 2003). However, much of this work is in the context of single agent decision-making. There is no prior work, to our knowledge, that characterizes the complexity of workflows in a collaborative setting.

The planning community has also seen advances in the problems of measuring the distance between plans (Roberts *et al.* 2014; Goldman and Kuter 2015), and generating diverse plan alternatives (Nguyen *et al.* 2012). However, very

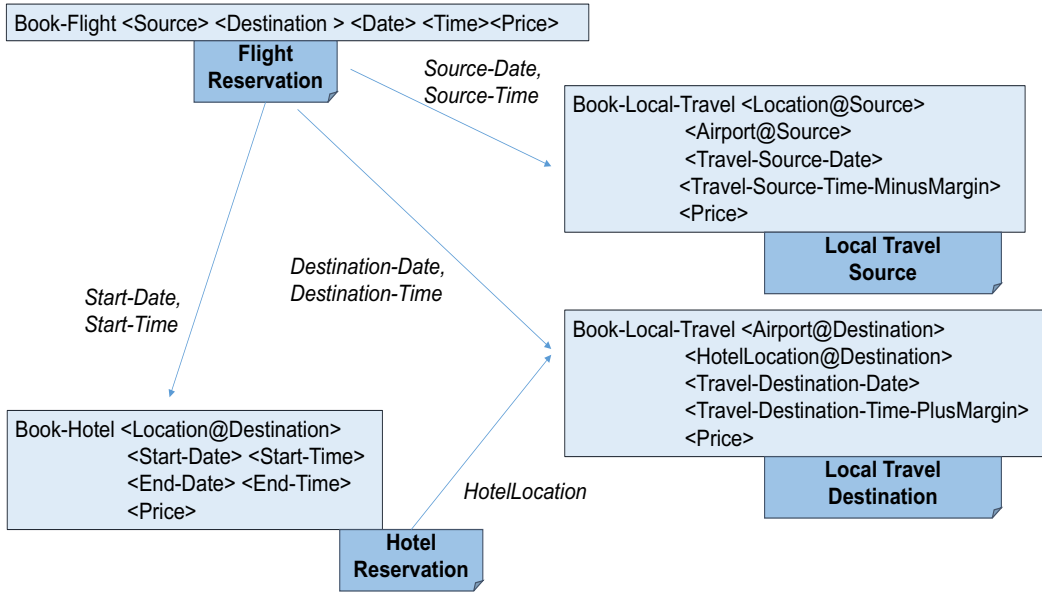


Figure 1: Actions, data (parameters) and constraint variables in a small travel example.

little research has focused on exactly what the different metrics that go into creating diverse plans should be. Such work has mostly looked at measures (cost, duration, robustness, etc.) that treat the plan as an artifact disconnected from humans, who must execute, understand, or participate in that workflow. Humans typically perceive complexity both from interaction issues, as well as from the actions in a workflow. Indeed, there is a long history of prior work from a linguistic and structural perspective for the former (Liao *et al.* 2017). However, there has been no focus on creating a class of metrics that attempt to define the complexity of a plan or workflow. We intend this paper as a challenge to the community to do exactly that.

3 Workflow Complexity: Example Usecases

We described the Travel Booking usecase in Section 1; here, we describe some other collaborative examples to highlight complexities that an automated decision making system can help reduce.

Scheduling Meetings A common collaborative task in the workplace is deciding a meeting time and venue, given a topic. This mundane task is complicated by the fact that there are different roles for participants in the meeting, hard and soft scheduling constraints, and limited access to participant information which changes with context. In such scenarios, setting up a meeting between colleagues who are at the same level organizationally may be more complex than one convened by the head of the organization – in the former there may be more hard constraints and various alternatives have to be considered, while in the latter everyone is likely to mark their (conflicting) constraints as soft. An automated

agent (Cranshaw *et al.* 2017) can play a crucial role in improving the efficiency of meeting scheduling¹. Specifically, it can verify participants and roles, identify potential conflicts from existing schedules, ask (the fewest number of) people to re-visit their constraints, and explain alternative time-slots.

Evaluating Hiring Choices Another workplace example is the evaluation of a set of candidates by a multi-disciplinary panel of experts. The experts may evaluate the candidate’s technical skills, non-technical (soft) skills, organizational fit, HR concerns, career progression etc. Depending on the role, the process may involve many interview rounds, evaluations, and discussion. Further complexity is added by variations in the evaluation scales, disagreements among the experts, and relative weights of selection criteria. An automated agent can make this process more efficient by formalizing the contributions of the experts, focusing the team on key decision factors, retrieving relevant candidate data, eliminating human bias, and providing justifications to the stakeholders when asked.

