Predicting the Priority of Social Situations for Personal Assistant Agents

Kola, Ilir; Tielman, Myrthe L.; Jonker, Catholijn M.; van Riemsdijk, M. Birna

DOI
10.1007/978-3-030-69322-0_15

Publication date
2021

Document Version
Final published version

Published in
PRIMA 2020

Citation (APA)
Kola, I., Tielman, M. L., Jonker, C. M., & van Riemsdijk, M. B. (2021). Predicting the Priority of Social Situations for Personal Assistant Agents. In T. Uchiya, Q. Bai, & I. Marsá Maestre (Eds.), *PRIMA 2020: Principles and Practice of Multi-Agent Systems - 23rd International Conference, 2020, Proceedings* (pp. 231-247). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 12568 LNAI). Springer. https://doi.org/10.1007/978-3-030-69322-0_15

Predicting the Priority of Social Situations for Personal Assistant Agents

Ilir Kola[1](✉), Myrthe L. Tielman[1], Catholijn M. Jonker[1,2], and M. Birna van Riemsdijk[3]

[1] Delft University of Technology, Delft, The Netherlands
{i.kola,m.l.tielman,c.m.jonker}@tudelft.nl
[2] Leiden Institute of Advanced Computer Science, Leiden, The Netherlands
[3] University of Twente, Enschede, The Netherlands
m.b.vanriemsdijk@utwente.nl

Abstract. Personal assistant agents have been developed to help people in their daily lives with tasks such as agenda management. In order to provide better support, they should not only model the user's internal aspects, but also their social situation. Current research on social context tackles this by modelling the social aspects of a situation from an objective perspective. In our approach, we model these social aspects of the situation from the user's subjective perspective. We do so by using concepts from social science, and in turn apply machine learning techniques to predict the priority that the user would assign to these situations. Furthermore, we show that using these techniques allows agents to determine which features influenced these predictions. Results based on a crowd-sourcing user study suggest that our proposed model would enable personal assistant agents to differentiate between situations with high and low priority. We believe this to be a first step towards agents that better understand the user's social situation, and adapt their support accordingly.

Keywords: Social situations modelling · Adaptive personal assistants · Machine learning techniques · Explainable AI

1 Introduction

Artificial agents that play the role of personal assistants are increasingly becoming part of everyday life (e.g. [14]). These agents have focused on representing internal aspects of the user, such as their values, goals, or emotions [25]. However, research in social science suggests that human behaviour is shaped both by these internal aspects, as well as by the situation someone is in [16]. Situations have a physical aspect (e.g., where it takes place) and a social one (e.g., who

This work is part of the research programme CoreSAEP, with project number 639.022.416, which is financed by the Netherlands Organisation for Scientific Research (NWO).

T. Uchiya et al. (Eds.): PRIMA 2020, LNAI 12568, pp. 231–247, 2021.
https://doi.org/10.1007/978-3-030-69322-0_15

is involved). We focus on the latter: our goal is to build methods which allow personal assistant agents to model the social situation of a user, and use that information to reason about how to provide socially-aware support.

The need for enabling intelligent support agents (such as personal assistants) to understand the social situation of the user has been acknowledged as one of the main open questions in agent research [12,32]. Existing work on modelling social context focuses on modelling the social practices of a situation (e.g. [7]), or the place where the interaction is taking place (e.g. [20]). In our approach, we model situations from the perspective of the user of the personal assistant agent by modelling how the user relates to the people in that situation on a number of relevant dimensions. This complements [7], which models the social practices *of a situation*, while we focus on modelling the perspective of an individual *on that situation*. This requires additional social features to describe social relations between people that go beyond their roles in the situation. Based on information about how the user relates to the social situation, we investigate how an agent can interpret that situation in order to determine desired actions that can support the user. Regarding our technical approach, we combine the strengths of existing work: we propose an explicit model of a social situation (similar to [7]), and combine it with learning techniques to derive new information (similar to [20]).

To illustrate our approach, we take the example of a personal assistant agent which helps busy users manage their agenda automatically (e.g. [21]). We consider each meeting to be a social situation. The agent takes as input situation cues (e.g. the setting of the meeting, such as a work meeting) and relationship features (e.g. the quality of the relationship, such as a very positive relationship) [15]. Based on this the agent determines which meeting the user would likely want to attend when two meetings overlap. If the user is too busy to respond to meeting requests themselves, the agent can take this decision for the user. The agent may then inform the user about this choice while noting which aspects of the situation led to this choice. This is a first step towards enabling the agent to explain its decisions to the user.

In this paper we investigate the building blocks that would be needed to create such a personal assistant agent. First of all, we need a way to determine which meeting is considered to be more important to the user. To facilitate this process, we quantify the importance of each meeting by assigning it a numerical value to which we refer as the *priority* score of the meeting. Our assumption is that people implicitly follow this priority score when deciding about conflicting meetings by choosing the one with the highest priority. This will be evaluated via our research hypothesis:

RH - When choosing between two meetings, people select the one with higher priority in the majority of the cases.

