AGENT BASED NEGOTIATION FOR INCENTIVE DRIVEN PRIVACY PRESERVING INFORMATION SHARING

A Thesis

by

YOUSEF RAZEGHI

Submitted to the
Graduate School of Sciences and Engineering
In Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in the
Department of Computer Science

Özyeğin University
August 2019

AGENT BASED NEGOTIATION FOR INCENTIVE DRIVEN PRIVACY PRESERVING INFORMATION SHARING

Approved by:

Assistant Professor Reyhan Aydoğan, Advisor
Department of Computer Science
Özyeğin University

Associate Professor Hasan Sözer
Department of Computer Science
Özyeğin University

Assistant Professor Ali Haydar Özer
Department of Computer Engineering
Marmara University

Date Approved: 20 August 2019

To my family.

ABSTRACT

While customizing their services, companies usually use their users' data. According to the new regularization, it is required to get the permission of their users to be able to store and share their users' private data. The current approaches rely on requesting access rights by providing some incentives. The customers can only accept or reject the possible incentive offered by the companies exchange for giving access rights. This thesis introduces an agent-based, incentive-driven, and privacy-preserving information sharing framework. One of the main contributions of this thesis is to give the data provider agent an active role in the information sharing process and to change the currently asymmetric position between the provider and the requester of data and information (DI) to the favor of the DI provider. Instead of a binary yes/no answer to the requester's data request and the incentive offer, the provider may negotiate about excluding from the requested DI bundle certain pieces of DI with high privacy value, and/or ask for a different type of incentive. We show the presented approach on a use case and conduct a user experiment. Questionnaire responses showed that participants like the idea of negotiation on their information sharing policies with the companies.

Furthermore, this thesis proposes an acceptance strategy using deep reinforcement learning for automated negotiating agents. In the automated negotiation literature, most of the acceptance strategies are based on some predefined rules. In contrast, this thesis proposes to use reinforcement learning in order to learn when to accept opponent's offer. Our experimental evaluation shows that the developed acceptance strategy performed as well as AC-Next acceptance strategy.

ÖZETÇE

Şirketler, hizmetlerini özelleştirirken, genellikle kullanıcıların verilerini kullanmaktadırlar. Yeni düzenlemeye göre, kullanıcıların özel verilerini depolamak ve paylaşmak için kullanıcıların iznini almak gerekmektedir. Mevcut yaklaşımlar, bazı teşvikler sağlayarak erişim hakları talep etmeye dayanmaktadır. Müşteriler, şirketlerin sundukları teşviklere karşılık erişim haklarını talep ettiklerinde cevap olarak sadece olumlu ya da olumsuz olarak dönüş yapabilmektedirler. Bu tez, etmen temelli, teşvik odaklı ve mahremiyete dayalı bir bilgi paylaşım sistemi sunmaktadır. Bu tezin ana katkılarından biri, veri sağlayıcısına bilgi paylaşım sürecinde aktif bir rol vermek ve sağlayıcı ile veri kullanıcısı arasındaki mevcut asimetrik pozisyonu veri sağlayıcısı lehine değiştirmesidir.

Veri kullanıcısının veri talebine ve teşvik teklifine evet / hayır cevab şekli yerine, sağlayıcı, talep edilen veri paketinden, yüksek mahremiyet değeri olan bazı verileri tekliften çıkararak pazarlık yapabilir veya farklı bir teşvik türü talep edebilir. Sunulan yaklaşımı bir kullanım senaryosu üzerinde gösterip kullanıcı deneyi yapılmıştır. Anket cevapları, katılımcıların, bilgi paylaşım politikalari hakkında firmalarla müzakere etme fikrini tastiklediklerini göstermiştir.

Ayrıca, bu tez, otomatik müzakere yapan etmenler için derin pekiştirmeli öğrenmeyi kullanan bir kabül stratejisi önermektedir. Otomatik müzakere literatüründe, kabül stratejilerinin çoğu önceden tanımlanmış bazı kurallara dayanmaktadır. Farklı olarak, bu tez, rakibin teklifini ne zaman kabül edeceğini öğrenmek için pekiştirmeli öğrenmeyi kullanmayı önermektedir. Deneysel değerlendirmemiz, geliştirilen kabül stratejisinin, AC-Next kabül stratejisi kadar iyi performans sağladığını göstermektedir.

ACKNOWLEDGEMENTS

> "Success is really nothing more than the progressive realization of a worthy ideal. This means that any person who knows what he is doing and where he is going is a success. Any person with a goal towards which he is working is a successful person."
>
> – Earl Nightingale

I want to express my gratitude to the people who helped me during my Master studies. I appreciate their consideration and attitude toward me. First of all I want to mention that I admire the founder of Özyeğin University, Mr. Hüsnü Özyeğin. In my opinion a great man who has created a wonderful place to learn and develop ourselves deserves the best thanksgiving. To be honest Özyeğin University not only changed my vision but also changed my life and destiny, I will not ever forget the time that I spent in this university. To my family, thank you for encouraging me in all of my pursuits and inspiring me to follow my dreams. I am especially grateful to my parents, who supported me emotionally and financially. I always knew that they believed in me and wanted the best for me.

I would like to give thanks to my advisor Assist.Prof. Reyhan Aydoğan for her persistent and precise support of my Master study. Her accuracy and patience in research inspired me so much, I got a strong motivation in each meeting we had. She was so supportive and diligent in guiding me through research problems, under her supervision I had the chance of meeting and collaborating with great people. I feel very fortunate to have such a great advisor during my Master study.

I am so grateful of another great person who I want to give my thanks to her, the Assoc.Prof. Pınar Öztürk from Norwegian University of Science and Technology. My advisor introduced me to Prof. Öztürk during her sabbatical visit at Özyeğin University and I had the opportunity to have a collaboration with them in preparing a paper on the issue of privacy preserving information sharing systems. This collaboration mainly shaped the outline for first part of my thesis and research focus.
Along with my advisor and Prof. Öztürk, I truly appreciate the consideration of my thesis committee: Assoc.Prof. Hasan Sözer and Assist.Prof. Ali Haydar Özer.
I am also thankful to Assoc.Prof. Murat Şensoy, I have learned a lot from him in various courses such as Introduction to Deep Learning, Semantic Web Engineering, Artificial Intelligence, and Data Science with Python. I appreciate his attitude toward students which is always motivating them to learn more and study precisely. He was very supportive and inspiring to me during my Master study.
I would like to extend my thanks to Assist.Prof. Melih Kandemir for his valuable time and feedbacks during my research on applying reinforcement learning for negotiation.
I extend my thanks to M.Sc. student Ozan Yavuz for his collaboration with me in our research project on applying reinforcement learning for negotiation, he is a very smart student and I am sure he will achieve his goals in future.
I extend my thanks to the administrative staff, Ms. Aylin Muhaddisoğlu, Ms. Gizem Bakır, and Ms. Gizem Tekin Akbaş at the Graduate School of Engineering and Science at Özyeğin University. I truly appreciate the time and effort they invested on the issues related to my education and graduation.

Finally I appreciate the valuable time and effort of people who participated in the research experiment which is presented in this thesis.

Contents

List of Tables

List of Figures

Chapter I

INTRODUCTION

> "If we knew what it was we were doing, it would not be called research, would it?"
>
> – Albert Einstein

In order to increase their sales and establish long-term strong relations with their customers, companies try to customize their services and consequently aim to provide well-targeted services. Customization of services requires understanding customers' behavior patterns (e.g., their consumption habits) and also incorporating their personal needs. To carry out service customization, they need to record and use their customers' data regarding their interaction with their system as well as some personal data (e.g., age, gender etc.). Although personalization has a reciprocal benefit for both companies and humans, it often causes privacy violation. Consumer behavior research reports that people are reluctant to share their personal DI [1] due to fear for possible illegal and unethical usage of their DI [1, 2].

Sometimes people can anticipate what the possible uses and harms may be while other times they don't even know for what *purposes* their DI may be used. This is the type of uncertainty called as "unknown unknowns" by the US Secretary of Defense Donald Rumsfeld [2], or "Knightean uncertainty" [2] in economics. This type of uncertainty is also one of the main reasons for stakeholders' inclination not to reveal their data. It has been stated that a person's willingness to share his/her data [3] is determined by the risk belief and the enticement beliefs. *Privacy calculus* is a validated theorem that studies

[1] From now on we use DI to refer to "data and information" but also interchangeably use only "data" or only "information" as well.

[2] http://papers.rumsfeld.com/about/page/authors-note

the agents which influence a person's decision and how these agents interact with each other [4, 5]. Approaches based on the privacy calculus theorem commonly rely on the dissuasive techniques in order to prevent data disclosure [4, 3, 6]. They mainly act upon user answers to prepared questionnaires. Humans can show different behaviour than the responses they gave to questionnaires which is called as the *privacy paradox* [7, 8, 9].

Existing information sharing practices deployed by the companies do not involve the specific constructs and the process necessary for dealing with differences between individuals. For instance, a data requester company proposes the same incentives to everybody regardless of individuals' preferences and beliefs. Another main downside of the information acquisition practice employed by the companies is that the data provider can only accept or reject the possible incentive offered by the data requester. To have to give only a binary option type of answer (i.e.,accept/reject) for giving to data requesters, access right to their data may create uncertainty and hence reluctance on data provider's side. Being in a passive situation where the data provider is involved neither in the selection of the data pieces to disclose nor the amount of the incentive blocks the information disclosure.

To deal with the aforementioned issues with the existing information sharing approaches, this master thesis proposes a negotiation framework in which data providers and data consumers can negotiate on what to to-be disclosed/shared with the data consumer, under what conditions, and what type of incentive to be provided by the data consumer. The proposed approach inherently takes into consideration that people vary in what they consider as secret and risky data and how they value it. Furthermore, a preference elicitation tool is developed for the data provider and consumer to elicit their preferences. The proposed negotiation framework is implemented for illustrating the proposed idea on an example from Telecommunication domain. On the chosen scenario, a user experiment in which an automated negotiating agent on behalf of a telecommunication company is negotiating with a human playing the role of customer, has been conducted and analyzed. The analysis of this experiment showed that human participants preferred the idea of negotiating for information sharing policies than giving only a binary option type of answer (i.e.,accept/reject) to company's requests on sharing their information. It is observed that people are sensitive about their GPS and call logs while they are easy to reveal their age, marital status and educational level. In the user experiments, most of the participants reached an agreement before the given deadline - less than 10 minutes where the deadline is 15 minutes.

Apart from designing and developing a negotiation framework for information sharing, this thesis also pursue the research question of learning when to accept other side's offer. Accordingly, this thesis proposes to use deep reinforcement learning to learn when to accept. In the literature, acceptance strategies are mostly relying on predefined rules taking into account number of factors such as remaining time and the utility of the offers. Our proposed strategy on the other hand aims to learn whether to accept its opponent's offer or to make a counter offer by reinforcement signals received after performing an action. Thus, instead of applying predefined acceptance conditions, our agent uses epsilon greedy action selection and learns its own acceptance strategy and improves it over time. We compare the performance of this acceptance strategy with the AC-Next acceptance strategy. The evaluation results show that our agent can perform as well as this state-of-the-art acceptance strategy.

Chapter II

PROPOSED FRAMEWORK

"The only way to do great work is to love what you do. If you haven't found it yet, keep looking. Don't settle."

– Steve Jobs

Based on our previous work [10], we propose an information sharing platform where data providers (e.g. customers) and data consumers (e.g. companies) negotiate on what type of information to be shared, with whom and for how long the requested information should be kept, and what kind of incentives will be provided for individuals in exchange for sharing their data. In this work, we focus on bilateral negotiation between a company agent and a data producer agent (i.e., individual) as demonstrated in Figure 1.

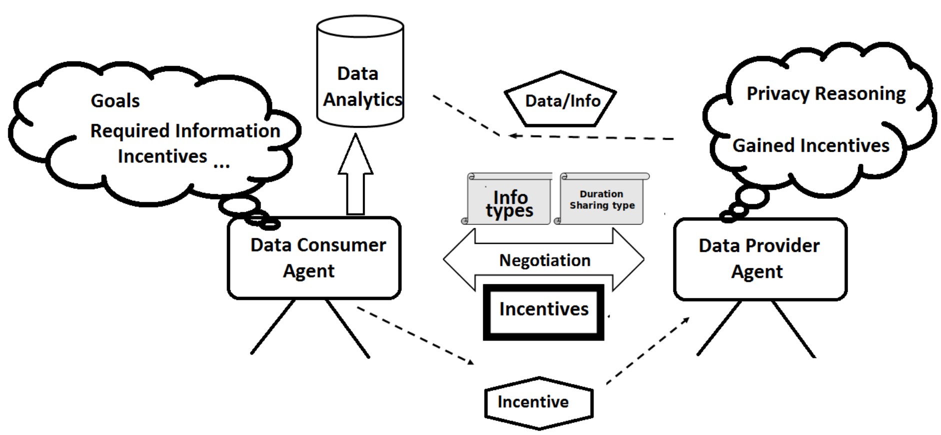


Figure 1: Proposed negotiation framework

In the proposed framework, both agents have their private knowledge base so that they can reason on their stakeholders preferences and beliefs while making their offers or deciding whether or not to accept the given offer. Additionally, the company agents have a database comprising the information gathered from various individuals. This database is often used for data analytic to infer knowledge for revising their business models or to create new models and revenues.

There is a shared ontology representing the types of information (e.g. GPS, age, etc), possible permission duration (e.g. six months, one year, and so on), possible sharing options such as "company only" or "company and third party" as well as the possible types of incentives (e.g., free SMS/Internet in the telecommunication use case in section 3) relevant for a specific domain. Hence, the content of this ontology is domain-specific for providing common understanding between agents.

The negotiation between the company agent and the individual is governed by the Alternating Offers Protocol [11, 12]. The company initiates the negotiation with an offer and the individual may accept this offer or make a counter offer. This process continues in a turn-taking fashion till the termination condition is met such as reaching a mutual agreement or deadline.

In this negotiation, a bid[1] structure can be formalized as follow: $o < I, d, s, p >$ where

- I denotes a set of information/data types under trade-off.
- d stands for duration of the contract (i.e. how long the data will be accessed or shared with the data provider) where $d \in D$ and D is the set of all possible predefined duration options.
- s denotes the sharing policy; particularly it specifies with whom the data will be shared where $s \in S$ and S is the set of all possible stakeholders.
- p denotes the incentives/promotions to be given for the shared information and $p \in P$ and P is the set of all possible promotions defined in the system.

In the shared ontology, the set of all possible information types, T is defined formally

[1]In this paper, the words of bid and offer are used interchangeably

and $I \subseteq T$. As mentioned before, $Agent_C$, the data consumer initiates the negotiation with an offer such as o_1 =< {Age, Birth date}, "1 year", "shared with only company", "3-month free phone call" >. $Agent_P$ needs to evaluate this offer to make its decision, either to accept or to make a counter offer.

The proposed information sharing approach is founded on the following two constructs. First, a data consumer (e.g. a company) must have specific *goals* and a purpose for wanting to access the concerned DI. **Utility of information** for the data consumer depends on how much this DI is needed to achieve, for example, a company's business goal(s). Second, a data provider (e.g., customer or individual as referred before) must have a motivation **for sharing her personal/private DI**. In the following part, we explain how the agents make their evaluation.

2.1 Data Consumer Agent's Reasoning

In automated negotiation, agents mostly assess the underlying bids according to their utilities. By following this paradigm, we introduce an expected utility function for the data consumers based on to what extent their goals are satisfied with the given DI and for what costs. For this purpose, the data consumers should model the *goals of their owners* and map each goal with a set of *DI types* required to achieve that goal. The knowledge base of the consumer agent contains those goals (e.g., business goals if a company) and the map of these goals with the type of DI necessary to achieve these goals as well as the importance of each goal for the data consumer agent. Furthermore, the knowledge base should keep the costs of incentives to be provided the data provider in order to persuade her/him to disclose that DI.

For the data consumer agent, the utility of an offer depends on two values: *the value of the bundle of information types within the given conditions (i.e, duration and to whom to be shared*) and *the value of the cost of the incentives.* Note that the range of those value functions should be the same. Therefore, the expected utility of an offer for the data consumer can be estimated as follows:

$$\mathscr{E}\mathscr{U}(o < I, d, s, p >) = Value_{Info}(I, d, s) - Value_{cost}(p) \tag{1}$$

When the data consumer agent ($Agent_C$) generates its offer, it ensures that the content of the information bundle, I, is sufficient for achieving its targeted goals. Similarly, when it evaluates the data provider's counter offer, it first checks whether the proposed bundle meets its goals. This is because the provider ($Agent_P$) may have excluded some contents of the information because of the privacy reasons.

The set of $Agent_C$'s goals are represented as $G = \{g_1, g_2, ...g_k\}$ where k denotes the total number of goals. A goal $g \in G$ relies on some information to be achieved. For instance, $i_1, i_2, i_3 \in I$ are necessary for achieving g_1 while achieving g_2 may require only i_3. $Agent_C$ aims to obtain this information from $Agent_P$ through the negotiation. On the opposite side, if the parts of the information has high privacy value for $Agent_P$, then, $Agent_P$ may try to avoid to provide this data naturally or ask for more incentives.