Human-Robot Teaming Planning for human-robot teaming (HRT) (Talamadupula *et al.* 2010; Chakraborti *et al.* 2016b) considers the problem of humans and robots in goal-oriented environments, and the planner’s mediation through control of the robotic agent. HRT scenarios usually involve extensive interaction between the human and the robot. Automated mediation can make the teaming more efficient in

¹This is distinct from the actual scheduling problem, which is to find a satisfying assignment given everyone’s constraints – our problem considers the workflow of scheduling the meeting.

USECASE	METRIC								
	NT	IT	AD	FO	Com	EC	PC	MC	
Travel Booking	L	H	H	H	H	H	M	M	
Scheduling Meetings	H	M	M	L	H	L	H	H	
Evaluating Hiring Choices	L	H	H	L	H	M	H	H	
Human-Robot Teaming	M	H	M	M	L	M	L	L	
Medical Treatment	H	L	L	L	H	M	H	H	
Personal Finance	M	M	H	L	H	H	H	H	

Table 1: Workflow complexity metrics and their footprint; H - High, M - Medium, L - Low; Metrics described in Section 4.

a number of ways, including coordination to reduce communication (Talamadupula *et al.* 2014), and restricting the number of agents that a human has to deal with.

Deciding a Medical Treatment Plan Another illustrative collaborative task, from the area of health, is deciding a medical plan for a person given a health condition (initial state) and a desirable new condition (goal state). For example, if a pregnant person has to be operated on for a planned child birth, specialists of the concerned medical fields need to coordinate specific procedures; schedule it with relevant nursing staff; complete insurance formalities; and reserve resources like the operation room. Some of these processes follow standardized or regulated workflows, while others are case-specific depending on patient risk factors, etc. Furthermore, the data in such scenarios must be controlled due to confidentiality and regulatory reasons (Leyens *et al.* 2017). An automated decision maker can help by focusing the attention of the medical team on ensuring compliance, examining risk factors and medical requirements, and avoiding costly mistakes that may foreclose future remedial actions.

Personal Finance Increasingly, personal finance has emerged as an area of great opportunity as well as challenge for decision making systems and decision assistants. Use-cases like buying a house, saving for retirement, or filing one’s taxes are important decisions with long-term life implications. A number of characteristics must be considered including the various alternatives available, their costs (both immediate and future), legal and compliance issues, etc. A specific example of such a decision making scenario is an automated tax assistant – such an assistant must be aware of the tax code which prescribes various rules and regulations that must be followed, and must recommend the best tax plan while optimizing a number of metrics including minimizing amount paid as tax, minimizing the complexity of the plan, and maximizing compliance (to minimize the chances of audits and fines).

4 Metrics

We now list some metrics from prior work that can be adapted to the problem we consider. Chakraborti *et al.*

(2016a) provide a framework for studying and evaluating interaction between human and robot team-members in goal-oriented environments. Some useful metrics that can be adapted from that work are:

1. **Neglect Tolerance (NT)**: How long the agent is able to perform well without human intervention.
2. **Interaction Time (IT)**: Time spent in communication.
3. **(Robot) Attention Demand (AD)**: Measures the attention demanded by the agent.
4. **Fan Out (FO)**: Communication load on the humans; proportional to the number of agents.
5. **Compliance (Com)**: How well the actions of an agent convey its intention to comply.

Separately, Keller *et al.* (2007) consider the problem of workflow complexity relating to configuring Information Technology (IT) infrastructure, e.g. a web application. They define configuration complexity as “the complexity of carrying out a configuration procedure as perceived by a human system manager”; and track information along three dimensions, which are (respectively) analogous to control flow, data flow, and space complexity in software engineering:

6. **Execution Complexity (EC)**: Number of actions and context switches.
7. **Parameter Complexity (PC)**: Number of parameters used by actions, and their usage variations.
8. **Memory Complexity (MC)**: Number of configuration values which need to be remembered along the workflow, and over the actions.

These measures are all relevant from a human-agent collaboration perspective, as they relate to the effort needed to review a plan and to gain human trust.

5 Discussion

In Table 1, we present the above metrics juxtaposed with their footprint in the collaborative domains introduced in Section 3. The footprint itself is quantized into three categories – **High (H)**, **Medium (M)**, and **Low (L)**. We address

a number of points in relation to the table. First and foremost, the table should be read column-wise, for each metric. Second, the High/Medium/Low annotations denote the typical or average-case profile for that metric in the respective usecase, and may vary depending on the specific problem instance etc.

Third, these values do not represent any intrinsic *goodness* – high neglect tolerance is good in scenarios like Human-Robot Teaming, because it shows that the automated agent is more independent; while low compliance might be a bad thing if the human wants constant confirmation or reassurance from the agent, like in medical treatment scenarios. However, these can easily switch depending on the domains and users in question: medical professionals may want a less independent agent (lower neglect tolerance), while meeting scheduling agents may not be required to show all the steps of their work. A general rule-of-thumb is that if the metric profile of a particular usecase is reflected in the plans that a planner produces, overall team success is more likely.

