

Modeling Agents, Roles, and Positions in Machine Learning Project Organizations

Rohith Sothilingam¹ and Eric Yu^{1,2}

¹ Faculty of Information, University of Toronto, Toronto, Canada

² Department of Computer Science, University of Toronto, Toronto, Canada

Abstract. As Machine Learning (ML) continues its emergence across numerous industries, software teams and organizations face new challenges beyond those found in conventional software projects. The design of data science teams in ML software projects can vary substantially based on the organization's maturity, personnel availability, and their relationship with customers. In an empirical case study of three ML software project organizations, we examined variations in project team designs using i* models. We consider the usefulness of the concepts of Agents, Roles, and Positions defined in the original i* framework to support the analysis of complex organizational relationships. We illustrate how the Position concept helps distinguish the different ways in which each ML software project organizes its team to meet specific needs.

Keywords: Organization modeling, data science, project organization, roles

1 Introduction

Most ML projects and organizations are at the early stages of the maturity curve, compared to other types of software projects [1]. The software development lifecycle for ML is similar to agile software engineering processes, with the cyclical nature of workflows, involving the collaboration between people from diverse backgrounds and skillsets [2].

Organizations assign sets of related responsibilities to people based on their capabilities and project needs. This is known as role mapping. There are many ways to allocate roles. To allocate roles effectively, roles should be assigned to people with skillsets matching those required of the role. Without accurate role mapping, business outcomes will be hindered due to a mismatch between project needs and the qualifications and abilities of team members [7].

Unlike in more mature areas of software development, roles in ML projects are still evolving and ill-defined. Furthermore, there is a shortage of trained personnel available to ML software organizations [4]. As revealed in recent empirical literature, technical ML engineering roles are not enough to cover the diverse types of expertise needed in the ML lifecycle [9]. Organizations often improvise in ML projects to allocate responsibilities to available personnel who do not entirely have the right types of expertise [9].

In an empirical case study, we identified a number of key issues regarding team design at three ML project organizations. In this paper, we consider how i* modeling might help analyze such issues. For example:

- Can goal evaluation in an i* model help identify the underlying factors behind unsuccessful ML projects?
- Can people issues such as dissatisfaction among team members be analyzed and addressed with the help of i* modeling?
- Is the i* concept of Position useful for designing ML project teams in which team members cover multiple roles?

2 Agents, Roles, and Positions in i*

The concepts of Agent, Role, and Position were introduced in i* for modeling complex organizational relationships [8]. A Role is an abstract characterization of a social actor. An Agent, which can *play* (one or more) Role(s), represents a physical entity, such as a person. The Position concept mediates between Agents and Roles so as to provide an abstraction for a bundle of roles that is typically allocated to a single Agent. The Agent is said to *occupy* the Position, while the Position *covers* the set of Roles. These terms are capitalized in this paper to refer to the i* concepts, to be distinguished from their general English usage. In everyday usage, it is common to speak of a person being hired into a role, such as a Project Manager or a Data Scientist. In i*, these would be treated as Positions, if each of them encompasses multiple Roles, such as Assigning Tasks, Monitoring Progress, and Evaluating Performance [8].

In ML projects, job roles often require a diversity of skillsets and knowledge areas [4] [6], including application domain knowledge [7]. Practitioners are required to occupy multiple Roles, which involve different skillsets and expertise [9]. An Agent hired to occupy a Position should possess the competencies and skills required to fulfill all the Roles covered by the Position. Previous work have noted how job titles at different organizations can vary based on responsibilities and expertise required [3] [9].

3 Model-based analysis of ML project team design

We conducted an empirical study where we studied three ML project organizations. The organizations differ in size, level of maturity in ML projects, and the types of products and services offered. Organization A is a large international financial organization which develops e-commerce payment systems globally. Organization B builds advanced ML systems for customers, drawing on research in deep learning. Organization C builds AI systems to help organizations screen candidates for hiring.

To compare the team structure design of these organizations, we use a simplified i* model showing Agents, Roles, and Positions and how they are associated with each other through *plays*, *occupies*, and *covers* links, while omitting strategic dependencies and rationales (Fig. 1). We arrived at the configuration of i* Agents, Roles, and Positions at each organization from interview data. We assigned each job title to be an i* Position. We named i* Roles for each set of responsibilities expected for each Position.

As we learned about the professional background, training, and skills of team members, we selected suitable naming for the Agents.