As it is usually the case, each goal may have different importance level for $Agent_C$. Therefore, a weight value, m_i is associated to each goal $g_i \in G$, denoting the importance of that goal for $Agent_C$. The sum of the goal weights is equal to one; $\sum_i^k m_i = 1$. In this context, a goal is considered as *satisfiable* if $Agent_C$ has access to the information required for satisfying that goal.

We define the value of information bundle $Value_{info}(I, d, s)$ as an additive function where U_I denotes the utility of given information bundle, U_D denotes the utility of duration - how long the given data would be available/shared, U_S denotes the utility of the sharing policy (i.e. with whom to share with). The value of utilities ranges between zero and one. The importance of each issue is denoted by w and the sum of the weights should be equal to 1. Note that it is assumed that there is no preferential dependencies among those issues. That is, irrespective of the content of the information bundle, the data consumer would always prefer longer duration and more comprehensive sharing policy over shorter duration and more limited sharing policy. Someone may think that the magnitude of those evaluation values (i.e. utilities) may be different depending on the information bundle. For instance, the utility of "sharing with both company and a third-party" might be higher for a particular set of information types than the utility of that option for another set of information types due to the business requirements. In such a case, generalized additive utility functions can be used where the utility of each issue value needs to be elicited for each possible information bundle combinations.

$$Value_{info}(I,d,s) = U_I(I) \times w_I + U_D(d) \times w_D + U_S(s) \times w_S \tag{2}$$

The utility of information bundle is defined $U_I(I)$ as the weighted sum of the satisfiable goals with I.

$$U_I(I) = \sum_{i}^{k} m_i \times Satisfiable(g_i, I) \tag{3}$$

where $Satisfiable(g_i, I)$=1 if I comprises all data that g_i requires; otherwise, $Satisfiable(g_i, I)$=0

Note that this formulation assigns a value to the entire bundle and does not consider each data item constituting the bundle separately. The rationale behind this is that a specific data item may be worthless without having other one(s). For example, consider that $Agent_C$ needs both i_1 and i_2 for achieving g_1 meaning that the lack of either i_1 or i_2 jeopardizes g_1. However, we conceive that more refined methods are needed to handle the data interdependency in a more sophisticated way.

Recall that an offer consists of four components: information bundle, duration, sharing policy and incentives offered to the data provider agent. The incentive incurs a cost. Without doubt, $Agent_C$ aims to minimize its cost. The value of the cost of the incentives for $Agent_C$ is a function, which maps the cost of the incentives for the obtained information types I to a real value between zero and one [0, 1]. High value means that it is highly costly to provide the chosen incentive in exchange for the information bundle. Note that the value of cost would be less than one if the cost of the incentive to be provided by $Agent_C$ is less important for the $Agent_C$ than the value of information to be provided by $Agent_P$.

2.2 Data Provider's Reasoning

People show significant differences regarding which type of DI[2] has high secrecy level, perception of the risk of sharing certain DI, as well as how they value their private DI. Certain types of DI are considered secret by everybody, such as personal id numbers while there are many differences across individuals (or individual companies) regarding secrecy of other types of DI. Similarly, certain DI may be perceived as bearing high threat for privacy breaches by some people while others may feel quite relaxed about the same DI. Hence, people value their DI differently. To sum up, an effective information sharing approach should be sensitive to individuals' peculiarities.

We define the notion of privacy in terms of two components. The first is **desire for secrecy** and captures that the DI owner may be reluctant to share a certain pieces of DI content just because she likes to keep it for herself. An example is that a person may not want others to see her falling down from a horse. Sharing a video record of this event would not lead to any harm but would make her uncomfortable. The second component relates to the **risk/fear of harm** and uncertainty about possible unethical and improper usage of DI by others without her consent. These together determine the **privacy value** attained by the data provider/owner to a certain DI content. Therefore, the knowledge base of the data provider agent has its preferences and beliefs about the secrecy and risk value of the knowledge in its knowledge base.

During the negotiation, the data provider agent takes into consideration both the utility of the incentive provided by $Agent_C$ and the level of privacy violation incurred while sharing its personal/private information requested by $Agent_C$. Accordingly, equation 4 shows how $Agent_P$ estimates the expected utility of a given bid where c_p,c_d and

[2]Note the distinction between DI and DI types where the former refers to the data itself while the latter is about the type of data, where 'age' is a DI type and '68' is a corresponding piece of DI.

c_s are the coefficients for expected trade-off between incentive and privacy violation, duration and sharing policy respectively. Note that the sum of those coefficients is equal to one for normalization purposes. We consider that $Agent_P$ has a utility function, $Value_{Incentive}(p)$, which maps each potential incentive to a real value $[0,1]$ according to its user's needs or interests.

$$\mathscr{E}\mathscr{U}(o < I,d,s,p >) = c_p \times [Value_{Incentive}(p) - Value_{privacy}(I)] + c_d \times U_D(d) + c_s \times U_S(s)) \tag{4}$$

While estimating $Value_{privacy}$, the value of privacy violation, $Agent_P$ considers level of secrecy of the information and how risky (i.e., harmful consequences) is to share the requested information. Secrecy has a psychological aspect and has to do with a person's preference to keep a certain personal information for herself, independently from whether it may be used against herself. As such secrecy is rather individual. In the proposed framework, the level of secrecy for each information type will be elicited from the data provider. For instance, the data provider might be more reluctant to share the identity number than the phone number, $SL(identitynumber) > SL(phonenumber)$. Higher values denote more reluctant the provider is to share the underlying information.

In the proposed framework, the provider also specifies $Risk(x)$ denoting how risky to share each information item $x \in T$ from the view point of itself. Considering the risk of each information type separately is a straightforward approach; however, the level of the risk may depend on what data types are shared together. Therefore, some providers may asses the risk level of each information type and their combination differently. While sharing particular data itself may not be risky but sharing it with other data may reveal sensitive information for them. For example, one provider may assign a low level of risk to “occupation” and a medium level of risk to “GPS”; but a high risk if we share them together since she may think that the data requester may find out the company she works for. For this reason, the framework enables the providers to specify the sharing risk for a subset of information types (i.e., $Risk_{DP}(Y)$ where $Y \subset T$) as well as for each information types separately (i.e., $Risk(x)\ \forall x \in T$).

Accordingly, Equation 5 shows how the value of privacy is estimated in our framework where $Risk(x)$ denotes how risky to share the requested item x from the point of data provider, $Risk_{DP}(Y)$ denotes the risk of sharing a subset of information types, $Y \subset T$, and SL represents the normalized secrecy level of the given information type. Note that DP is a set of subset of information types - denoting the risk information dependencies. For instance, if a data provider thinks that it is more risky to share both "GPS" and "occupation " than sharing them individually, she defines a dependency such as $\{GPS, occupation\}$.

$$Value_{privacy}(I, d, s) = max(\max_{x \in I}(SL(x) * Risk(x)), \max_{Y \subseteq I \wedge Y \in DP \wedge z \in Y}(SL(z) * Risk_{DP}(Y))) \tag{5}$$

In our formulation, we choose to take the value of maximum privacy violation instead of taking the average privacy violation of each information in the given bundle. This is because the bundle may consist of information types whose privacy violation might be very high and very low, then the average may not capture how significant the violation is accurately.

2.3 Negotiation Strategy

Any utility based negotiation strategy can be employed by the agent. In this section, the bidding and acceptance strategies used in our experiments are described as follows.

- **Bidding function**: As bidding strategy, we pick the stochastic time-based concession strategy [13]. According to this strategy, negotiating agent has two functions for determining the *upper boundary* and the lower boundary for target utility. The bidding mechanism works in a way such that at the beginning of the negotiation all of the possible bids are created and then are sorted based on their expected utility values. Based on the ratio of remaining time to the deadline, the functions calculates the lower and upper boundaries for the target utility. The agent is supposed to make an offer whose utility is between lower and upper target utility. Our agent chooses one of the candidate offers whose utility is between lower and upper boundary.

Equation 6 presents the formula for determining the lower and upper boundary where r denotes the scaled remaining time $r \in [0,1]$ and P_0, P_1, P_2 are the maximum value, the curvature of the curve, and minimum value respectively.

$$TU(t) = (1-r)^2 \times P_0 + 2 \times (1-r) \times r \times P_1 + r^2 \times P_2 \tag{6}$$

Figure 2 depicts our bidding functions used in our experiments. For the lower bound we consider 0.7, 0.94, and 0.5 as coefficients. While 1.0, 0.94, and 0.94 are considered as coefficients for the upper bound. Note that the deadline is set to 15 minutes.

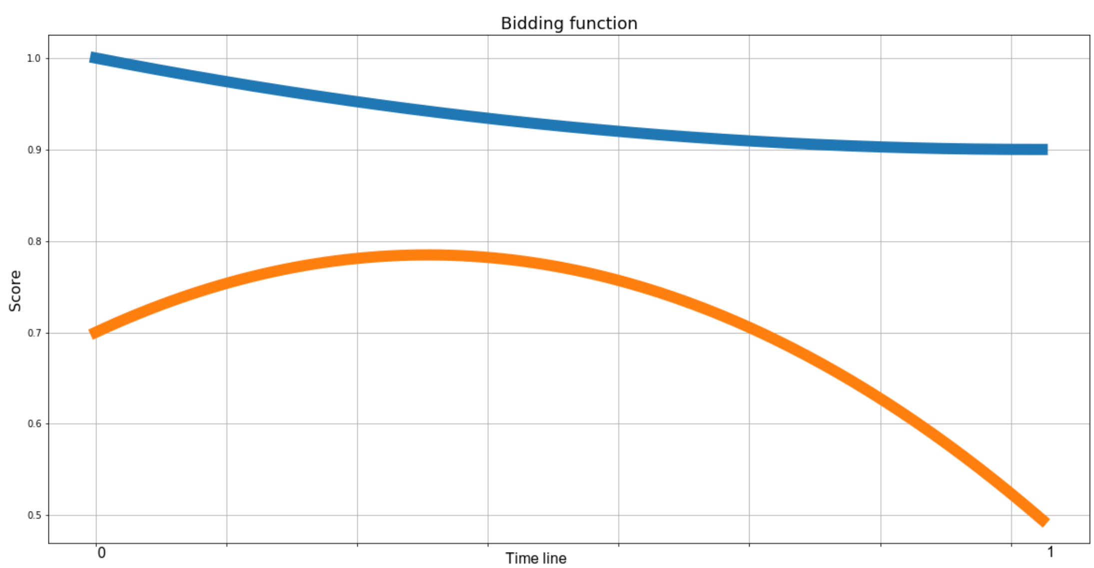


Figure 2: Bidding functions

- **Acceptance strategy**: As an acceptance strategy, we adopt the AC_{next} [14] acceptance strategy. Following this strategy, the agent accepts its opponent's offer where the utility of that offer is greater than or equal to the target utility of its next offer.

Chapter III

CASE STUDY AND EVALUATION

> "Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."
>
> – Marie Curie

As a use case, we consider a telecommunication company, which aims to do some data analytics on their customer's data in order to gauge customer needs and satisfaction better, and accordingly to provide more targeted services/products for their customers. According to the laws, they need to ask for their customers' permission to store and use their personal/sensitive data. In order to get their customers' permission, they may offer some incentives such as "1GB Fee Internet", "100 SMS for one month", and so on. A customer may accept this offer or reject it. When the customer rejects to give permission to the company regarding his/her personal data, the conversation ends in most of the cases.

However, we suggest a more interactive way of information sharing for such kind of scenarios. That is, the company (i.e., data consumer) may initiate a negotiation process with their customers (i.e. data providers) in a bilateral fashion and they together decide what to share for how long and with whom as well as the incentive to drive sharing. In order to develop such a mechanism, we first need to define the types of information of interest, possible duration values (e.g. one year, three years etc), stakeholders for sharing policy (e.g. only company, third party) and the kind of incentives the company may provide in exchange for the requested information types. Afterwards, the company and their customers should be able to express their preferences as explained in the previous chapter.

In our setup, a human customer negotiates with an agent representing the telecommunication company. In the following part, we first introduce the software tool we developed for preference elicitation and negotiation. Afterwards, we describe our experimental set up and present the experimental results elaborately.

3.1 *Preference Elicitation and Negotiation tool*

Similar to other negotiation frameworks such as *Genius* [15] and *Pocket Negotiator* [16], this framework provides stakeholders an interface to describe the underlying negotiation domain (i,e., negotiation issues and their possible values) and to express their preferences over those alternatives.

3.1.1 Domain Specification

In our scenario, we have four issues: the bundle of information types (i.e. the set of all possible information types under negotiation), promotion to be provided by the data consumer, how long the data to be shared and with whom.

- **Information specification**: Defining the information types is a major part of the negotiation framework. Information types are the major issue of the negotiation. Figure 3 depicts the dialog form to define the information types of the domain. As seen, there are 14 information types defined in our scenario.

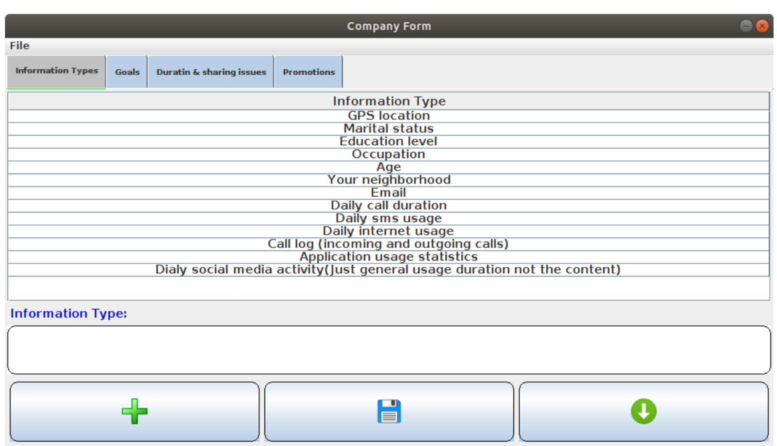


Figure 3: DI Type- Information type specification

- **Duration and sharing policy**: The duration of the contract denoting how long the access rights are given and sharing policy specifying with whom the data will be shared also plays an important role in the negotiation. We have two sharing policy options: (1) sharing just with company (2) sharing with company and third party. In addition we have 4 sharing duration options: (1) six months, (2) one year, (3) three years, and (4) five years.

- **Promotion packages**:

  Table 1 shows 12 promotion packages defined in our use case scenario.

Package name	Duration	SMS	Call hours (Minutes)	Internet quota
Package-1	1 Week	100	60	1
Package-2	1 Month	50	30	1
Package-3	1 Month	100	60	-
Package-4	1 Month	-	30	2
Package-5	1 Month	500	-	-
Package-6	1 Month	-	100	-
Package-7	1 Month	-	-	4
Package-8	3 Months	-	30	2
Package-9	3 Months	500	-	-
Package-10	3 Months	-	100	-
Package-11	3 Months	100	60	-
Package-12	3 Months	-	-	4

Table 1: Promotion packages

3.1.2 Preference Elicitation

We first explain the preference elicitation phase for the data consumer (i.e. telecommunication company) and then show how we elicit the preference of the data provider.

3.1.2.1 Data consumer

The data consumer first defines his goals and associates the required information types with the specified goals accordingly as well as specifying to what extent the goals are important for the company. Afterwards, the company provides cost information for the promotion packages.

- **Goal definition**: The utility of a given offer for the company is decided by the goals met by the content of the offer as described in the previous chapter. For each goal, company defines the information types that are required. Figure 4 and Figure 5 demonstrate the dialog forms to specify what information is needed for each goal of the company and determine the importance of the goals. The weight values are in the range of 0 and 100.

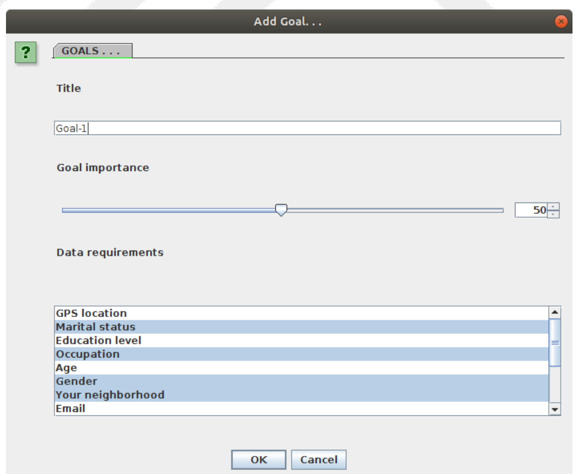


Figure 4: Defining company goals

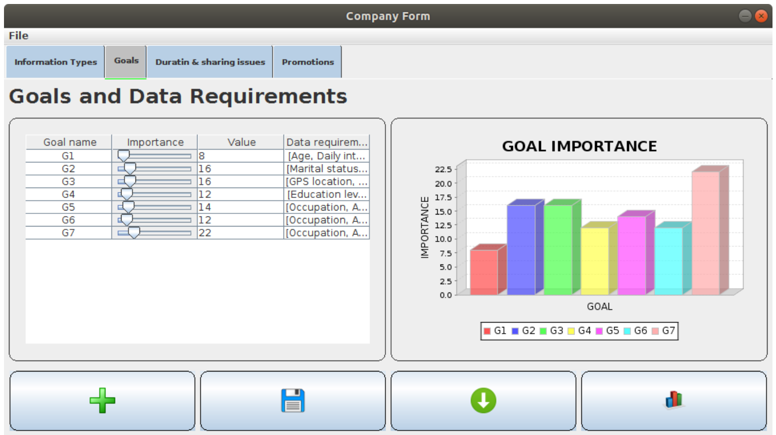


Figure 5: Goal specification

Figure 6 depicts how the preferences of the company is elicited regarding sharing policy and duration. Since the cost of the promotion packages are used in estimation of the expected utility of the company agent, the company representative specifies the cost for each individual components of the packages by the form shown in Figure 7. Figure 8 shows the form where specification of both content and cost of promotion packages.