The task of the agent now becomes to learn a model which predicts the numerical priority of meetings. We explore whether we can tackle this task by using machine learning techniques on a data set containing information on hypothetical meeting scenarios collected from multiple people. This leads to our first research question:

RQ1 - Can we use machine learning techniques to predict the priority of social situations based on situation cues and relationship features?

In our view for human-centered personal assistants, the ability of the agent to explain its decisions to the user is a fundamental requirement. This is because in such a system, it is important for the user to trust the suggestions of the agent. Lim et al. [17] suggest that in socio-technical applications, users trust the agent more when they understand why the agent has selected attending a specific meaning. Making the decisions of the agent explainable consists of three parts: the agent should be able to determine the internal processes that led to a certain suggestion, to generate an explanation based on them, and to present this explanation to the user [22]. Our focus is on the first part: we explore methods that allow the agent to determine which features of a social situation contribute to the prediction of priority. The other parts will be explored in future work.

Our predictive model is built using information from multiple people, however people can have differing preferences. To achieve more personalization, we extend the model by including personal values as input features for our predictive model. Values are considered to be a driving factor in human behaviour [8], so we explore their role in helping better predict the priority of social situations:

RQ2 - Does adding information about the personal values of users as input features to the predictive model increase that model's accuracy of prediction of the priority of social situations?

The rest of this paper is structured as follows: In Sect. 2 we present our approach for tackling the research questions and hypothesis. In Sect. 3 we introduce background knowledge related to the concepts we use. Sect. 4 presents a crowdsourcing user study conducted to collect data for building and evaluating our models, which is done in Sect. 5. Section 6 concludes this paper.

2 Proposed Approach

We propose an architecture that allows personal assistants agents to model the user's social situation, and use this information to predict the priority of this social situation, or, in other words, predict how important this specific situation is. A high-level depiction of the architecture is presented in Fig. 1[1]).

Overall, the framework works as follows: In the offline stage, a supervised learning algorithm takes as input multiple social situations from different users, described in terms of their social relationship features and situational features. The learning target is the priority of these situations. This forms our prediction model. During run time, the personal assistant agent is provided with the features of two different meetings which overlap. Using the priority prediction model, it determines the priority score of each meeting, it keeps on the schedule the meeting with the highest priority, and informs the user. At this point, the agent also determines the features that have the highest impact on this prediction, which will in future work lead to generating explanations.

[1] Icons used in Fig. 1 were made by Freepik and retrieved from www.flaticon.com.

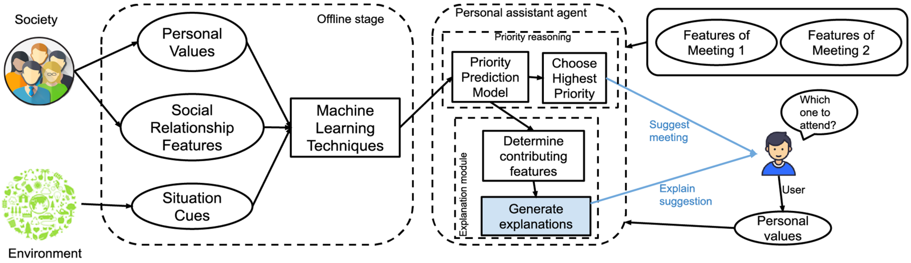

Fig. 1. High level representation of the proposed architecture. Circles represent the modelled concepts, whereas boxes represent learning/reasoning steps. Items marked in blue are concepts that we do not explicitly tackle in this work. (Color figure online)

In order to provide more personalized support, we add to the model information about the personal values of the user. The assumption is that people with similar value preferences will also assign more similar priorities to specific situations. For instance, users who value achievement and success might give a higher priority to work meetings. Therefore, having the value information as an input in our model can potentially lead to better predictions.

A key concept that we use in this work is assigning a numerical score to the priority of social situations. Using this approach, as opposed to directly choosing between two conflicting meetings from a set of input features, has several advantages. First of all, using priority can facilitate the explanations given to the user by the personal assistant agent: the agent first tells the user which meeting has the highest priority, and secondly it explains why. Furthermore, having a numerical representation of priority comes with technical benefits, since the task of learning preference rankings from pairwise choices can be computationally intractable [5].

3 Concepts and Methods

In our vision, personal assistant agents should be able to provide human-centred support. This means the support actions should be transparent and intelligible. This guides our choices from two points of view: concepts (Sect. 3.1) and techniques (Sect. 3.2) used for modelling. This means we should use techniques that allow insight into their decision making process, combined with concepts that are understandable to the users. For this reason, we combine explainable machine learning techniques with concepts from social sciences. Using explainable machine learning techniques means that, when given a set of features which model a social situation, the model is able to output both a prediction as well as which features contributed to this prediction. Lim et al. [17] show that such a procedure improves the intelligibility of context-aware intelligent systems. The set of features that we use to model social situations is borrowed from social science literature. Since these are concepts that we use in everyday life, they should be understandable to the user. In this section, we present the rationale behind the concepts and techniques that we use.