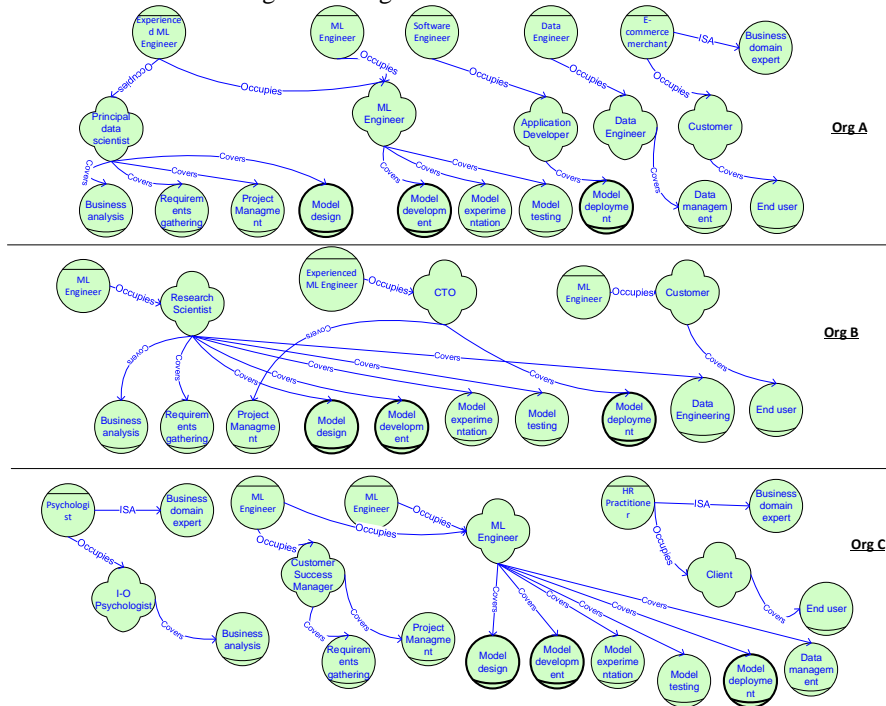


Fig. 1. Actor-Association Model of Three ML Project Organizations

We note that there is much commonality among the Roles found in the three organizations: End user, Business analysis, Project management, Model development, Model experimentation, Model design, Model testing, and Model deployment. However, there is considerable variance in how the Roles are grouped into Positions. In Organization A, the Roles of Model Design, Model Development, and Model Deployment are covered by different Positions, and are played by different classes of Agents. In Organization B, the first two Roles are covered under the same Position (Research Scientist), whereas in C, all three Roles are covered by the ML Engineer Position, which is occupied by a ML Engineer Agent, i.e., someone with the competencies and skills of an ML engineer.

In Organization B, the Research Scientist Position, occupied by an ML Engineer Agent, covers the Role of several technical Roles, such as Model Deployment. At Organization C, the Business Analysis Role is covered by an I-O Psychologist Position occupied by a Psychologist Agent. This analysis shows that Role assignment can differ substantially based on how Positions are defined in each organization. The design of the Positions may be constrained by the existing skills and competencies available and the kinds of talent they can attract, recruit, and retain.

The diverse expertise required of Roles which a Position covers will determine what expertise is required of an Agent occupying the Position. In Fig. 1, we can see that Organization C has introduced a Psychologist Agent who is a specialization of Business

Domain Expert. Using modeling, Organization C can evaluate their project team design and analyze how well this Agent can satisfy the I-O Psychologist Position based on the Roles it covers. The Psychologist Agent is a business domain expert, providing them with the right expertise to perform the Business Analysis Role.

The validity of the empirical findings are limited by our interview data, which have been obtained through one individual at each organization - the Principal Data Scientist, the CTO, and founder and CTO at Organizations A, B and C respectively.

4 Analyzing project team design

In this section, we use i* Strategic Rationale modeling and goal evaluation to analyze a past issue in Organization C (Fig. 2) and how it was subsequently addressed (Fig. 3). The detailed i* modeling allows us to identify the specific underlying factors behind why this organization's customers were not satisfied in their early history. We use slightly heavier border elements for highlighting tasks typically seen in ML projects. Using this convention, we can see how typical ML activities appear in i* models, to analyze their dependency relationships.