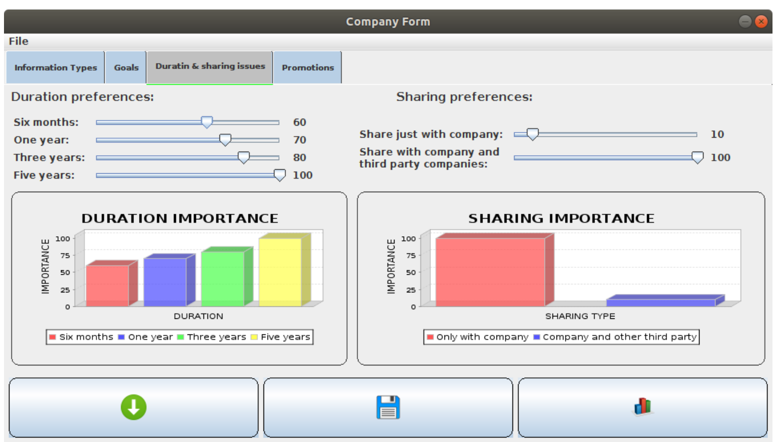


Figure 6: Duration and type of information sharing

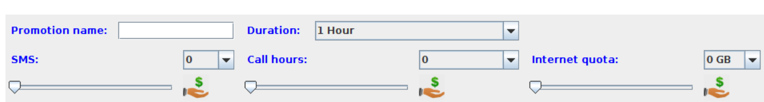


Figure 7: Unit cost specification for each offered free service in promotions

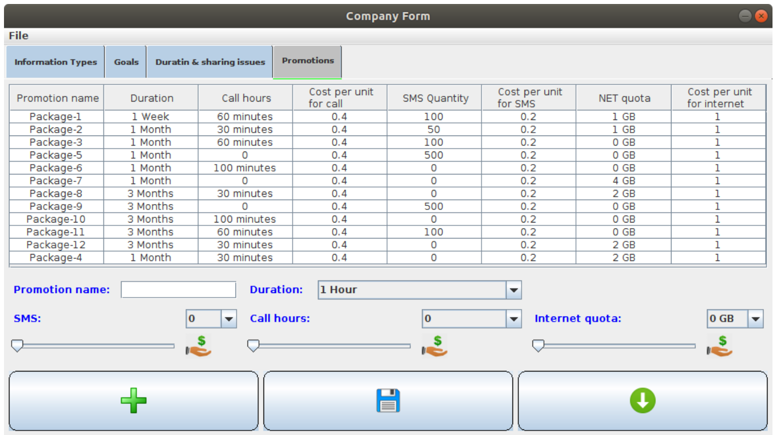


Figure 8: Incentive - Promotion Specification

3.1.2.2 Data provider

The data provider in our case customers can specify their preferences in three steps. The data provider first specifies her preferences regarding the sharing risk and secrecy level of each predefined information type, sharing policy (whom to share) and duration as shown in Figure 9

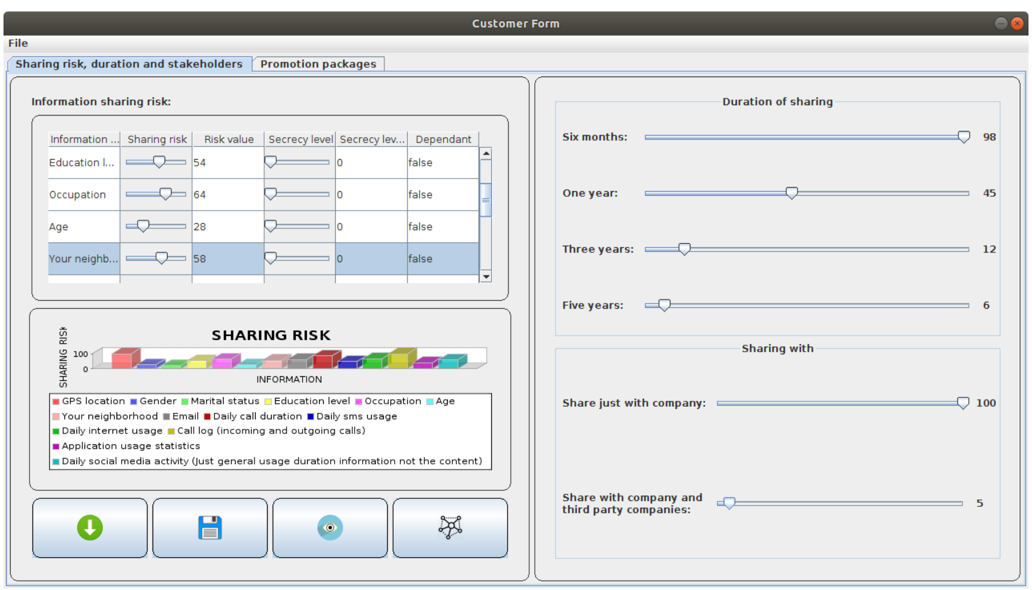


Figure 9: Customer's preferences regarding sharing risk, secrecy level, duration, and type of sharing

As explained in the previous chapter, there may be some information types which are more harmful when shared together. The bundle of these items are called dependency in our framework. We let the data provider define their dependencies as shown in Figure 10. For those dependencies, they need to specify sharing risk value separately.

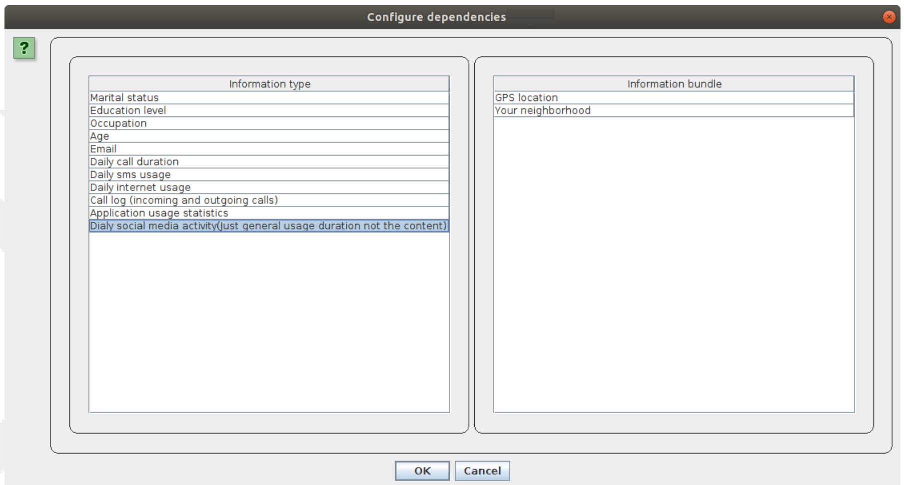


Figure 10: Defining dependency

Finally, the data provider should specify their preferences on promotion packages. Each individual can indicate a value between 0 and 100 for each of promotion packages which are presented by the company. Note that high values mean that they are more preferred.

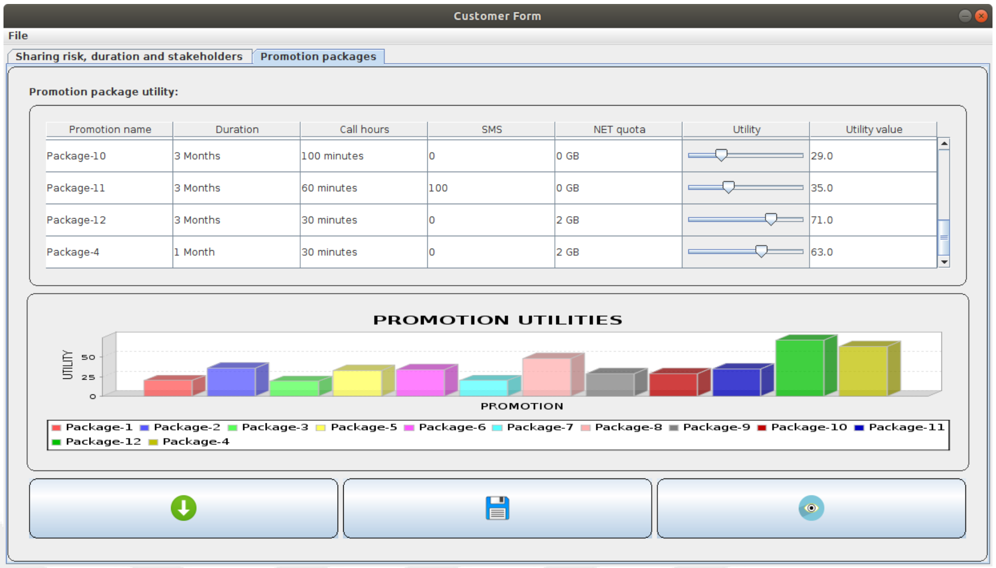


Figure 11: Customer's preferences on promotion packages

3.1.3 Negotiation Tool

Using the negotiation tool, the company agent and individuals can interact with each other to reach a consensus on information sharing domain. The main user of the negotiation interface is the human user. The domain, company profile and user profile are uploaded by the human participants. However the company profile is not visible for the human user.

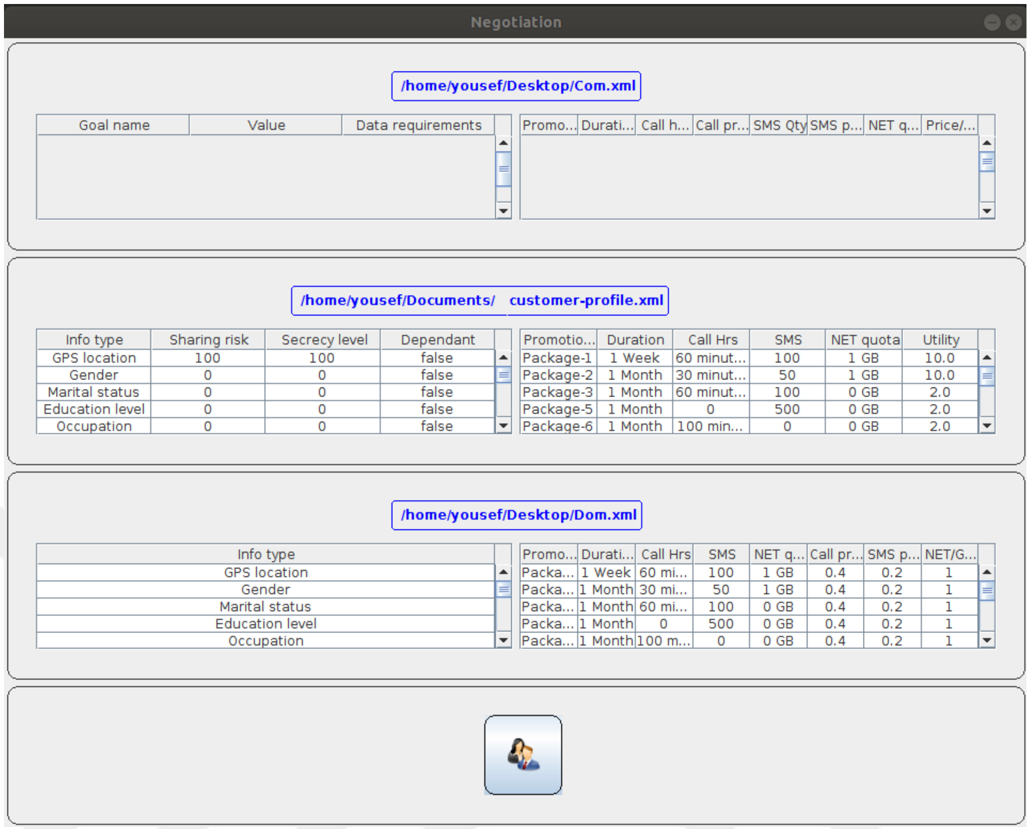


Figure 12: Negotiation start

After pressing the start button at the bottom of the page as seen in Figure 12, the negotiation starts with an offer made by the company agent. At the first step, the company agent offers a bid with maximum utility for itself as seen in Figure 14. If user accepts the offer, a message appears and negotiation ends. On the other hand, if the first offer gets rejected by human user then the user should make his/her offer to company agent by selecting the values for each negotiation issue in the given form as seen in Figure 13 and 15.

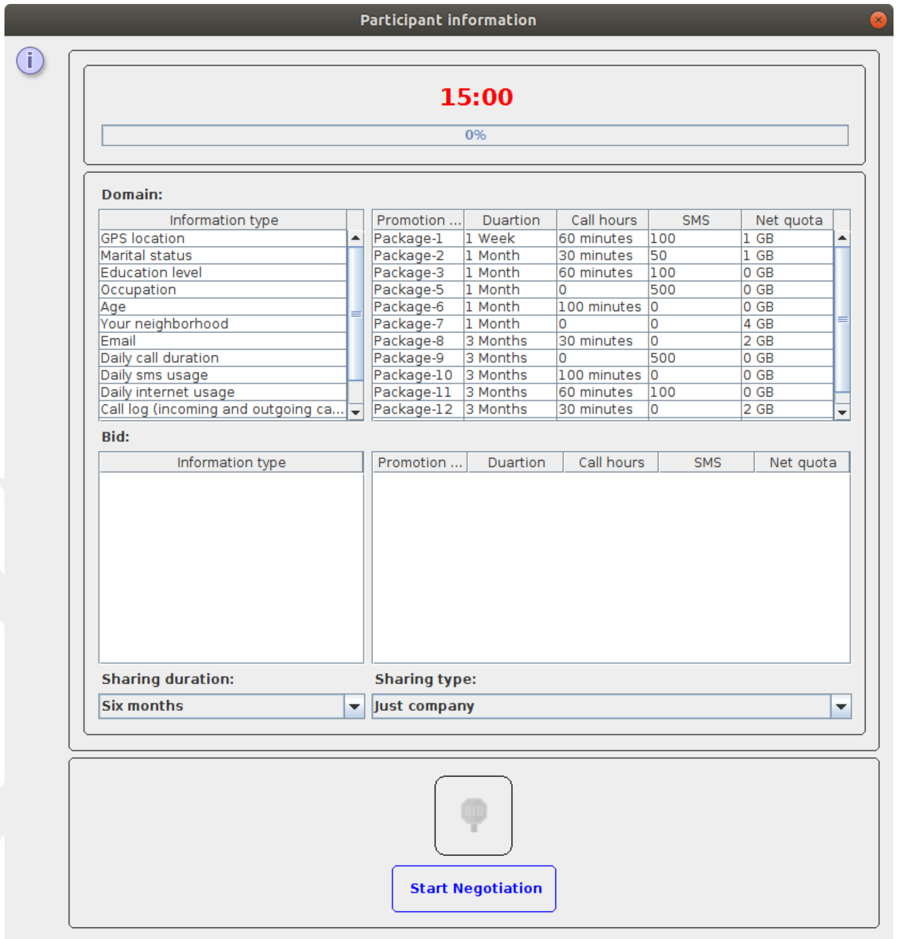


Figure 13: Making counter offer

Each offer from the company side appears in a specific window which represents the contents of the offer(i.e., information type, offered promotion package, duration, and the type of sharing information). User can rate the offers received from company and her/his own offers as well.