3.1 Social Science Concepts

Our focus in this work is on modelling social situations - situations involving our user and people from their social circle. Kola et al. [15] propose modelling social situations involving two people as a combination of social relationship features, which represent how the two people are related to one another, and situation cues, which represent the circumstances in which the situation takes place. In this approach, the set of features that describes a social situation is based on social science literature that aims at modelling the relevant aspects of social relationships and situations. These concepts can be both concrete and objective (e.g., geographical distance between the user and the other person, for how long they know each-other), as well as subjective (e.g., quality of the relationship between the user and the other person).

Personal values represent key drivers of human decision making [27,29]. Friedman and colleagues [8] define values as “what a person or group of people consider important in life”. People hold various values (e.g. wealth, health, independence) with different degrees of importance. The most prominent models of human values were proposed by Rokeach [27] and Schwartz [29]. In our work, we use the model proposed by Schwartz since it offers validated measurement instruments with fewer items than Rokeach, which makes them more suited to online surveys. Furthermore, Schwartz builds on the work of Rokeach and other researchers, so there is overlap in their proposed value lists.

3.2 Machine Learning Methods

A predictive model for a personal assistant has to fulfil two main requirements. Firstly, it should be able to achieve satisfying accuracy for smaller data sets, since acquiring large amounts of data from human subjects can be challenging. Secondly, in order to provide human-centred support, the algorithms should provide insights on which features influence a specific prediction. Ensemble methods [6] are a family of machine learning techniques which fit these requirements. These techniques combine predictions from multiple learning algorithms in order to increase accuracy. The idea is to combine accurate and diverse weaker learners, in order to exploit the strength of each learner. Ensemble methods are shown to perform better than the individual learners they consist of [6]. Furthermore, they are shown to perform well and generalize better for smaller data sets [26]. When the base learners are decision trees, it is also possible to have insights into the features that led to a certain decision, as we will show further on. Another advantage is their accuracy when dealing with structured data. A recent survey [28] shows that ensemble methods have won different machine learning competitions, thus demonstrating high predictive power.

Some of the more successful methods are random forests [3] and gradient boosting machines [9]. Random forests are a specific example of bagging methods

[2]. In bagging, each learner is built independently over a random sub-sample of the data, and the decision is made by aggregating the outputs. The sub-sampling procedure reduces the variance of the method, which is usually a problem for decision trees. In addition, in random forests, for each split of the tree only a subset of the features is considered, this way we avoids the possibility of all the trees selecting the same features and ignoring others. Gradient boosting machines are an example of boosting methods [9]. In boosting, weaker learners are trained sequentially (and not in parallel like in bagging), and each new learner tries to correct its predecessor. Gradient boosting achieves this by fitting the new predictor to the residual errors of the previous ones.

Understanding why a model makes certain predictions is a general goal in machine learning, especially when it comes to providing human-centred support. Lundberg and Lee [18] propose a unified framework for interpreting predictions, called Shapley Additive Explanations (SHAP). The benefits of using this framework are that it provides both global interpretability for the model (i.e. which are generally the most important predictors), as well as local interpretability (i.e. which are the most influential predictors for each individual observation). We use this framework in order to gain insight into the predictions of our model.

4 Crowd-Sourcing User Study

In this user study we gather data for constructing and evaluating our models. The study was approved by the ethics committee of Delft University of Technology.

4.1 Choice of Concepts

Features of Social Situations. In order to set up the user study, we need to define a set of features that will be used to model social situations. Our starting point is the feature set proposed in [15], where social situations are described through a set of relationship features and a set of situation cues. Since their feature set is based on a limited number of social science models, we start by conducting a more extensive literature review. Then, we conduct an exploratory pre-study in order to investigate what aspects of a social situation people take into account when determining how important that situation is. Thus, our feature set is evaluated both from a theoretical and practical perspective.

In our literature review, we found five comprehensive models which aim at describing aspects of dyadic social relationships, three of which were not taken into consideration in [15]. The results are presented in Table 1.

Next, in the exploratory pre-study, we collect answers from 33 participants through Amazon Mechanical Turk[1]. Our goal was to explore which features do participants find important when thinking about the priority of social situations. First, participants were asked to describe five social situations in which they participated in the past week. They were instructed to provide at least the

[1] https://www.mturk.com/.

time, location, activity and role of the other person, in order to ensure they were thinking about concrete situations. Furthermore, these suggested activities provide the basis for the formation of the hypothetical scenarios in our main user study (Sect. 4.2). Then, they were asked to consider which aspects of the situation play a role in determining the priority they would assign to a social situation, and the relative importance of these aspects towards determining priority. This question was asked separately for the relationship features and the situation cues. In both questions, participants were free to add answer options, and for each feature they ranked its importance (i.e., how much is it taken into account) in determining the priority of the situation on a 5-point Likert scale ranging from ‘Not at all important’ to ‘Extremely important’. The set of relationship features that had an importance of 3 or higher and were mentioned by at least 20% of the participants, are marked with a + in the last column of Table 1.