In the past (Fig. 2), the Business Analysis Role was assigned to the ML Engineer Position, which was occupied by an ML Engineer Agent who lacked the Business domain expertise (Resource element) required for the Business Analysis Role. Using goal evaluation, we can see that the End user Role's goal of Business objectives satisfied could not be fully achieved due to the insufficient Business domain expertise of the ML Engineer Agent.

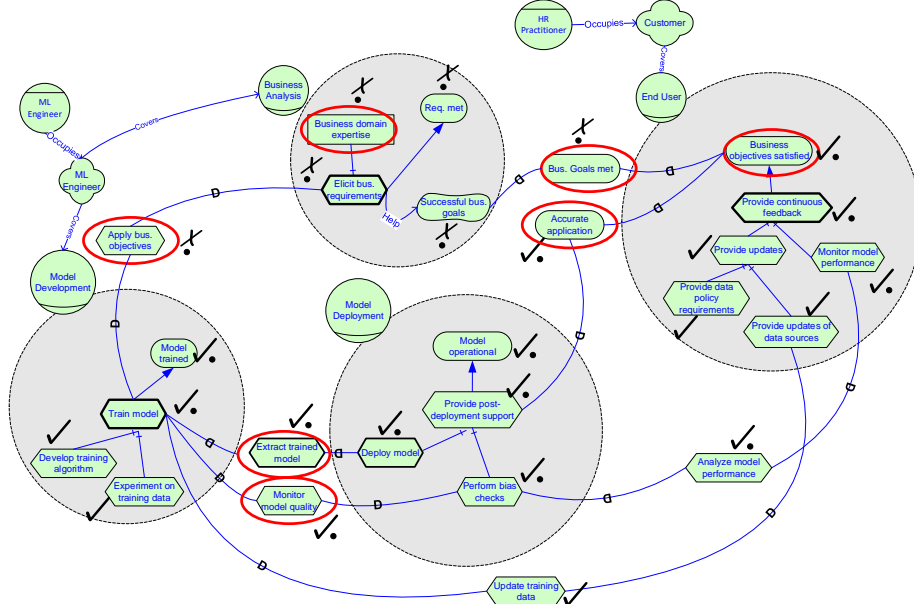


Fig. 2. Identifying insufficient role mapping

5 Conclusions and future work

In this paper, we used modeling to identify how ML or data science teams can vary substantially in their team design. We demonstrated the use of the i* concept of Position, not included in the iStar 2.0 Core, for modeling complex organizational relationships, and as a step toward addressing the challenge of mapping roles in ML projects to the right people based on expertise. Using modeling, we were able to identify the lack of business domain expertise as the underlying factor contributing to why a ML project organization was facing challenges with their customer satisfaction. By identifying where specifically failure is occurring, organizations can diagnose challenges in team design in greater detail, through the early detection of the problem. In future work, we plan to consider expertise and domain knowledge of Agents in an i* extension to help improve the analysis of Role mapping.

References

1. Akkiraju, R., Sinha, V., Xu, A., et al. (2018). Characterizing machine learning process: A maturity framework. arXiv preprint arXiv:1811.04871
2. Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4), 105-120.
3. De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, 54(5), 807-817.
4. de Sá Sousa, H. P., & do Prado Leite, J. C. S. (2017). Requirement patterns for organizational modeling. In *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)* (pp. 252-259). IEEE.
5. Fazel-Zarandi, M., & Fox, M. S. (2012). An Ontology for Skill and Competency Management. In *Proceedings of the 4th Conf. on Formal Ontologies in Information Systems (FOIS)* (pp. 89-102).
6. Linden, A., Idoine, C., Hare, J., & Brethenoux, E. (2018). Staffing Data Science Teams: Map Capabilities to Key Roles. Retrieved October 15, 2019, from <https://www.gartner.com/document/3888468?ref=TypeAheadSearch&qid=d8a9c1b58ba53d2ab0&q=Staffing Data Science Teams: Map Capabilities to Key Roles>.
7. Saltz, J. S., & Grady, N. W. (2017, December). The ambiguity of data science team roles and the need for a data science workforce framework. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 2355-2361). IEEE.
8. Yu, E. S., & Mylopoulos, J. (1994). Understanding “why” in software process modelling, analysis, and design. In *Proceedings of 16th international conference on software engineering* (pp. 159-168). IEEE.
9. Zhang, A. X., Muller, M., & Wang, D. (2020). How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *arXiv preprint arXiv:2001.06684*.