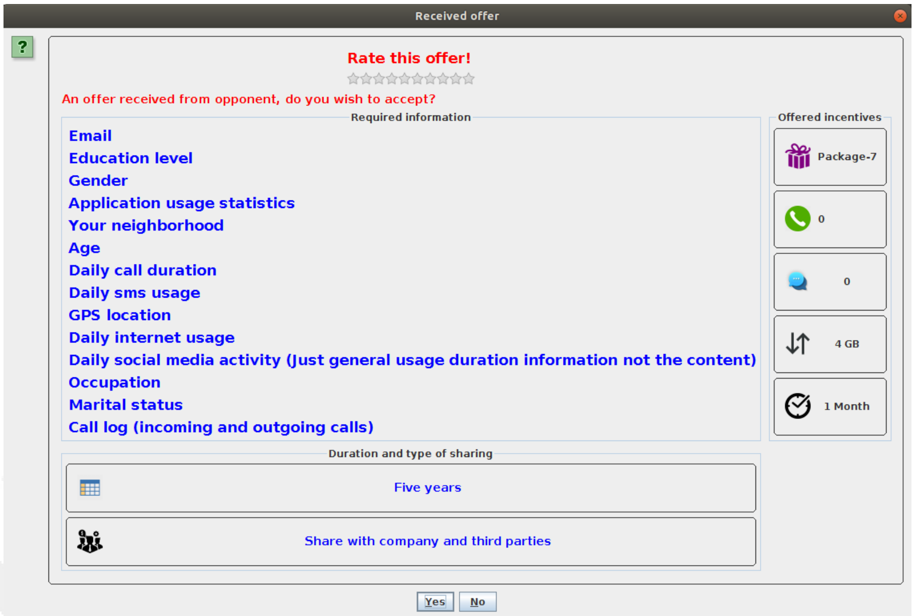


Figure 14: Company agent's offer

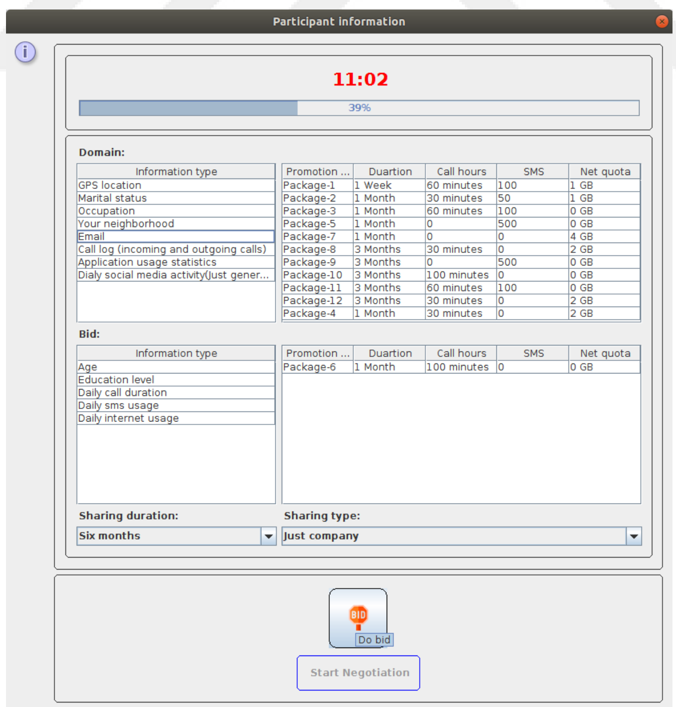


Figure 15: Customer's counter offer

3.2 Experimental Setup and Analysis

We have conducted a user experiment to evaluate the proposed framework. There were 25 participants in our experiment in which each user plays the role of the data provider and specify their preferences and negotiate with the company agent on the telecommunication scenario explained in the previous section. The participants are selected among undergraduate and graduate students of the science and engineering faculty at Özyeğin University. Participants' age ranges from 22 to 30 and there were 20 male and 5 female individuals. After the preference elicitation and negotiation phase, they are asked to fill in a questionnaire form 3.2.4 consisting of 17 questions regarding their negotiation and experience with the tool. The results regarding the experiment and the questionnaire will be presented in the upcoming sections.

3.2.1 Data Consumer's Profile

For this experiment, we created a profile for the company agent. In our scenario, the company have seven goals as follows:

- G1 with a weight value of 8: Age, daily internet usage
- G2 with a weight value of 16: Marital status, occupation, age, daily call duration
- G3 with a weight value of 16: GPS, your neighborhood, occupation
- G4 with a weight value of 12: Email, education level, occupation, age, daily social media activity
- G5 with a weight value of 14: Application Usage Statistics, occupation, age, gender
- G6 with a weight value of 12: Call list, Occupation, Age, Gender
- G7 with a weight value of 22: Age, Occupation Daily call duration, Daily SMS usage, Daily Internet usage, Application Usage Statistics

In our pilot studies, we observed that human participants feel more comfortable to specify their preferences between zero and 100 rather than a real number between 0 and 1. Therefore, we use that scale for all components of the expected utility. That is, the expected utility for the agent is between zero and 100. Accordingly, we specify the utility values for other components of the offers as follows. Briefly, the company prefers 5 years over 3 years, and 3 years over 1 year and 1 year over 6 months. It prefers the sharing policy with sharing with third company over sharing with only company.

- Duration (5 years) = 100
- Duration (3 years) = 87
- Duration (1 year)= 75
- Duration (6 months)= 60
- Sharing policy(company only) = 80
- Sharing policy(company with third company) = 100

The weight values in estimating the overall expected utility are listed as follows: 0.7 for goal satisfaction; 0.2 for duration and 0.1 for sharing policy.

To calculate the value of cost, we calculate the cost of each promotion cost (See Table 3.2.1) and normalized it by the following formula $\frac{cost-min(c)}{max(c)-min(c)} \times 100$ where $min(c)$ and $max(c)$ represent the minimum and maximal cost in package costs.

Package name	Cost	Package name	Cost	Package name	Cost	Package name	Cost
Package 1	45.0	Package 2	23.0	Package 3	44.0	Package 4	14.0
Package 5	100.0	Package 6	45.0	Package 7	4.0	Package 8	14.0
Package 9	100.0	Package 10	40.0	Package 11	44.0	Package 12	14.0

Table 2: Cost of promotion packages for company

3.2.2 Preference Elicitation

Human participants play the role of the data provider. Before their negotiation, we elicited their preferences. Afterwards, a detailed analysis was conducted. Figure 16 demonstrates the heat map for the sharing risk level for each individual participated in the experiment. As demonstrated using the legend on this figure, the darker color means that the participant assigns higher risk value to that specific information type. Based on our observation, we can claim that there are certain information types such as GPS and Call log, which have a high sharing risk value for almost all of the individuals. We can find out some patterns, which categorize the individuals regarding their attitude toward information sharing risk. While some participants behaved in an ignorant fashion (e.g. participant 18), others acted more conservatively on their privacy (e.g., participant 3).

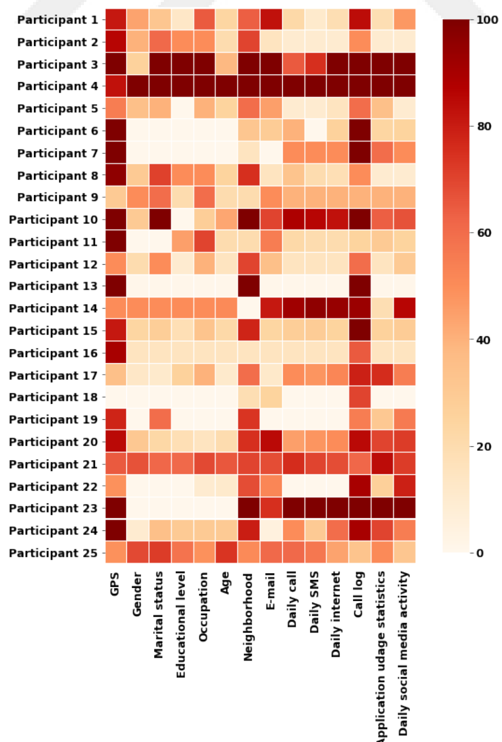


Figure 16: Information sharing risk map

Each information type would have different sharing risk and secrecy levels depending on individual preferences and with respect to desired application. For example, "gender" would be underestimated by most of the individuals since of lack of knowledge on how some recommender systems may be fed using biased data. Figure 17 demonstrates the secrecy level for each individual participated in the experiment. In Figure 17, the darker tones mean that the participant considers that specific type of information more personal and secret. We can claim that there is a certain difference between the participants sharing risk preferences and secrecy level but there can be some correlations in some cases, since some people (e.g. GPS for participant 3) consider both secrecy level and sharing risk equally as well.

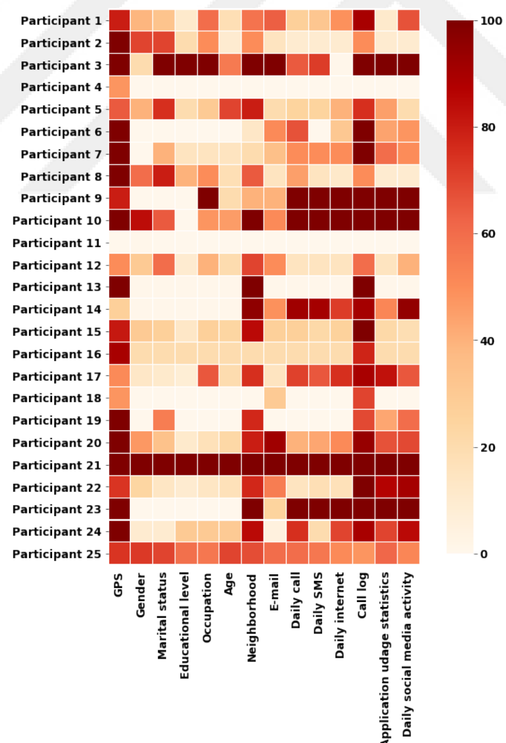


Figure 17: Information secrecy level map

Figure 18 demonstrates the mean and standard error on the collective preferences of individuals. This is an attempt to generalize participants preferences as whole group and understand group behaviour toward the information privacy preferences. This illustration reveals that three most important information type regarding both sharing risk and secrecy level for all users are the GPS location, neighborhood, and Call log respectively.

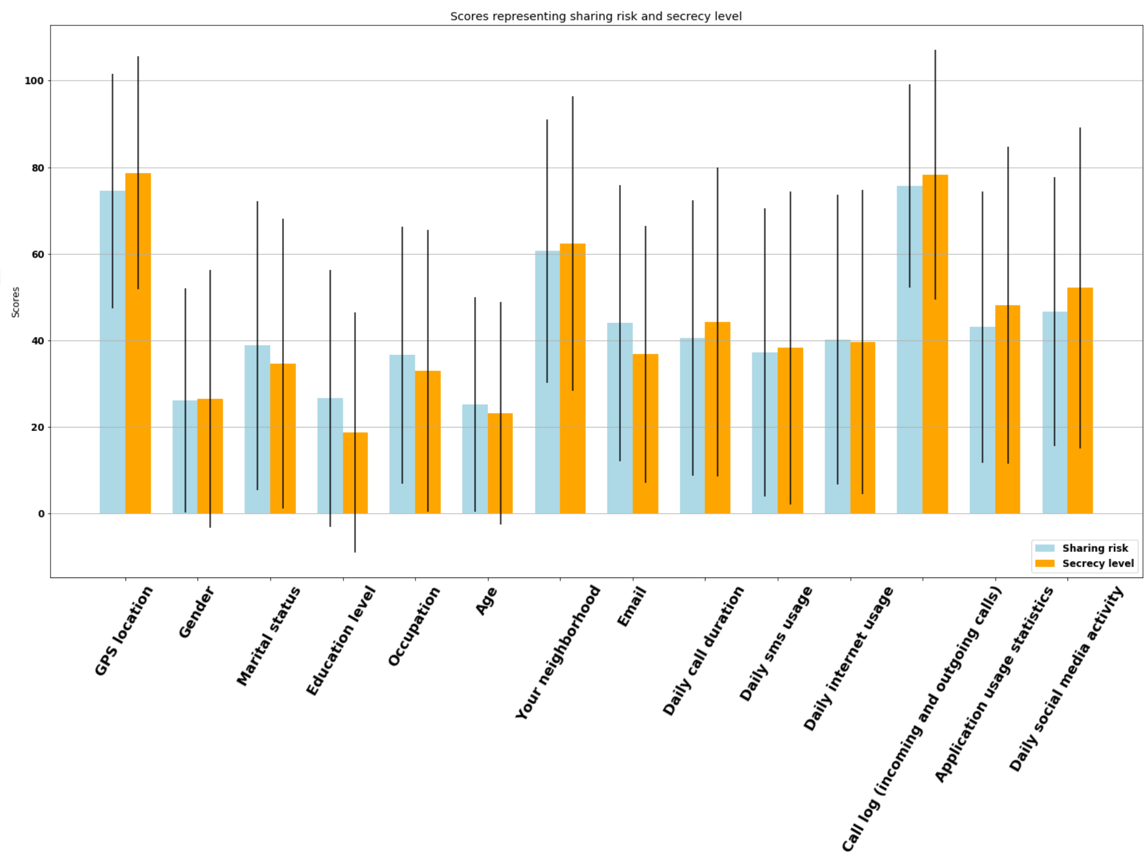


Figure 18: Sharing risk and secrecy level preferences

Figure 19 illustrates the mean and standard error for the utility values that participants assigned for utility values of the duration and sharing policy. As expected, people preferred shorter duration over longer one and preferred company only over company with third companies option.

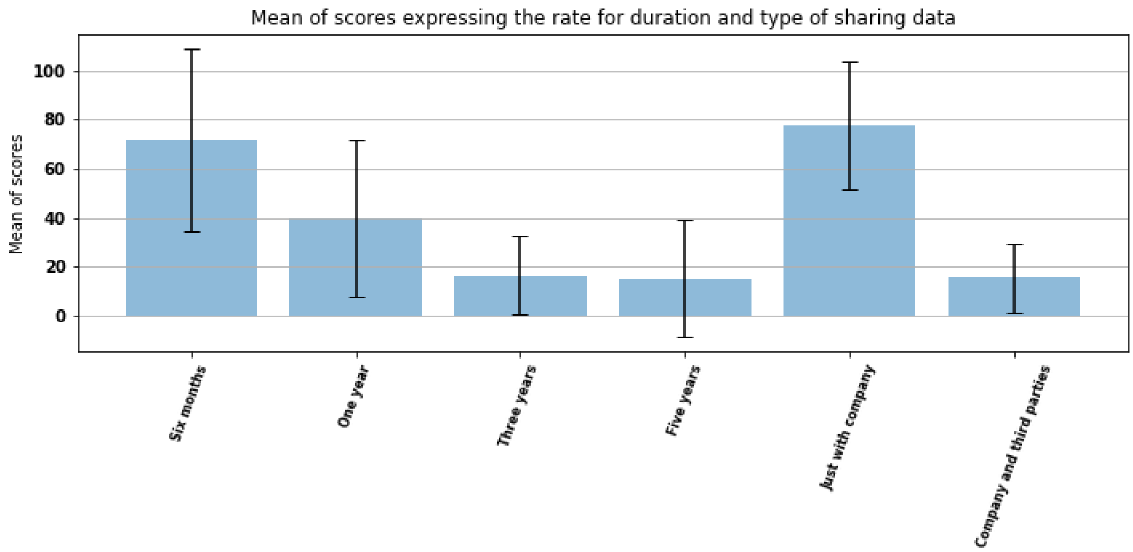


Figure 19: Sharing duration and type

Figure 20 illustrates the mean and standard error for the values that participants indicated to represent their desire to get each specific promotion package in exchange for the privilege of information usage. The contents of the packages is represented in Table 1. In this figure we can observe that users are mostly desired to get packages that include higher internet service quota or higher call duration and SMS in case of lower internet quota.

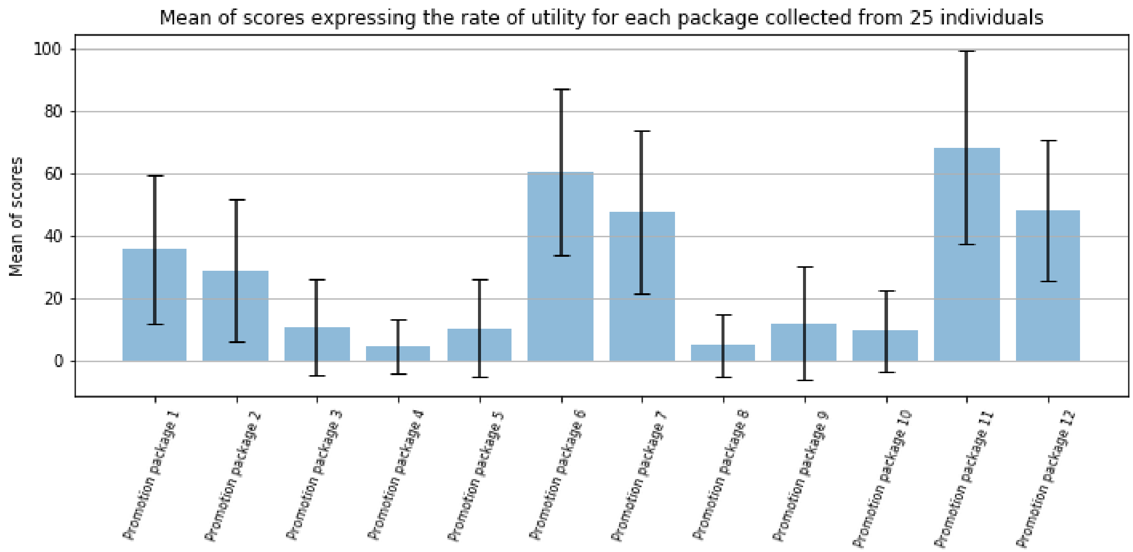


Figure 20: Utility preferences on promotions

Some participants indicate some information bundles which they assume if they share them together in any offer there would be more risky. These bundles are named as dependencies, Table 3.2.2 and Figure 21 illustrate the contents of dependencies and preferences of participant regarding sharing risk of dependencies.