We now discuss the metrics from Table 1 in the context of creating new metrics that define the complexity of plans or workflows in terms of the interaction issues, as well as the complexity of the actions that constitute those workflows. The first set of metrics informally represent interaction issues: Neglect Tolerance (NT), Interaction Time (IT), and Attention Demand (AD) are related to each other, and are concerned with the demands that a workflow imposes on the user/human via the agent’s roles in the workflow. Similarly, IT and Fan Out (FO) offer a measure of the communication that is expected from the user, and how many different agents the user must accommodate (the assumption being that communication load increases as a function of the number of such agents). The second set of metrics represents the complexity of the actions in the workflow itself: while a scenario that involves scheduling meetings might feature a number of possible alternative workflows and each action might consist of multiple parameters, other scenarios like human-robot teaming might in fact feature relatively fewer alternatives and action parameters. These are all important to track in the final plan that is generated for the human-agent team, since they contribute to the difficulty of explaining the workflow and its constituent parts (as required).

6 Conclusion & Future Work

We conclude by reiterating that the metrics we discuss in this paper differ from the traditional metrics used in the planning community, which apply specifically to actions and goal-states; the optimal profiles for these metrics are instead at least partially determined by the usecase in question. We would like to use these as a starting point in ultimately creating metrics that explain the complexity of the workflow cumulatively from the perspective of the agent that must understand, explain, or execute it. Our hope is that this paper will spur action in two directions: (1) the post-processing of plans from existing planners to take cumulative plan complexity metrics into account; and eventually, (2) the creation of new planners that can handle such complexity metrics directly in the state-space search and plan synthesis processes.

References

- Tathagata Chakraborti, Kartik Talamadupula, Yu Zhang, and Subbarao Kambhampati. A formal framework for studying interaction in human-robot societies. In *AAAI 2016 Workshop on Symbiotic Cognitive Systems (SCS)*, 2016.
- Tathagata Chakraborti, Yu Zhang, David E Smith, and Subbarao Kambhampati. Planning with resource conflicts in human-robot cohabitation. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 1069–1077. International Foundation for Autonomous Agents and Multiagent Systems, 2016.
- Justin Cranshaw, Emad Elwany, Todd Newman, Rafal Kocielnik, Bowen Yu, Sandeep Soni, Jaime Teevan, and Andrés Monroy-Hernández. Calendar. help: Designing a workflow-based scheduling agent with humans in the loop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2382–2393. ACM, 2017.
- Robert P Goldman and Ugur Kuter. Measuring Plan Diversity: Pathologies in Existing Approaches and A New Plan Distance Metric. In *AAAI*, pages 3275–3282, 2015.
- A. Keller, A. B. Brown, and J. L. Hellerstein. A Configuration Complexity Model and Its Application to a Change Management System. *IEEE Transactions on Network and Service Management*, 4(1):13–27, June 2007.
- Lada Leyens, Matthias Reumann, Nuria Malats, and Angela Brand. Use of big data for drug development and for public and personal health and care. *Genetic Epidemiology*, 41(1):51–60, 2017.
- Q. Vera Liao, Biplav Srivastava, and Pavan Kapanipathi. Tailoring Conversational UX through the Lens of Dialogue Complexity. In *CHI Workshop on Conversational UX Design*, 2017.
- Tuan Anh Nguyen, Minh Do, Alfonso Emilio Gerevini, Ivan Serina, Biplav Srivastava, and Subbarao Kambhampati. Generating diverse plans to handle unknown and partially known user preferences. *Artificial Intelligence*, 190:1–31, 2012.
- Mark Roberts, Adele E Howe, and Indrajit Ray. Evaluating Diversity in Classical Planning. In *ICAPS*, 2014.
- Biplav Srivastava and Jana Koehler. Web Service Composition - Current Solutions and Open Problems. In *In: ICAPS 2003 Workshop on Planning for Web Services*, pages 28–35, 2003.
- Kartik Talamadupula, J Benton, Subbarao Kambhampati, Paul Schermerhorn, and Matthias Scheutz. Planning for human-robot teaming in open worlds. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 1(2):14, 2010.
- Kartik Talamadupula, Gordon Briggs, Tathagata Chakraborti, Matthias Scheutz, and Subbarao Kambhampati. Coordination in human-robot teams using mental modeling and plan recognition. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 2957–2962. IEEE, 2014.
- Wil M.P. van der Aalst and Kees van Hee. Workflow Management: Models, Methods, and Systems. In *ISBN:978-0262720465, MIT Press*, 2004.