The relationship features that we use to model social situations are marked in bold in Table 1. We select the aspects which appear in at least two columns of the table. To that set, we add two more features, namely the age difference between the user and the other person, and whether the two have the same or different genders. Despite not being directly relationship features, age and gender appear as relevant aspects of social relationships in most research from social sciences [4,24], so we believe this warrants their addition to our model. These features are not included in Table 1, since it exclusively contains relationship features.

Table 1. Different aspects of social relationships present in the literature as well as in the exploratory pre-study. The items written in bold text form our set of social relationship features. Items marked with an asterisk are the features proposed in [15].

Relationship feature	[24]	[4]	[1]	[11]	[23]	Pre study
Role*	+	+	+	−	+	+
Contact Frequency*	+	+	+	−	+	+
Geo-distance*	+	−	+	−	−	+
Years known*	+	+	+	−	+	+
Hierarchy*	−	−	−	+	+	−
Relationship quality*	−	−	+	−	−	+
Depth of acquaintance*	+	+	−	+	−	+
Formality level*	−	−	−	+	+	−
Trust level*	−	−	−	−	−	−
Shared interests	+	+	−	−	+	+
Communication aspects	−	−	−	−	+	−
Reciprocity	−	−	−	+	−	−
Complexity	−	−	−	+	−	−

When it comes to situation cues, we use the ones proposed by Kola et al. [15], since the literature review and exploratory pre-study did not reveal new elements that warrant addition. Thus, our set of situation cues consists of: setting (work, family, sports, casual), event frequency (regular, occasional), initiator (user, other person, neither) and help dynamic (giving, receiving, neither).

Personal Values. For a list of personal values to elicit from participants, we turn to the European Social Survey [30]. It consists of a list of 18 statements (two for each universal value group - Self-direction, Stimulation, Hedonism, Achievement, Power, Security, Conformity, Benevolence and Universalism) that describe features/qualities of a person (e.g., "Thinking up new ideas and being creative is important to him/her/them. He/She/They like(s) to do things in his/her/their own original way."), where each statement represents a personal value (e.g., creativity). The subjects were asked to assess how similar they believe this person is to them, on a scale from 1 (Not like me at all) to 6 (Very much like me). The original survey consists of 21 values, however, we removed the statements of the value group "Tradition" since its values (devotion, religion) do not fit with the type of scenarios that participants were presented with. Furthermore, in the category "Security" we replaced the statement for the value National Security with the statement for the value Health for the same reason. The statement for the value Health was taken from an extended version of this survey which consists of 40 items [30].

4.2 Method

Participants. We recruited 302 subjects on the online crowd-sourcing platform Prolific Academic. Using a crowd-sourcing platform allowed us to efficiently obtain a large sample size in a short amount of time. Respondents received monetary compensation for the time they spent, as per the platform policies. After eliminating the ones who did not pass at least two of our three attention checks, our data consists of answers from 278 subjects. 149 of them are female, 127 are male, and two participants selected the option "other" when asked about their gender. The average age of the subjects is 36.2 years old (SD = 12.3).

Procedure. Subjects answered an online survey[2]. After being briefed about the purpose of the study, they were presented with its four parts. In the *first part*, subjects were asked about their relationship with five people from their social circle. The questions were the relationship features that are marked in bold in Table 1. Ideally, we wanted the subjects to select people with whom they have different types of relationships. Kola et al. [15] suggest that when left without guidance, subjects tend to select people closer to them. This, in turn, leads to less variety and a more imbalanced data set. To avoid this, we pre-determined

[2] The survey questions and the data can be found in the supplementary materials in https://doi.org/10.4121/13176923.

some of the features as follows: the first person the subject selects had to be a family member. The second person had to be one of their (current or past) direct supervisors or managers. The third person had to be someone with whom they have a negative or very negative relationship. The fourth person had to be one of their friends. The last person had to be someone that the subject does not know very well. Subjects were instructed to simply provide us with the initials of these people. This way, on one hand anonymity is preserved, and on the other hand, we could refer back to these people in the next parts of the experiment.

In the *second part*, subjects were presented with eight hypothetical social situations, which were meeting scenarios involving one of the people from the first part (selected randomly). We used hypothetical situations, since this gives us control over the types of situations subjects are presented with, ensuring a wide variety. To make the situations seem realistic, we presented subjects with activities that are common for people in their daily lives. Meeting situations were formed by combining situation cues: setting, activity within setting, event frequency, initiator, and help dynamic, as described in Sect. 4.1 (E.g. "You have a weekly work meeting with your team leader where you expect to get feedback on a project that you are working on."). Activities are not part of our situation cues, however, we included them in the description of the scenarios in order to make them more concrete. These activities were collected in the exploratory pre-study described in Sect. 4.1. The activities were grouped into settings, and for each setting, we selected the ones that were suggested more often: four for the casual setting, three for the work setting, three for the family setting, and two from the sports setting, for a total of twelve activities. We selected more activities for the casual setting and less for sports, to reflect the proportions of activities mentioned by the participants of the exploratory user study. Each subject was presented with eight of these twelve activities. Subjects were asked what priority they would assign to each situation on a 7-point Likert scale (ranging from Very Low to Very High). Furthermore, they were asked how likely they are to encounter a similar situation in their daily life on a 5-point Likert scale (ranging from Very Unlikely to Very Likely). This information is used to assess whether the hypothetical scenarios seem realistic to the subjects.