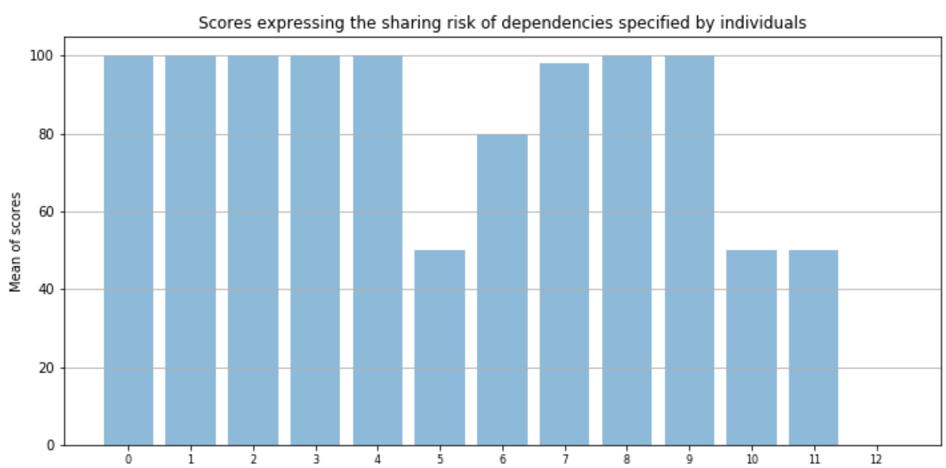


Figure 21: Dependency sharing risks

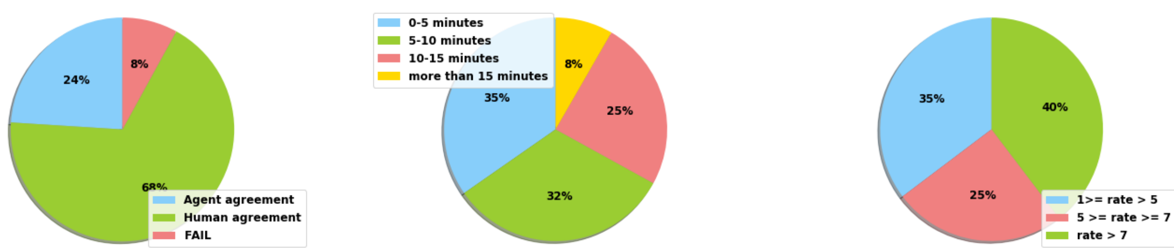
Order	Contents of dependency	Sharing risk
0	GPS location , Your neighborhood, Call log (incoming and outgoing calls), Occupation	100
1	GPS location, Occupation, Marital status, Daily internet usage	100
2	GPS location, Marital status, Occupation, Your neighborhood	100
3	Daily internet usage, Application usage statistics	100
4	GPS location, Marital status, Your neighborhood, Call log (incoming and outgoing calls), Occupation	100
5	GPS location, Your neighborhood	50
6	Occupation, Email, Daily internet usage, Daily social media activity (Just general usage duration information not the content)	80
7	Gender, Marital status, Age	98
8	GPS location, Your neighborhood, Email	100
9	GPS location, Your neighborhood	100
10	Your neighborhood, GPS location, Gender	50
11	Email, GPS location, Your neighborhood	50
12	GPS location, Call log (incoming and outgoing calls), Daily social media activity (Just general usage duration information not the content)	0

Table 3: List of dependencies

3.2.3 Negotiation

There were 25 negotiation sessions in our experiment where human participants negotiated based their own preferences. Among all sessions, only two of them resulted in failure because of reaching time deadline without an agreement. In six sessions agent accepted human negotiator's offer while 17 sessions ended with an agreement from human side as represented in Figure 22(a). It is obvious that the majority of successful sessions ended with an agreement from human side.

Figure 22(b) demonstrates the categorization of negotiation sessions regarding the duration of negotiation. It is obvious that most of the participants could reach an agreement with our agent in 10 or less than 10 minutes from the beginning of session. Figure 22(c) demonstrates the percentage of the user rating scores on the last offer (i.e., agreement in case of successful negotiation). It is observed that most of participants (i.e., 65 percent) are well satisfied with the result of negotiations.



((a)) ((b)) ((c))

Figure 22: Experiment details

Figure 23 demonstrates the distribution of utility values gained by agent and user ratings in each negotiation session. Note that orange bars show the user ratings while blue bars denote the utility of the agent. Note that user ratings are subjective evaluation. In case of unsuccessful negotiation, the utility gained by parties is taken as zero.

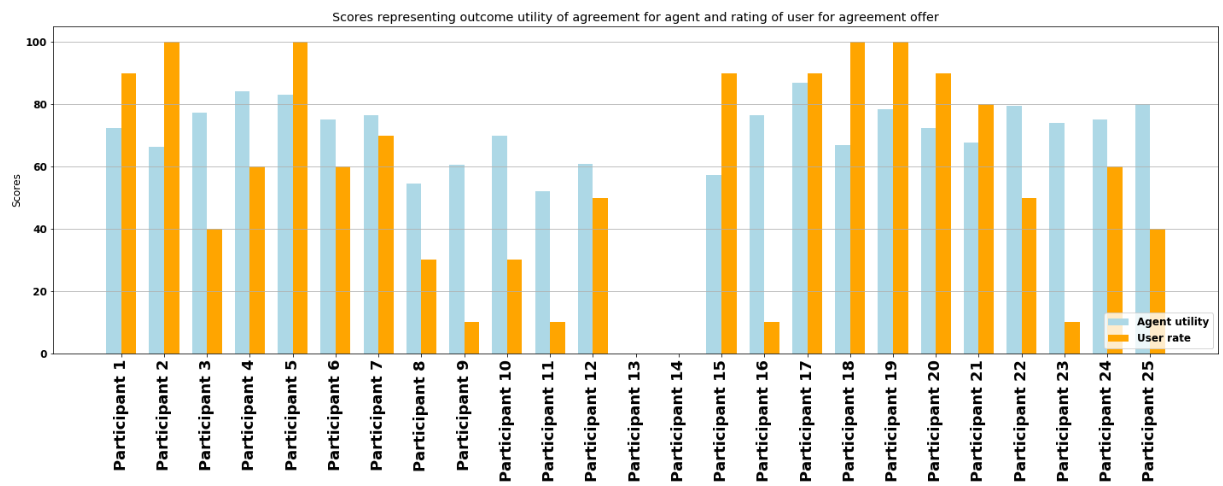


Figure 23: Distribution of utilities in negotiation sessions

Figure 24 depicts the willingness of each participant to share specific information type during the negotiation. This factor is calculated using the frequency of appearing each specific information type in the offers that the participant sent to our agent. In other words, it demonstrates the behavior of the participant in negotiation rather than their preferences which is depicted in Figure 16 and Figure 17. In Figure 24 the, darker tones show the information types that the participant have more tendency to share while lighter tones represent the information types that user avoid to share with the company. It can be observed that people are so sensitive regarding certain data type (e.g., GPS and Call log) while they are fine with sharing data types such as "Age" and "Educational level". Additionally, some conservative behavior patterns are observable (e.g., participant 13, 11 and 9).

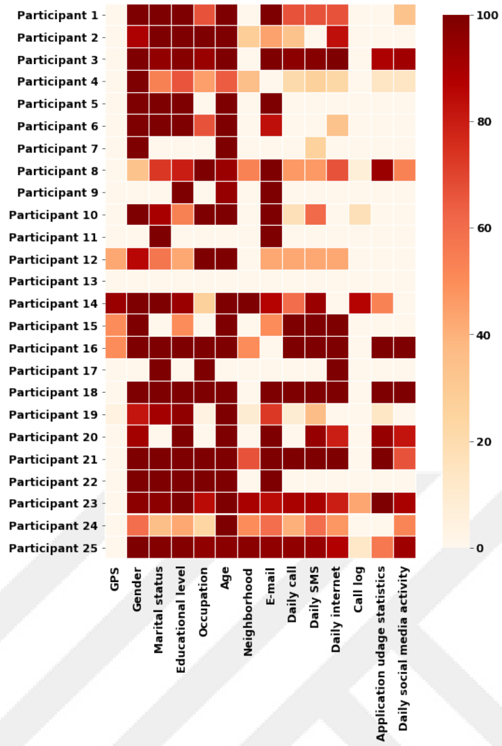


Figure 24: Willingness to share information types based on negotiation logs

Figure 25 demonstrates the percentage of the frequencies for each information type in the human participants' offers. People were mostly inclined to share "Age", "Gender" and "Marital status" while they avoided to share their "GPS" and "Call log".

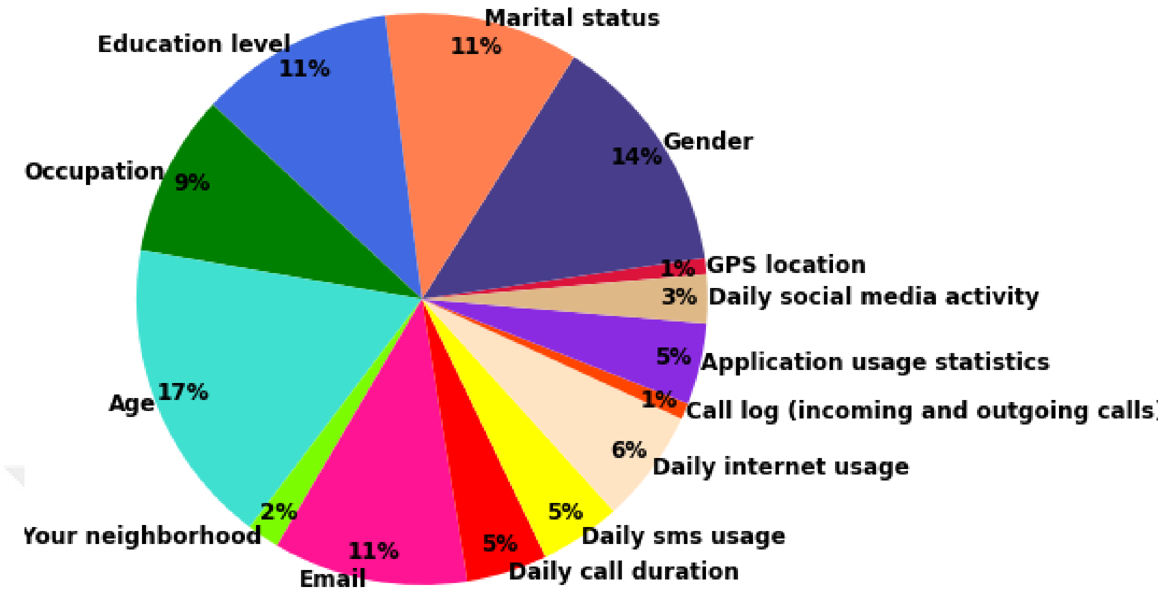


Figure 25: Ratio of each information type in human offers

Figure 26 demonstrates the most frequent promotion packages in user bids. The percentage of each package in this figure represents the desire of participant toward getting that package. Lower values show the packages that are so less popular among others. Other package names that are included in this figure did not appear in user bids when they negotiated with the agent. We observed that most of them are interested in getting package 12 and package 7, which provide them more internet service than other packages.

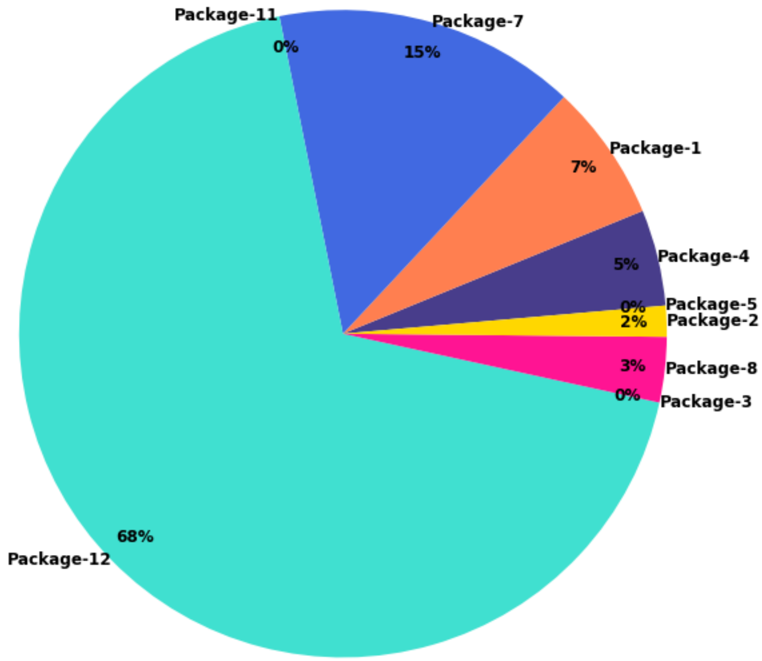


Figure 26: Ratio of each desired promotion package

Figure 27 demonstrates the ratio of sharing policies preferred by participants in negotiation. Based on our observation, participants were inclined to share their data only with the main company rather than third party companies.

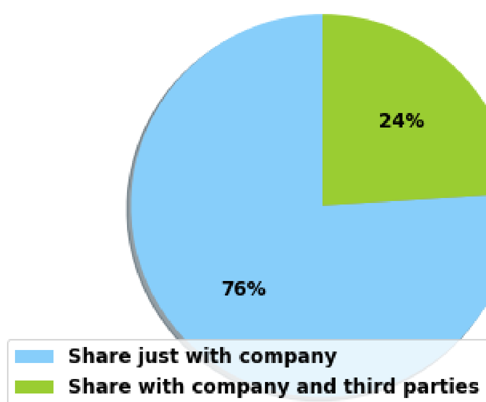


Figure 27: Ratio of each sharing policies

Figure 28 depicts the percentages of each value for duration appeared in human participants' offers. It is clear that people are mostly like to share their data for shorter time periods.

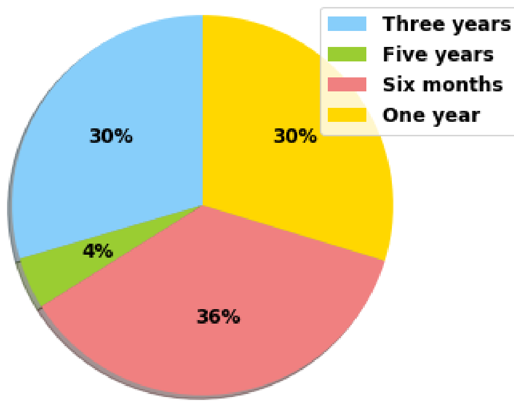


Figure 28: Ratio of each sharing duration

3.2.4 Questionnaire

Figure 29 shows the average ratings given by the users to our questionnaire consisting of 9-point scaled questions after their negotiation (1 for strongly disagreement whereas 9 for strongly agreement). The statements in the questionnaire are listed in Table 3.2.4. The aim of this questionnaire is to understand what they think about the proposed framework and their experience with our preference elicitation and negotiation tools. Participants responses are visualized separately in the appendix A.

The average rating for the question "I like the idea of negotiating about the information sharing policy (i.e., types, duration, etc.) and incentives/promotion packages." is high (7.5 out of 9). Similarly, a similar response was gathered to the question "I prefer to be able to make a counter offer as an additional option

to accept or reject the company's offer." Those ratings support the idea of our negotiation framework in information sharing.

Participants on average considers that our agent made reasonable offers during their negotiation. The average rating of the participants to the 14^{th} question is also high (around 6).

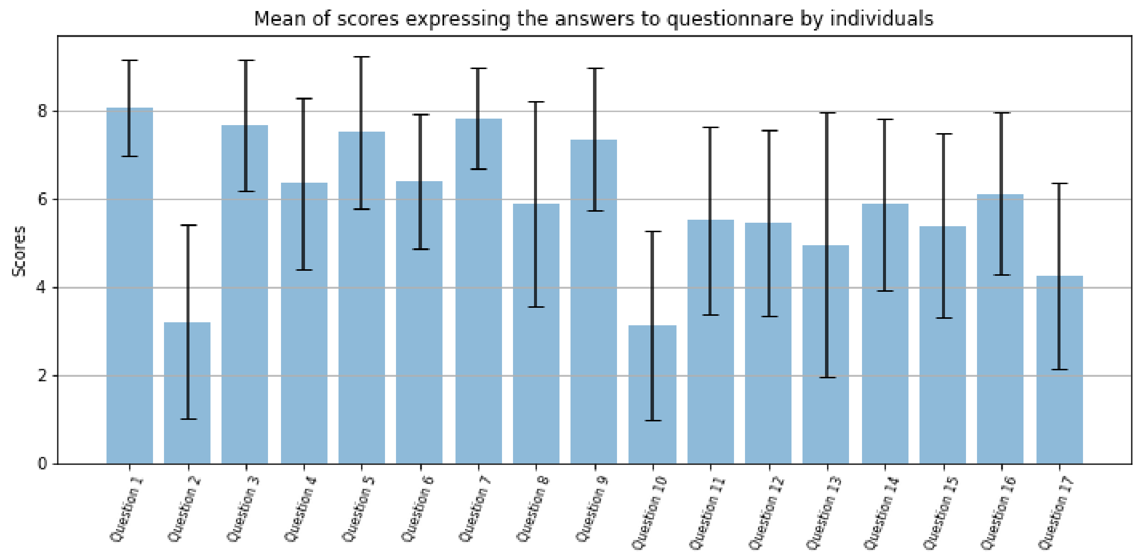


Figure 29: Individuals answers to questionnaire

Order	Question
1	The instructions provided to me for the experimental negotiation were clear.
2	It was not clear to me how I should use the preference elicitation tool.
3	It was clear to me how I make my bids in the given negotiation tool.
4	Specifying my preferences in the given tool increased my awareness of privacy.
5	I like the idea of negotiating about the information sharing policy (i.e., types, duration, etc.) and incentives/promotion packages.
6	I will be confident if a software agent negotiates with the company on behalf of me after eliciting my privacy preferences.
7	I prefer to be able to make a counter offer as an additional option to accept or reject the company's offer.
8	Assessing my privacy preferences was more challenging process than I thought.
9	I would prefer to customize the content of the promotion during the negotiation rather than selecting the promotion items from predefined set.
10	It does not make sense to me to negotiate on information sharing policies and incentives.
11	I would not let a software agent negotiate about information sharing policy on behalf of me.
12	I would prefer negotiation with a software agent instead of a real person.
13	My preferences on sharing information policy would be the same for any context. It does not matter whether it is telecommunication or hospital.
14	My opponent made reasonable offers during the negotiation.
15	My opponent took my privacy concerns into account.
16	My opponent takes my previous offers into account while making its current bids.
17	My opponent was not collaborative at all to find a mutual agreement.

Table 4: Questionnaire

Chapter IV

DEEP REINFORCEMENT LEARNING FOR ACCEPTANCE STRATEGY

"If one day, my words are against science, choose science."

– Mustafa Kemal Atatürk

Automated negotiation [17] is an important study field in Artificial Intelligence, where intelligent agents negotiate on behalf of their users on multiple issues with the aim of maximizing their own utility. Bidding strategy[18], opponent modeling [19, 20, 21] and acceptance strategy [14] are the main challenges in automated negotiation. Agents exchange offers consecutively between each other to reach an agreement in a given negotiation scenario. This interaction is governed by a certain protocol determining the rules of encounter. Alternative offers protocol [12] is one of the most widely used protocols in bilateral negotiation. According to this protocol, an agent initiates the negotiation with an offer and its opponent can accept or reject this offer. If the opponent accepts the current offer, negotiation ends with an agreement and the utility of the agreement for each agent is calculated with respect to their preference profile. Otherwise, the opponent agent takes the turn and makes a counter offer. This process continues in a turn-taking fashion until reaching an agreement or negotiation deadline for that session. If the predefined deadline passes and there is no agreement, each agent gets the reservation value (i.e., BATNA). The turn taking fashion of taking actions in automated negotiation makes it an appropriate environment for applying Reinforcement Learning [22] algorithms, where agent can learn the best action to be taken based on the feedback given during these interactions.

Reinforcement learning (RL) is the process of finding optimal policy in an environment based on feedback coming from environment in response to the agents' actions.

In other words, agent learns from its experience. In RL, a state is a definite and immediate situation in which the agent finds itself. An action is any possible move that the agent can take in the given state. The goal of the agent is to select the action maximizing its performance. In RL, agents receive a signal from their environment, which indicates the extent of the success or failure of the agent's actions. Reward function maps the actions to rewards with respect to the feedback coming from environment. The environment is the space which the agent moves in. The environment takes the agent's current state and action as input, and outputs the reward and next state. The agent aims to learn the policy that maximizes its reward while interacting with the environment. In this work, a negotiating agent employs reinforcement learning in order to determine whether it should accept its opponent's offer.

Existing works on acceptance strategies for automated negotiation are mostly based on predefined rules that takes remaining time and utility of the negotiation outcome into account. For instance, AC-next is one the most widely used acceptance strategies where an agent accepts its opponent's offer if the utility of the opponent's offer is higher than the utility of its next offer. Furthermore, these predefined rules can be combined to form more complex acceptance strategies [14]. In this work, we aim to develop an acceptance strategy that learns when to accept opponent's offer using reinforcement learning while some recent works employ reinforcement learning in order to learn what to bid. A reinforcement learning agent receives a feedback from its environment for each action it takes and this feedback can be formulated as reward or cost. The agent has to Figure out what it did that made it get the reward or cost, which is known as the credit assignment problem [23]. In RL problems, the environment is modeled as a Markov Decision Process (MDP) with inputs (actions sent by the agent) and outputs (observations and rewards sent to the agent). The sequential flow of operations in the negotiation framework makes it an appropriate test bed to implement and assess the RL algorithms on agents. Since our aim is to design and develop a domain independent acceptance strategy which can be applicable to any given negotiation scenarios, GENIUS [24] (**G**eneric **E**nvironment for **N**egotiation with **I**ntelligent multi-purpose **U**sage **S**imulation) environment is chosen as our negotiation platform. An advantage of the GENIUS environment is embedding the BOA framework [25], which enables developing negotiation components (i.e., Bidding strategy, Opponent modeling, and Acceptance strategy) separately. There are a variety of acceptance strategy modules developed using the BOA framework in GENIUS. This framework enables researchers to develop and study individual parts of the negotiation strategy.

The primary goal of our approach is the development of an acceptance strategy which can engage and achieve reasonable results with different opponents regardless of the negotiation domain. During the negotiation, it is observed that the agent learns what to accept over time and improves its performance. The main contribution of this thesis is developing and exploiting a domain independent acceptance strategy using reinforcement learning. The developed strategy is compared with AC-next acceptance strategy and in some cases the proposed approach performs better than the AC-next while in other cases it achieves close performance to AC-next.

Chapter V

PROPOSED ACCEPTANCE STRATEGY

> "People who are crazy enough to think they can change the world, are the ones who do."
>
> – Rob Siltanen

Our negotiation environment consists of two agents negotiating on a multiple issues (e.g., selecting travel destination, deciding on location and etc) under a certain time limit. The agents take their actions in a turn taking fashion by alternating offers protocol [12]. The action set consists of accepting the opponent's offer and making a counter offer which means "rejecting the counter offer". The preferences of agents are represented by means of additive utility function where overall utility of an outcome is calculated by weighted sum of the each individual utility value of each issue. In the given system agents know only their own utility function. That is, they do not know their opponent's utility values. The goal of the agents is maximizing their total utility gained from the accepted offers. If the agents do not reach an agreement before the deadline, each agent gets the reservation value. In this work we proposed to use a Deep Q Network (DQN) to learn when to accept.

In Q-learning there is a table consisting [state-action] pairs, the Q function is used in order to determine the value of state-action pairs. For deciding which action to perform, with a probability equivalent to 1 - ϵ the agent acts greedily and picks the action that maximizes the outcome utility. This is called exploitation, the agent performs actions which lead to maximizing the return. Then, the Q function is updated based on the action performed.

In this work the Q function gets updated in terminal state transition right before an accept. However, when the agent sends an offer as a counter offer there is an uncertainty about the acceptance of the offer by opponent. Therefore, the agent needs to wait for opponent's response to make sure that the offer is accepted or rejected. In case of acceptance the Q function gets updated and next negotiation session starts. In order to train our DQN, we calculate the state action values with Q-learning update rule and provide feedback to the neural network after each action performed. Q-learning algorithm is used to determine which offers should be accepted or counter-offered. Q-learning is a value based reinforcement learning algorithm where agent receives a reinforcement signal called reward, the agent is up to maximizing the reward based on taken actions with respect to target actions. The agent learns the state action values of its actions with Q-learning updates and performs its actions based on ϵ greedy policy. ϵ is a constant value between 0 and 1 that determines the probability of choosing random action and 1 - ϵ determines the probability of choosing action which its state-action value is the maximum. By this method, for each action that agent performs it explores the environment with ϵ probability and exploits the environment based on its experience with $1-\epsilon$ probability.

Our general approach is Q-learning update which is done after each step transition and is as follows.

$$Q(S,A) \leftarrow Q(S,A) + \alpha[r + \gamma max_a Q(S^{'},A) - Q(S,A)]. \tag{7}$$

In Equation 7 the Q(S, A) is the Q function which outputs the Q-Value of the next state according to the current state and action taken. S' is the transitioned state, r stands for the immediate reward received by transitioning from state S to S'. γ represents a constant which scales the Temporal Difference (TD) approximation effect. The expression $[r + \gamma max_a Q(S',A) - Q(S,A)]$ is called TD error which is the amount of update on Q(S, A) and is reduced by the α which is the learning rate. However, unlike the tabular approach, we use a neural network to approximate Q function which is known as semi gradient Q-learning.

Our state consists of five elements represented as a tuple:

$$< \Delta O, D, MNU, R, C >$$

where

- ΔO: The difference between target utility and received offer utility

- D: Scaled remaining time

- MNU: The agent's next offer utility

- R: The target utility

- C: Current utility value of the opponent's offer

The agents can perform actions whether to accept the given offer or to make counter-offer. The immediate reward that agent receives after each step transition is as follows:

$$Immediate\_reward = \left\{ \begin{array}{cc} -2^{|target\ utility - final\ value|} & \text{(target utility>received utility)} \\ 2^{|target\ utility - final\ value|} & \text{(target utility<received utility)} \\ 0 & \text{(The step transition is non-terminal)} \end{array} \right\}$$

The reward function (Equation 5) we considered for the acceptance strategy is called in transition steps. The transition steps result in *terminal* or *non-terminal* states and for non-terminal states the reward is zero. Non-terminal states are the exchanging of counter offers between agents. Terminal states are the states that one of agents accepts an offer from its opponent and session ends or when the session reaches to deadline without an agreement. The reward in terminal states is calculated using an exponential function with respect to the difference between received utility and target utility. Note that target utility is the utility value which agent aims to achieve in negotiation. If the target utility is greater than received utility, the reward is negative otherwise the reward is positive. Our framework is demonstrated in Figure 30.

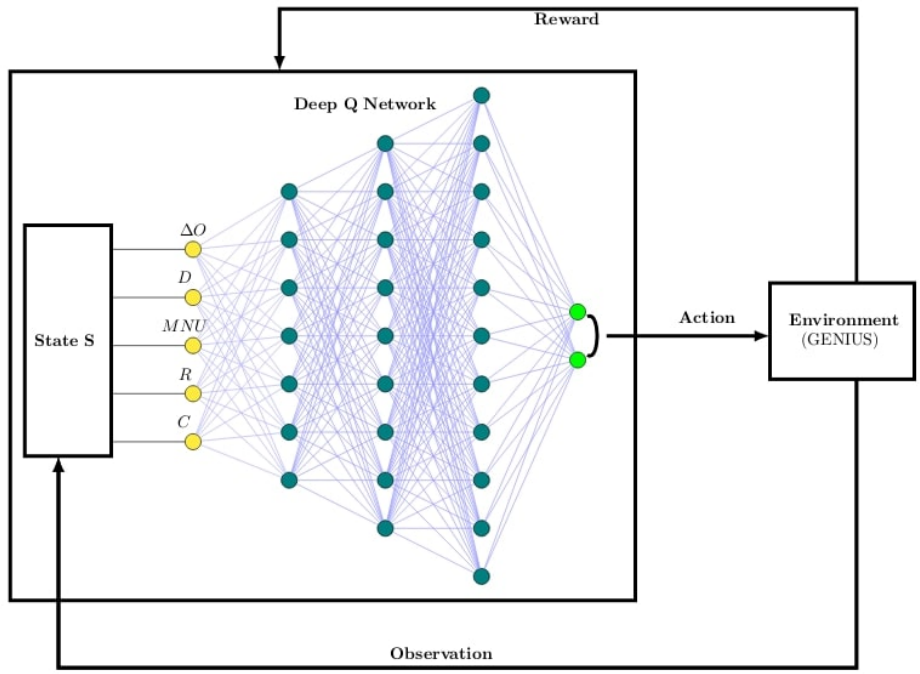


Figure 30: Proposed Negotiation Architecture

As illustrated in the Figure 30, there is a feedback loop between the agent and the environment in the concept of RL. State S goes through the DQN and after passing the hidden layers the outcome of the model are two float values ranging between 0 and 1 which defines the possibility of the best action to take in state S. The action with highest value goes through environment and environment gives a reward signal with respect to the action and next state which again goes through the DQN in the form of state S. This feedback loop shapes the learning process of the model. Algorithm 1 illustrates the proposed approach for our acceptance strategy, in algorithm the NN is an abbreviation for Neural Network.

Determine_Acceptability() is called after receiving an offer from the opponent and does the decision making at non-terminal state transition updates. The difference between our method and tabular method is the fact that we do not keep a table for each state action pair. Instead we give state as input to neural network and specify the Q(S,A) as the output of neural network. This method enables learning in continuous state space.

Algorithm 1 Acceptance Strategy Based on Q-Learning

```
procedure DETERMINE_ACCEPTABILITY()
    if previousState == null then
        previousState←getCurrentState()
    else
        currentState←getCurrentState()
        < V_Accept, V_Reject >←NN.predict(previousState)
    end if
    if V_Accept > V_Reject then
        maxValuedAction←Accept
    else
        maxValuedAction←Reject
    end if
    immediateReward←0
    < V_Accept, V_Reject >←immediateReward+γ*currentState[maxValuedAction]
    NN.train(previousState, < V_Accept, V_Reject >)
    With 1-ϵ probability do:
    Begin
        < V_Accept, V_Reject >←NN.predict(getCurrentState())
        if V_Accept>V_Reject then
            return Accept
        else
            return Reject
        end if
    End
    With ϵ probability do:
    Begin
        return Randomly(Accept|Reject)
    End
end procedure
```

5.0.1 Generalization and Regularization

The motivation is to propose an acceptance strategy based on deep reinforcement learning which can generalize over several domains and opponent agents in the bilateral negotiation. Therefore, first we trained the agent with several agents and domains to enable the DQN to learn various agent behaviors on different domains. The problem with training on different domains and against various opponents was that in each session our agent tried to learn that specific domain and agents possible bids. Due to environmental constraints, setting sessions which can include different domains and several opponents at the same time was impossible. We overcome this problem by generating virtual states and feeding them through experience replay memory which is used already for previous negotiation sessions. Experience replay memory is a reservoir of any desired number of transitions to be sampled from later for the agent to learn from. Existence of this memory separates the learning phase from gaining experience. In this approach learning is happened based on taking random samples from this memory. In our state definition each state is a tuple of 5 float values so by generating a dataset of possible states and feeding them to the network we simulate such conditions to enable generalization for DQN. To prevent overfitting in DQN we applied L2 regularization on the DQN. L2 regularization adds squared magnitude of coefficient as penalty term to the loss function.

Chapter VI

EVALUATION AND COMPARISON OF ACCEPTANCE STRATEGY

> "Anyone who stops learning is old, whether at twenty or eighty. Anyone who keeps learning stays young. The greatest thing in life is to keep your mind young."
>
> – Henry Ford

In order to evaluate the performance of the proposed approach, we implemented a negotiating agent adopting our RL-based acceptance strategy in BOA framework of GENIUS environment. This environment hosts a variety of negotiation scenarios (i.e., negotiation domain and a pair of preference profiles) and negotiating agents. In this platform, an agent can negotiate with an opponent in different negotiation scenarios. Recall that we adopt experience replay memory to maintain the state and action pairs from different negotiations. Therefore, the DQN learns the optimal acceptance conditions in different negotiation settings. In the following sections, we will describe how we have trained our RL-based agent (Section 6.0.1) and present the experiment results with respect to performance of agent in test environment (Section 6.0.2).

6.0.1 Training Session

For training purposes, a well-known negotiation domain – *England-Zimbabwe* [26]– is used. Figure 37(d) demonstrates the utility distribution of available bids for this scenario. For each profile, our agent negotiates with its opponent 600 times. That is, training data involves 600 negotiation sessions (i.e., epochs). Note that our agent plays both sides (600 times for England and 600 times for Zimbabwe); it makes 1200 negotiation sessions in total. The deadline of the negotiations is 180 seconds. If agents cannot reach an agreement until the deadline, they receive a zero utility.

In our setup, the opponent agent employs the *Gahboninho* [27] negotiation strategy. This agent is selfish and stubborn. At the beginning for a certain period of time, this agent insists on making bids with the utility of 0.9 utility. Afterwards, the agent becomes more selfish and hardheaded and at last moments Gahboninho concedes to avoid disagreement.

Recall that an agent consists of the following components in GENIUS:

- **Bidding strategy**: A model that maps the flow of negotiation to bids. This model defines the amount of concessions with respect to negotiation flow, target utility, remaining time, discount factor and any extra item which is considered to intervene.