In the *third part*, subjects were presented with five pairs of situations (from the second part), and for each pair, they were asked the following question: "Suppose that in a certain week you are very busy due to some other unexpected commitment, so you can attend only some meetings and cancel some others. Which of these two meetings would you attend?". Lastly, in the *fourth part* subjects answered the survey about personal values described in Sect. 4.1.

4.3 Description of Data

In order to be able to build a model that generalizes better, it is important to have a wide variety of data. Overall, we notice that this is the case for most of the social features. The roles were represented as follows: friends - 29.5%, family members - 26,31%, supervisors/managers - 21.3%, co-workers - 8.71%, neighbours - 5.53%, members of the same group - 3.02%. Features such as geographical distance (64.8% living less than 1 h away), depth of acquaintance (mean = 3.28, SD = 1.33), frequency of contact (mean = 2.91, SD = 1.4) and formality level (mean = 2.27, SD = 1.45) follow a similar distribution to the ones reported in Kola et al. [15], so we do not report them fully for space purposes. Relationship quality was on average slightly positive (mean = 0.55, SD = 1.26, on a scale where −2 = very negative, −1 = negative, 0 = neutral, 1 = positive, 2 = very positive). Fixing its value for one of the selected people led to more balanced answers for relationship quality as compared to the ones reported in [15].

When it comes to the priority of the scenarios, subjects assigned relatively high priorities. The average priority was 5.12 (on a 7-point Likert scale), with a standard deviation of 1.96. Participants found the scenarios on average to be relatively realistic (mean = 3.02, SD = 1.5, on a 5-points Likert scale), with 47.9% of the scenarios being 'Likely' or 'Very Likely'.

In the third part of the user study, we asked subjects to specify which meeting they would attend if they had to select between two meetings. We use this data to test whether subjects mostly select meetings which have a higher priority. In 25% of the cases, subjects were presented with two meetings which have the same priority, so we cannot use this fraction of the data to test our hypothesis. This is an unintended result of the experimental setup, and in future experiments this can be controlled beforehand. For the data in which it is possible to make a distinction, subjects select the meeting with a higher priority in 58% of the cases, and the one with lower priority in 42% of the cases. This result marginally supports our research hypothesis, however, 42% remains a large figure. One potential reason can be the noise in the data caused by the fact that we present subjects with hypothetical scenarios, since some of these scenarios are situations that subjects do not normally encounter in their lives. To test this assumption, we remove the meetings which subjects consider to be 'somewhat unlikely' or 'very unlikely' in part 2 of the experiment. In the remaining data, in 68% of the cases subjects select the meeting with the higher priority. This is significantly higher than 58% (Two-Proportions Z-Test, $p < 0.05$), which suggests unlikely meetings can be a source of noise. Further reasons why some subjects select the meeting with a lower priority will be explored in future work.

When asked about personal values, subjects reported on average the following scores (on a 6-points scale): Benevolence - 4.81 (SD = 0.93), Self-direction - 4.75 (SD = 0.93), Universalism - 4.7 (SD = 1), Security - 4.49 (SD = 1), Hedonism - 4.03 (SD = 1.09), Conformity - 3.96 (SD = 1.22), Stimulation - 3.87 (SD = 1.26), Achievement - 3.78 (SD = 1.28), and Power - 3.22 (SD = 1.36).

5 Predicting Priority of Social Situations

In the following subsections, we use the data from the crowd sourcing user study to explore our research questions.

5.1 Predictive Models and Results

In this section, we present the models that we use to predict the priority of social situations, and compare their performance. Models take as input the full list of social relationship features and situation cues. Subjects could assign priorities on a scale from 1 to 7, so we model this task as a regression task, since there would be too many classes to model it as a classification task for the amount of data that we have. This means, given a set of features, the model predicts a continuous score between 1 and 7. We believe this should not pose an issue although subjects were presented with discrete answer choices, since these choices were ordinal, and the concept of priority is in itself continuous.

As mentioned in Sect. 3, we use a random forest model as well as a gradient boosting machine model. Specifically, we use the RandomForestRegressor and XGBRegressor implementations from the Scikit-learn package in Python[3]. We split the data and randomly assign 80% to the training set and 20% to the test set. We perform parameter tuning by using cross validation on the training set. We report the performance of these models on the test set. For comparison we include a decision tree model, since this approach was previously used to predict the priorities of social situations [15]. Furthermore, we include three baseline predictors based on heuristics, namely: an algorithm which always predicts the mean priority score, an algorithm which predicts a random score between 1 and 7, and an algorithm which always predicts the most chosen class (in this case, a priority of 7). Including such baseline predictors is common practice for new machine learning tasks with no predetermined benchmarks (e.g. [10]).