- **Opponent modeling**: A learning approach to model the preference profile of the opponent.

- **Acceptance strategy**: A condition that determines whether to accept the received bid from opponent.

For our agent's bidding stragtegy we pick the bidding strategy of the AgentK [28]. AgentLG [29] and NTFT (i.e., Not Tit For Tat) are used for the opponent modeling and opponent modeling strategy respectively. As the acceptance strategy, our agent implements the RL based acceptance strategy proposed in this work. Our agent starts the negotiation by taking random actions and explores the action space during the negotiation to find the optimal acceptance strategy with respect to time-line. Agent decreases the ratio of exploration and exploitation over time and after a while begins to exploit the actions with maximum rewards (i.e., utility in our case).

In the training phase, we first analyze the utility of the agreements. Figure 31 demonstrates the average utilities of the agreements for our agent negotiating with *Gahboninho* for both profiles. Note that to have a tidy visualization we plot the mean of utility values per each 10 sessions. Recall that training phase consists of 600 session per each profile. It can be seen that the performance of our agent gets better after getting certain amount of experience (i.e. 25 negotiation sessions in this case). The increasing trend in terms of received utility indicates the learning capability of our agent. Furthermore, we can notice that there are some fluctuations on the utilities of agreements especially after the significant rise of the utility. Recall that most of the bidding strategies have a stochastic nature. Therefore, even the same agents negotiate with each other; they may end up with a different negotiation outcome. This fluctuation can be explained due to the stochastic nature of bidding strategies. Another observation is that our agent learned to wait more during the negotiation to get better offers from the opponent over time. Recall that our opponent starts conceding when approaching the deadline.

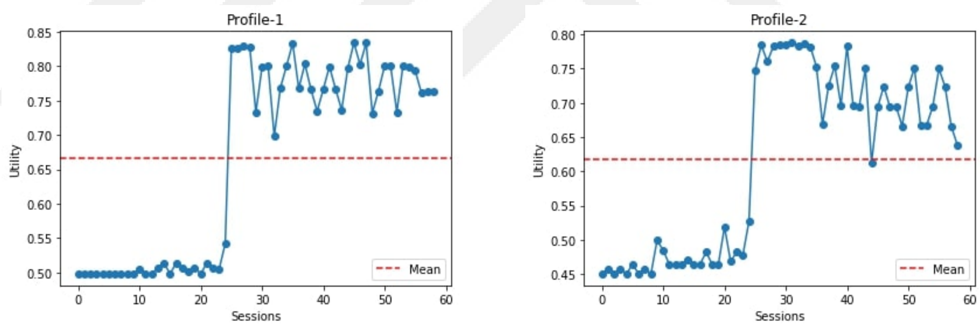


Figure 31: Utility changes in - *RL Acceptance Party*

In Figure 32 the results of K-Means clustering algorithm on the utilities achieved in training demonstrates a significant change in behavior after a time period. Agent begins to take random actions and exploring for actions which can maximize the reward and accordingly the outcome utility, when the agent receives enough amount of experience during the training and discovers the best actions then starts to exploit them and achieve higher rewards and outcome utility respectively. This change as represented in Figure 32 shows the effective learning in agent behavior.

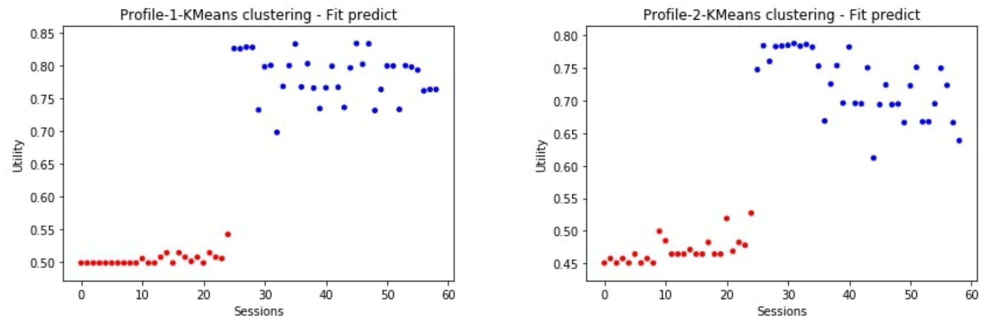


Figure 32: K-Means clustering on utilities achieved in training - *RL Acceptance Party*

As a baseline negotiation strategy, we used an agent which randomly negotiates. Figure 33 shows the average utility changes of the agreements for the randomly negotiating agent. When we compare these results with those received by the RL acceptance strategy shown in Figure 31, it is obviously seen that our agents receives much higher utility.

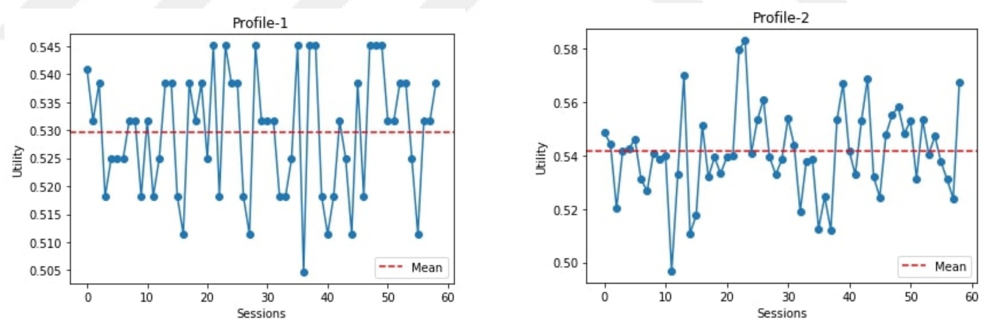
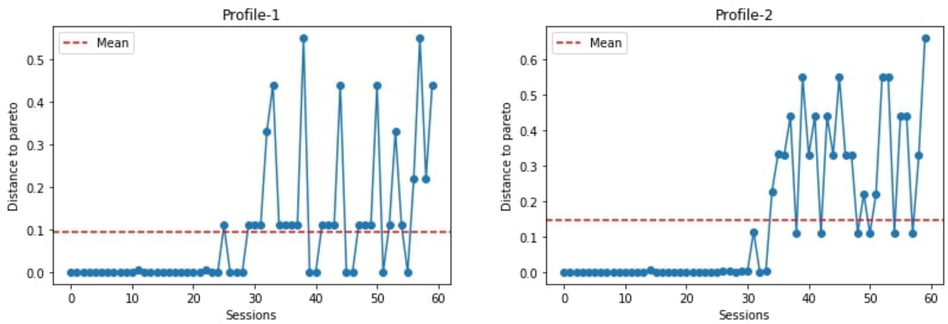


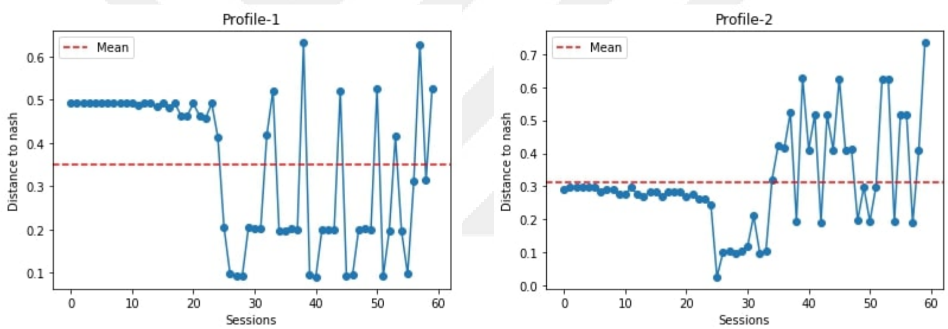
Figure 33: Utility changes in - *Random behavior agent*

Furthermore, we also analyze the negotiation outcome in terms of distance to Nash product solution, and distance to Pareto solutions, and social welfare (i.e., sum of agent's utilities). Assessing the training session results reveals a noticeable change in agent's behavior in terms of Nash, Pareto and Social Welfare metrics. Following results demonstrate the change in agent's behavior:



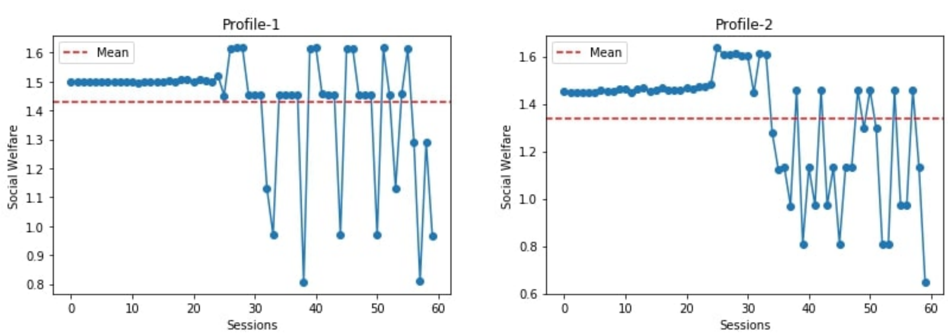
((a)) ((b))

Figure 34: Distance to Pareto



((a)) ((b))

Figure 35: Distance to Nash



((a)) ((b))

Figure 36: Social welfare in agent behavior

6.0.2 Test Session

In automated negotiation, it is important to design agents that can negotiate with different opponents. In order to assess the performance of the proposed RL-based acceptance strategy, in this phase our agent negotiates with different opponents on other negotiation scenarios. For opponent strategies, we pick the following six strategies: *Agent Smith* [30, 29], *Nozomi*[31], *Yushu*[32], *FSEGA* [33], *IAMHaggler* [34], and *Pars Agent* [35].

These six agents were selected as opponents which were among the top rated agents in the previous years at ANAC [36] competition. A brief description of opponent agents strategies is provided in order to get familiar with their approach:

- **Agent Smith:** This agent models opponent's preferences during the negotiation. It initially makes the best offer for itself (i.e., offer with the maximum utility). Afterwards, it compromises over time towards the interest of its opponent.

- **Yushu:** Using a combination consisting of ten last received bids and an estimation about remaining round, the agent calculates a target utility and makes its offer with that target utility. Note that Yushu also considers the minimum utility value it may accept while making its offers.

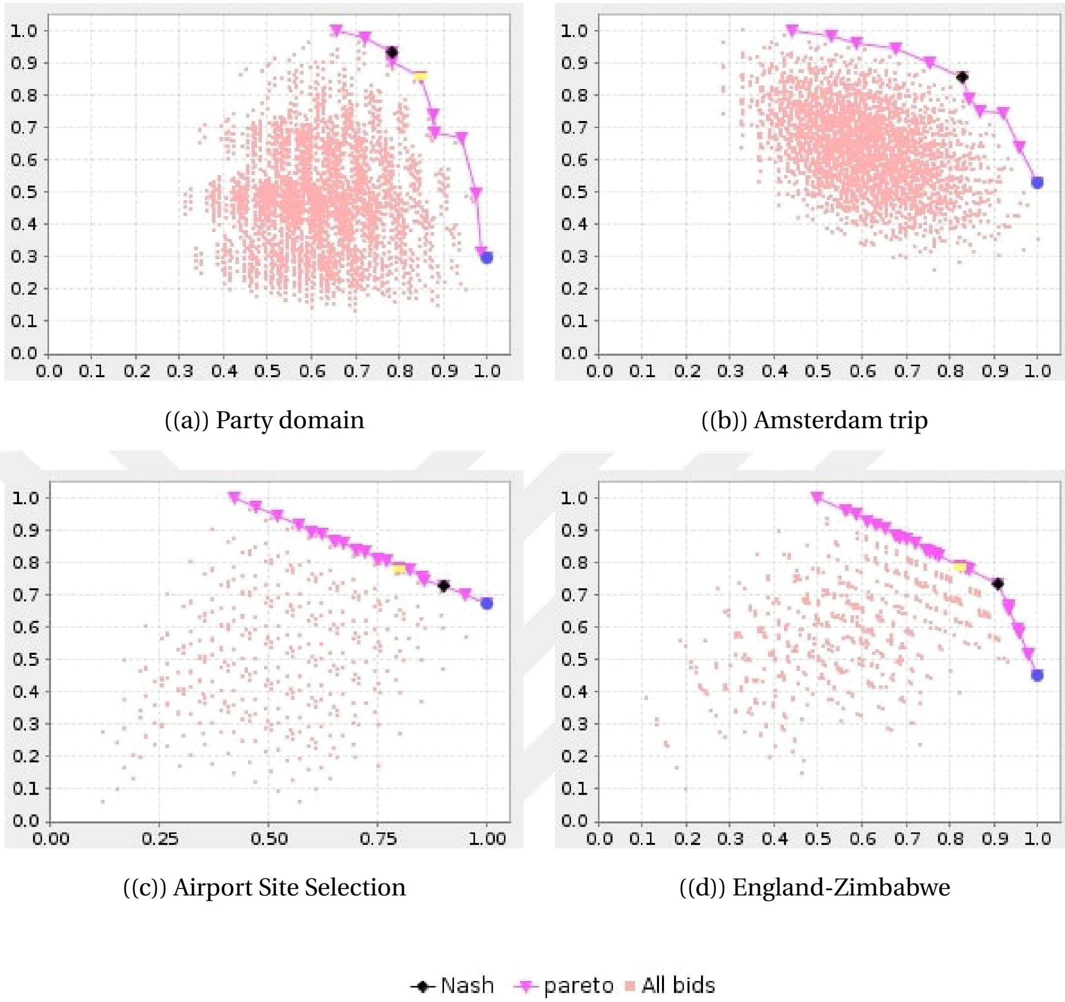
- **FSEGA:** It divides the negotiation into three phases. First 85 percent of the negotiations, it aims to model its opponents by analyzing the exchanged bids. In the second part- from 85 to 95, it does not concede. In the last phase (95-100), FSEGA employs a concession-based strategy due to time limit and accordingly sends bids which are just higher than reservation value. This agent always accepts the best available offer; otherwise, it offers a new bid.

- **IAMHaggler:** This agent constructs an opponent model using Bayesian learning method. As a starting point agent offers a bid with maximum utility and continuously selects a target utility based on various factors such as opponent model, remaining time, and received bid utility.

- **Pars Agent:** Pars agent employs a bidding strategy which is a combination of time-dependent, random and frequency-based strategies to make a bid with high utility which is close to the opponents offers. This behavior increases the possibility of reaching an agreement sooner. This agent has ranked 2nd in the individual utility category in the ANAC2015 [37].

- **Nozomi:** At the beginning Nozomi sends an offer with maximum utility. Based on the opponent model based on opponent's last offer and remaining time, it chooses to compromise or insist.

Three different negotiation scenarios (i.e., party, Amsterdam and airport) are used to evaluate the performance of the proposed acceptance strategy. We choose different domains than the one used in training in order to evaluate the generalization ability of the proposed RL model. The distribution of bids in test domains are demonstrated in (Figure 37(a), 37(b), 37(c)).



((a)) Party domain

((b)) Amsterdam trip

((c)) Airport Site Selection

((d)) England-Zimbabwe

Figure 37: Domain information

We compare the performance of our acceptance strategy with the performance of the AC-next [14] acceptance strategy, which is most widely used acceptance strategy in automated negotiation. When the agents employ the AC-next acceptance strategy, they accept the received offer if its utility is higher than the utility of the agent's coming bid.

In the experiments, we keep the BOA components for bidding strategy, opponent modeling and opponent modeling strategy same for those agents. We only change the acceptance strategy to compare their performance. Each negotiation is repeated 10 times and the deadline of each negotiation is set as 10 seconds. Since each scenario has two profiles, each acceptance strategy was tested on both negotiation profiles to assess the overall performance.

Figure 6.0.2 shows the average utilities of our agent for both acceptance strategies. According to those results, it can be seen that when our agent negotiates with the opponent except FSEGA, the performance of the RL acceptance strategy is almost the same with the performance of the AC-next strategy. Our agent reaches agreements with higher utilities when it employs the AC-next strategy for the Amsterdam trip scenario. When we study the utility distribution for Amsterdam trip domain in Figure 37, it seems that most of the bids are distributed on the right top side of the outcome space. In other words, most of the bids have the high utility for both sides. We observed that in many cases, agents fail to find agreement when our agent employs RL acceptance strategy for Amsterdam trip scenario. On the other hand, the RL acceptance strategy outperforms the AC-next strategy for Party and Airport Site Selection scenario where the bids are distributed over a wide-ranged area.