We start by reporting the Mean Absolute Errors, as well as the Mean Squared Errors for predictions on the test set. Results are reported in Table 2.

Table 2. Model errors in predicting the priorities of situations. Differences between predictions are statistically significant ($p < 0.05$). In bold, the best performing model.

Model	Mean absolute error	Mean Squared error
Random prediction	2.53 (SD = 1.76)	9.17
Predict most chosen class	1.84 (SD = 1.93)	7.1
Predict mean	1.56 (SD = 1.15)	3.72
Decision tree	1.81 (SD = 1.79)	6.21
Random Forest	**1.35 (SD = 1.02)**	**3.25**
XGBoost	1.43 (SD = 1.12)	3.34

As we can see from the results, the best performing model is the Random Forest model, followed by XGBoost. They outperform the Decision Tree model, as well as the baseline heuristic predictors that we used as a comparison.

[3] The code can be accessed under: https://github.com/ilir-kola/priority-social-situations.git.

In practical terms, it means our best model on average makes a prediction error of 1.35, on our 7 point scale. However, this number is just an average, so it gives limited insight into individual predictions. For this reason, we look more in detail into what does this error mean for the three best performing models from Table 2.

In general, our data set is to some extent unbalanced, since there are more situations which receive a high priority (i.e. 5, 6 or 7) compared to the ones receiving a low priority (i.e. 1, 2 or 3). In our specific domain - a personal assistant that manages the user's agenda - it is often more important to be able to distinguish a situation with a low priority from one with a high priority (or vice versa), rather than to be able to differentiate between two meetings with different degrees of high (or low) priority. This is a well-known controversy (e.g. [31]) arising from interpreting Likert scales as numeric intervals: a prediction error of 2 which confuses 'Slightly high' with 'Very high' does not have the same nuance as a prediction error of 2 which confuses 'Slightly high' with 'Slightly low', because of the change of category (from high to low) involved in the latter example. By dichotomizing our data into situations with high priority (i.e. with a priority higher than 4) and low priority (i.e. with a priority lower than 4), we can evaluate how often do the predictors assign a high priority to a situation with a low priority, as well as the other way around (similar to Type 1 and Type 2 errors). The algorithm which always predicts the mean (i.e., 5.12) always predicts a high priority, so it is always right for situations with a high priority, and always wrong for situations with a low priority. The Random Forest model and XGBoost perform equally well for high priority situations: none of them is classified to have a low priority by Random Forest, and only 2.17% of them by XGBoost. When it comes to situations with low priority, these models clearly outperform the heuristic predictor: Random Forest wrongly classifies only 30% of situations to have high priority, whereas for XGBoost the value is 29.5%.

Our results suggest that Random Forest and XGBoost outperform heuristic predictors both in absolute errors as well as when considered in the context of our application domain. Random Forest has a slight edge on XGBoost, however, the difference is not high enough so as to declare a clear winner.

5.2 Determining Important Features for Predictions

A key advantage of the machine learning models is the fact that it is possible to get insight into their decision process. This allows for the possibility to explain to the users which features led to a certain prediction, and adapt the model if needed. We use the TreeExplainer method of the SHAP package, which is based on the work of Lundberg and Lee [18].

From a global perspective, the most informative features are setting, relationship quality, age difference, and role. This means these are the features that mostly contribute to the predictions. However, when running the predictive model without the least important features (i.e., hierarchy, geographical distance, other person's gender), we notice a drop in accuracy. This suggests that all the features are to some extent important in predicting specific situations.

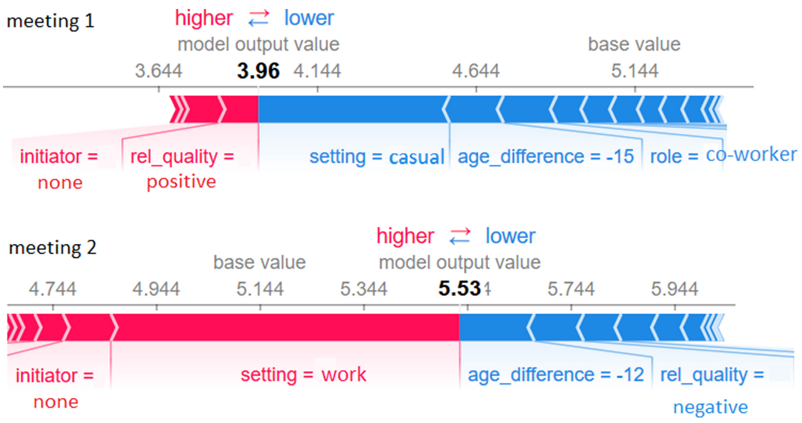

Fig. 2. Explaining the features that led to specific predictions. Larger bars have more impact on the decision. Features marked in red contributed to making the priority prediction higher, whereas the ones in blue lower. The text under the bar indicates the values of each feature for the specific situations. (Color figure online)

To illustrate the interpretations of individual predictions we use two specific social situations which our model had to predict (Fig. 2). In both situations, setting is the feature with the highest influence. We notice in this example that the work setting causes the meeting to have a higher priority, whereas the casual setting contributes to a lower one. As expected, a positive relationship quality makes the priority of the meeting higher, as opposed to the negative relationship quality. In both cases, meeting a younger person contributes to a lower priority.

This method allows for insight into the decision process of the agent, and can form a basis towards explaining the suggestion to the user. Miller [19] proposes that explanations in AI should be contrastive: people want to know why the agent suggested a certain action rather than another one. This is inherently part of our method, since the agent can explain to the user why one meeting was selected instead of another. Furthermore, people prefer an explanation that consists of a few causes rather than many. Using the SHAP package allows this, since it identifies the features with the highest impact.

5.3 Role of Personal Values in Predicting Priorities

We start our work by building a predictive model using data from multiple people, however, we want to explore whether it is possible to have some degree of personalization for the user. To achieve this, we explore whether adding information on the personal values of the users helps to increase the accuracy of the model. The underlying assumption is that users with similar value preferences will assign similar priorities to situations. This is based on the definition of values, which are considered to be drivers of behaviour. First, we train our Random Forest model with the original set of features, as well as 9 new features representing the score that the user assigned to each of value groups (Sect. 4.1), collected in the last part of the user study. The mean absolute error, in this case, is 1.38. This means that the quality of the predictions slightly deteriorates when adding information about values. One reason for this might be that adding 9 new features to the existing ones causes the model to have too many features, which can deteriorate performance. Another possible reason can be related to the salience of personal values in different situations. Schwartz [29] argues that in order for values to influence action not only should they be important to the actor, but they should also be relevant in that specific context. Kayal et al. also propose the use of domain values in order to reason about social commitments [13]. We check for this insight in our data. Some situations explicitly mention that the user is expected to help someone. Therefore, the value 'helpfulness' is salient in these situations. We notice that subjects who value helpfulness more, on average assign a significantly higher priority to situations where they have to help someone, as compared to subjects who value helpfulness less (6 vs. 5.09, $p < 0.01$ when performing the Mann-Whitney test). For meetings that do not involve giving help, the differences in the priorities assigned by these subjects are not significant. This suggests that certain values which are salient to the domain can potentially help predictions.

6 Conclusions

6.1 Contributions

In this work, we propose an approach which enables personal assistant agents to predict the priority of a user's social situation. The approach relies on concepts from social sciences which are used to model social situations, as well as machine learning techniques which are used to learn the priority scores from data from multiple people. First, we review literature from social sciences and propose a set of features which we use to model the social situations of a user. Then, we conduct a crowd-sourcing user study in order to gather the data needed to build our predictive models and evaluate our approach. The subjects' answers suggest that having a numerical representation of priority can in principle be used to help deciding which meeting to attend in cases of overlapping meetings. The results marginally supports our hypothesis (RH): 58% of the subjects select the meeting with a higher priority. This can form the basis for allowing a personal assistant agent to use its priority predictions to choose between the meetings.

Next, we show that ensemble models such as Random Forests outperform baseline models in predicting the priorities of social situations, especially when it comes to differentiating between situations with high and low priorities (RQ1). Furthermore, we present a procedure which enables the personal assistant agent to determine the features that contributed to the predictions, which in future work will be presented as explanations to the user. We envision that this, together with the fact that features are taken from social science literature and are therefore more understandable for people, can help achieving the vision for more transparent and intelligible personal assistant agents. Lastly, we test whether adding information about the personal values of the user can help us lower the prediction error (RQ2). Results show that in our setting this is not the case, since the mean absolute error of the model suffers a slight increase. However, insights from the data suggest that using personal values which are salient in the specific situation has the potential to be a more successful approach.

6.2 Limitations and Future Work

First of all, our experimental setup presented subjects with hypothetical scenarios. This was done to ensure variety in the data, however, this comes at the cost of the data being noisier, since some of the scenarios might be unlikely to actually occur, so the subjects might not answer consistently. It would be useful to conduct a user study in which subjects report all their social situations from a fixed period of time, in order to evaluate our models with more realistic data. Furthermore, asking subjects which meeting they would attend when two meetings overlap (part 3 of the user study) presented them with a binary choice, which does not inform us how certain they were about their selection. An alternative would be to provide participants with a slider, where they can state how inclined they would be to attend one of the meetings [13]. In future work, we aim to enable personal assistant agents to provide full explanations regarding their decisions to the user. This is based on our assumption that presenting the user with the social features that contributed to a prediction makes the work of the personal assistant agent more transparent. This has to be tested in practice. Next, we will investigate the possibility of a feedback loop between the user and the agent based on the explanations, in order to further personalize support. This would be important especially for cases where the subjects disagree with the agent's decisions. Lastly, we will explore whether our models can be used to predict other aspects of social situations other than priority.

References

1. Antonucci, T.C., Akiyama, H.: Social networks in adult life and a preliminary examination of the convoy model. J. Gerontol. **42**(5), 519–527 (1987)
2. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
3. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
4. Burt, R.S.: Network items and the general social survey. Soc. Networks **6**(4), 293–339 (1984)
5. Chevaleyre, Y., Koriche, F., Lang, J., Mengin, J., Zanuttini, B.: Learning ordinal preferences on multiattribute domains: The case of CP-NETs. In: Fürnkranz, J., Hüllermeier, E. (eds.) Preference Learning, pp. 273–296. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14125-6_13
6. Dietterich, T.G.: Ensemble methods in machine learning. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 1–15. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45014-9_1
7. Dignum, F.: Interactions as social practices: towards a formalization. arXiv preprint arXiv:1809.08751 (2018)

8. Friedman, B., Kahn, P.H., Borning, A., Huldtgren, A.: Value sensitive design and information systems. In: Doorn, N., Schuurbiers, D., van de Poel, I., Gorman, M.E. (eds.) Early engagement and new technologies: Opening up the laboratory. PET, vol. 16, pp. 55–95. Springer, Dordrecht (2013). https://doi.org/10.1007/978-94-007-7844-3_4
9. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of statistics, pp. 1189–1232 (2001)
10. Gu, S., Kelly, B., Xiu, D.: Empirical asset pricing via machine learning. Technical report, National Bureau of Economic Research (2018)
11. Heaney, C.A., Israel, B.A.: Social networks and social support. Health Behav. Health Educ. Theory Res. Pract. **4**, 189–210 (2008)
12. Kaminka, G.A.: Curing robot autism: a challenge. In: Proceedings of the 2013 International Conference on AAMAS, pp. 801–804. IFAMAAS (2013)
13. Kayal, A., Brinkman, W.P., Neerincx, M.A., Riemsdijk, M.B.V.: Automatic resolution of normative conflicts in supportive technology based on user values. ACM Trans. Internet Technol. (TOIT) **18**(4), 1–21 (2018)
14. Kepuska, V., Bohouta, G.: Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference, pp. 99–103. IEEE (2018)
15. Kola, I., Jonker, C.M., van Riemsdijk, M.B.: Who's that? - social situation awareness for behaviour support agents. In: Dennis, L.A., Bordini, R.H., Lespérance, Y. (eds.) EMAS 2019. LNCS (LNAI), vol. 12058, pp. 127–151. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51417-4_7
16. Lewin, K.: Field theory and experiment in social psychology: concepts and methods. Am. J. Sociol. **44**(6), 868–896 (1939)
17. Lim, B.Y., Dey, A.K., Avrahami, D.: Why and why not explanations improve the intelligibility of context-aware intelligent systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2119–2128 (2009)
18. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, pp. 4765–4774 (2017)
19. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. **267**, 1–38 (2019)
20. Murukannaiah, P.K., Singh, M.P.: Platys social: relating shared places and private social circles. IEEE Internet Comput. **16**(3), 53–59 (2012)
21. Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D.L., Morley, D., Pfeffer, A., Pollack, M., Tambe, M.: An intelligent personal assistant for task and time management. AI Mag. **28**(2), 47–47 (2007)
22. Neerincx, M.A., van der Waa, J., Kaptein, F., van Diggelen, J.: Using perceptual and cognitive explanations for enhanced human-agent team performance. In: Harris, D. (ed.) EPCE 2018. LNCS (LNAI), vol. 10906, pp. 204–214. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91122-9_18
23. Pabjan, B.: Measuring the social relations: social distance in social structure - a study of prison community. Acta Physica Polonica Series B **36**(8), 2559 (2005)
24. Phillips, S.L., Fischer, C.S.: Measuring social support networks in general populations. Stressful life events and their contexts, pp. 223–233 (1981)
25. Pinder, C., Vermeulen, J., Cowan, B.R., Beale, R.: Digital behaviour change interventions to break and form habits. ACM Trans. Comput. Hum. Inter. (TOCHI) **25**(3), 15 (2018)
26. Polikar, R.: Ensemble based systems in decision making. IEEE Circuits Syst. Mag. **6**(3), 21–45 (2006)

27. Rokeach, M.: The Nature of Human Values. Free Press, New York (1973)
28. Sagi, O., Rokach, L.: Ensemble learning: a survey. Wiley Interdisc. Rev. Data Mining Knowl. Discov. **8**(4), e1249 (2018)
29. Schwartz, S.H.: Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. Adv. Exp. Soc. Psychol. **25**(1), 1–65 (1992)
30. Schwartz, S.H.: Human values. European Social Survey Education Net (2005)
31. Sullivan, G.M., Artino Jr., A.R.: Analyzing and interpreting data from likert-type scales. J. Graduate Med. Educ. **5**(4), 541–542 (2013)
32. Van Riemsdijk, M.B., Jonker, C.M., Lesser, V.: Creating socially adaptive electronic partners: interaction, reasoning and ethical challenges. In: Proceedings of the 2015 International Conference on AAMAS, pp. 1201–1206. IFAMAAS (2015)