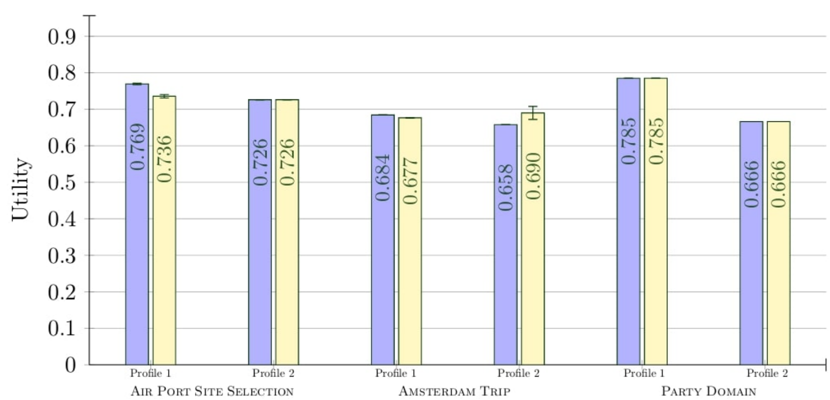


Figure 38: Yushu

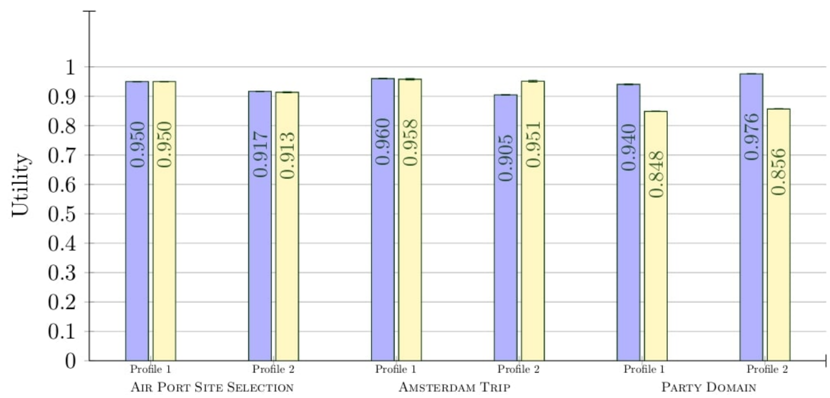


Figure 39: IAMHaggler

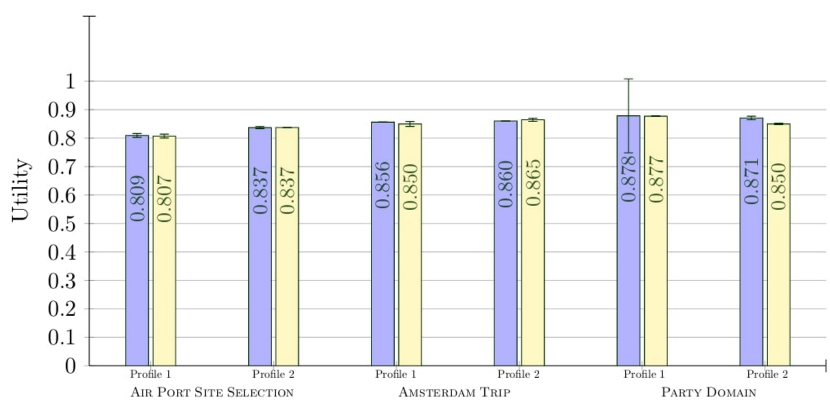


Figure 40: Nozomi

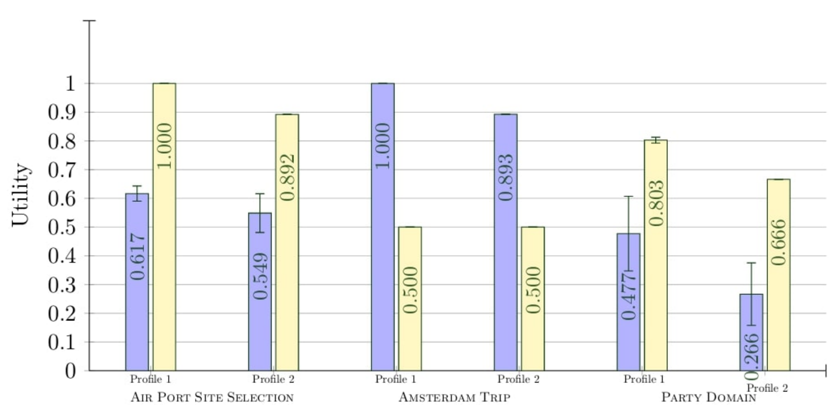


Figure 41: FSEGA

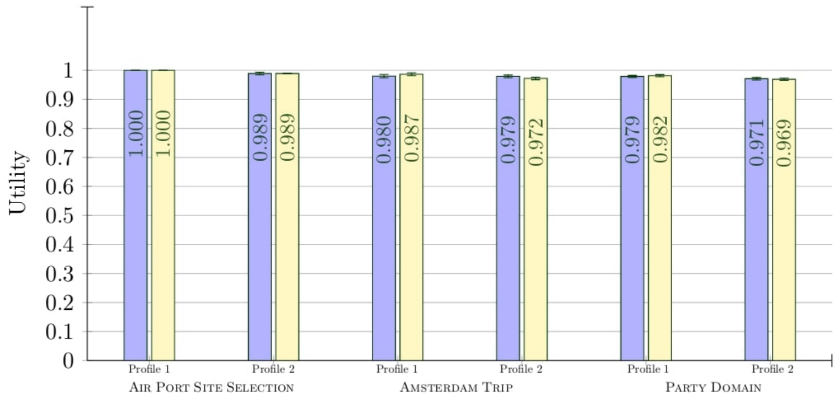


Figure 42: Agent Smith

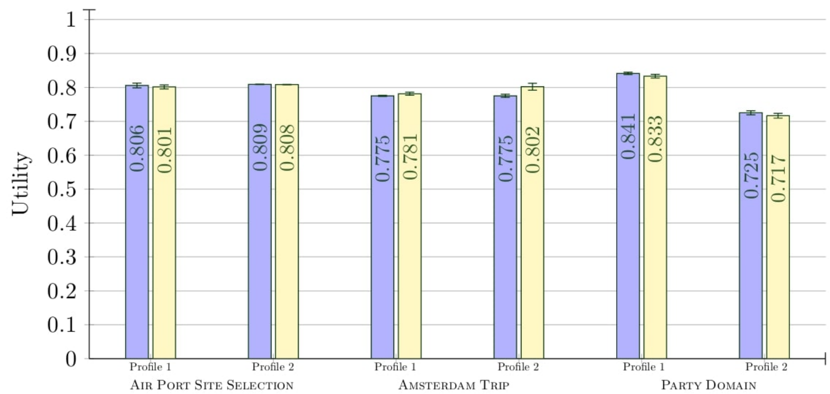
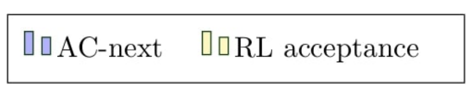


Figure 43: Pars Agent



Test results compared to AC-next acceptance strategy

Chapter VII

RELATED WORK

"The present is theirs; the future, for which I really worked, is mine."

– Nikola Tesla

In this chapter, we review related work in two categories: privacy 7.1 and reinforcement learning in automated negotiation 7.2.

7.1 Privacy

There have been a number of works focusing on service consumer's privacy concern about their personal data, which they need to provide in order to get some services from service providers for invoicing, shipment, etc. [38, 39]. Those works point out that service provider's traditional "take it or leave it" approach (i.e., service consumer needs to provide this information to get the underlying service) or "one-size-fits-all" approach (i.e., acting each service consumer in the same way without considering their sensitivity about their personal information may vary) would have a negative impact on user satisfaction. Therefore, they propose to a more flexible approach based on privacy negotiation.

El-Khatib presents a privacy negotiation protocol where the service provider and consumer negotiates on privacy policy. According to that negotiation protocol, the service provider (i.e.,data consumer in our case) initiates the negotiation with an offer and the service consumer (i.e., data provider in our case) could accept this offer or reject this offer with an explanation why the given bid is rejected. In this set-up, an offer contains how the consumer's information will be used (e.g. shared with other department or only shared with the billing office etc.) and a discount rate as an incentive.

Furthermore, Preibusch models this interaction like a dynamic game [39]. In that study, four types of users have been defined according to their characteristics: users who extremely concern about any use of data, users who only sensitive about their financial and health data, users who only concerns about personal data like address, phone, credit card, and users who do not care about sharing their data. Service provider initiates a negotiation with the service consumer if the consumer does not belong to the first category. That is, the provider negotiates if and only if there is a room for negotiation. Similar to the other work, service provider makes his offers and service consumer accepts or rejects them. The main difference is that the negotiation does not end when the service consumer accepts the given offer. It continues until one of the parties ends the negotiation. Although both studies present more flexible way of building privacy policy than the traditional approach, service consumer still is not as powerful as service provider is. Service consumer can only accept or reject an offer. On the other hand, they have the same bargaining power in our framework.

A more recent study also proposes a negotiation scheme for permission management [40]. Baarslag *et al.* suggests following a negotiation in which the service consumer (i.e, data provider) makes a partial offer regarding shared data and asks service provider to complete this partial offer with the remaining issues such as price discount. The service consumer may accept the given complete offer or make another partial offer. In the proposed framework, asking for completing the partial offer has a penalty (cost) for service consumer; in this way, the service provider avoids to reveal its entire cost structure. In contrast to previous works, service consumer has higher bargaining power but not same as the service provider since complete offers are always made by the service provider. That is, the service consumer may not ask for more discount or any other incentive. In addition, it may avoid making more partial offers and asking the service provider to complete the given bids since there is a cost associated with this process. From this perspective, our work differs from this study. Furthermore, our ways of evaluating the complete offers are different. In that study, an agent's preferences are represented by means of additive utility functions and the given cost is subtracted from this utility. Implicitly it is assumed that there is no preferential interdependencies exists between issues. However, we believe that there might be such interdependencies. For example, the evaluation of how the data used may depend on data type. Therefore, in our work we consider a number of factors to evaluate a bid such as the secrecy level of the given information, the risk of sharing it as well as the gained profit from the received incentives. In our work we only focus on the information type and the given incentives whereas they also consider other issues such as how the data will be used etc. That would be interesting to extend our work in that direction.

There are also other works focusing on detecting the privacy violation rather than preserving such as PRIGUARD [41] and PROTOSS [42]. Those works are complementary to our work. After negotiation, our information sharing framework may check whether both parties act in line with their agreements. In case of violation of the agreement, the agent violating the agreement may get penalized (e.g. a low reputation is assigned to that agent and so on).

7.2 Reinforcement Learning in Negotiation

There exist some attempts in order to establish a reciprocal information and service framework which can ensure both the data privacy of users and commercial concerns of the companies. These works rely on the two well known negotiation approach named as *take it or leave it* and *one size fit all.* both of those methods have a negative effect on user satisfaction since of ignoring user sensitivity and making data sharing compulsory by means of leaving no choice to get the service rather than sharing data.

El-Khatib [38, 39] presents a negotiation protocol for privacy in which the company and customer can negotiate on privacy policy. According to their protocol, the company initiates the negotiation with its offer and then the customer just can accept or reject this offer with a comment describing the reason of rejection.

Preibusch [39] models negotiation as a dynamic game. They categorize users in four groups with respect to their behaviour toward their data. First group are the users who extremely concern about any use of data, second group are the people who just are sensitive about their financial and health data. Third group consists of the people who just take into account the personal data like address, phone, credit card, and the last group are the ones who do not care about sharing their data. Service provider initiates a negotiation with the service consumer if the consumer does not belong to the first category. The company negotiates if and only if there is a room for negotiation. Like the other previous works, company makes offers and customer just accepts or rejects. Negotiation does not end when the customer accepts the given offer, it continues until one of the parties ends the negotiation. Customer still is not as powerful as company while both studies present more flexible way of building privacy policy than the regular methods. Customer can only accept or reject an offer. On the other hand, they have the same bargaining power in the framework presented in this thesis.

Baarslag *et al.* proposes a negotiation scheme for permission management [40]. They suggest following a negotiation in which the customer makes a partial offer regarding shared data and asks service provider to complete this partial offer with the remaining issues such as price discount. The customer may accept the given complete offer or make another partial offer. In the proposed framework, asking for completing the partial offer has a cost for customer; in this way, the company avoids to reveal its entire cost structure. While in this framework the customer has more bargaining power but again it is not equal on both sides. In addition there is a cost for customer when asking to complete an offer. In this thesis the focus is only on the information type and the incentives whereas they also consider other issues such as how the data will be used etc.

Lihong Chen *et al.* [43] applied Q-learning algorithm on their approach to improve the efficiency of negotiation. Their approach uses Q-learning to generate optimal negotiation strategy dynamically. Also there are differences regarding their experimental setup, in our case the agents get in negotiation directly with each other but in their approach there is mediator agent which makes offers to both buyer and seller party.

Papangelis *et al.* [44] used RL to learn multi-issue negotiation policy to design an agent which could confront with humans. They applied Q-learning with function approximation and to handle the vast state space, they consider different feature-based representations of state and action space. They have trained their agent against a simulated user (SU) which is a hand-crafted negotiation agent based on agenda paradigm(Rudnicky and Xu, 1999) [45] For the RL-based agent, their reward function is similar to ours; they give penalty if no agreement reached before the deadline, if agreement is reached the agent receives reward based on the values of final offer and the agent's preference. They used Q-learning to find an optimal policy, like our case they used *GLIE* approach to explore more when the agent is naive, and then exploit more as the agent learns. They evaluated their model by running 20000 episodes and analyzed the scores of agents obtained at each episode after negotiation ends. They have tabulated the success percentages which means best values for all issues involving in negotiation, where the RL-based agent got more than or equal to %35, %65 and %100. They also asked human raters to rate which agent (agenda-based or RL-based) performs better by providing the negotiation transcripts. According to their results, RL-based agent outperformed the agenda-based agent. Unlike our environment, their environment is dynamic; deadline and agents change during negotiation session, and utility values can fluctuate during negotiation. In our model however, deadline does not change arbitrarily and utility values of negotiated issues does not change during the negotiation.

Zou *et al.* [46] integrated genetic algorithm and reinforcement learning to determine an optimal strategy in negotiation, their motivation for applying this kind of technique is confronting with uncertainty issued from incomplete information about the opponent.

They stated that their technique achieves better results in case of efficiency, fairness and strategy convergence. also they mentioned that their approach achieves higher reward, shorter negotiation time and lower degree of greediness comparing to classical evolutionary models. Kröhling *et al.* [47] studied determining negotiation strategy of automated negotiation agent during negotiation. They introduced a conceptual entity as oracle which can be queried by its agent. Agent tells the current context during the negotiation and the oracle estimates the utility value of each strategy using Q function.

Rodriguez *et al.* [48] studied bilateral negotiations with emphasis on the electricity market energy contracts. They introduced context aware Q-learning approach for energy contracts. By using context aware Q-Learning they estimated the utility value of contract prices in order to prioritize opponent agents regarding the maximum outcome utility in a possible negotiation. They have deployed two phases of negotiation in their approach. First phase is the pre-negotiation and the second phase is the main negotiation. In their approach the opponent agents are prioritized for negotiation based on pre-negotiation results. Their approach concentrates on the energy domain however, our proposed strategy can perform in any negotiation regardless of domain.

Sunder *et al.* [49] stated that they have developed an agent which is able to negotiate on contracts in industrial scenarios. They leveraged *multi-agent reinforcement learning* to train two agents negotiating with each other. Their agents learn to show consistent behavior towards each other based on this approach. They developed another agent using reinforcement learning named as *meta agent* which learns to negotiate by negotiating with the agents trained before. They evaluated the meta agent in negotiations with human opponents.

Bakker *et al.* [50] focused on the offering strategy using the same BOA framework as we used in our study. Their reward function acts similar to our reward function, their function returns zero for transition steps that result in exchanging counter offers. In case of terminal states their function returns the utility value of the agreement as the reward. Their contribution is not limited to developing a offering strategy in addition they have developed a general RL framework to be used in BOA framework for autonomous negotiations.

In our study on the other hand, we use Q-learning to estimate the utility values of accepting or rejecting an offer for our party regardless of the party which our agent negotiates with. Because of the fact that our opponent may have various behavior strategy, we trained our agent by generating random state action pairs to help our agent generalize what a good or bad offer is. Thus, our study aims to make our automated negotiation agent learn an effective acceptance strategy.

Chapter VIII

CONCLUSION

> "To know that we know what we know, and to know that we do not know what we do not know, that is true knowledge."
>
> – Nicolaus Copernicus

In this work, we introduce a negotiation-based privacy preserving information sharing framework, in which data consumers offer some incentives in exchange for being authorized to store and use data provider's personal data. Different from other existing framework, the data provider (i.e., service consumer in e-commerce) has the same bargaining power with the data consumer (i.e., service provider in e-commerce). They can both make offers and accept an offer if they like it. Accordingly, we have developed a domain and preference elicitation tool for the proposed framework and illustrated it on a case study for the information sharing procedure between a telecommunication company and their customers.

The negotiation is about to decide what types of information to be shared with the data consumer and what incentives to be received by the data provider, how long to share the data and with whom. We conducted a user experiment to evaluate our framework. The analysis of this experiment showed that human participants like the idea of negotiating for information sharing policies rather than giving only a binary option type of answer (i.e.,accept/reject) to company's requests on sharing their information. It is observed that people are sensitive about their GPS and call logs while they are easy to reveal their age, marital status and educational level. In the user experiments, most of the participants reached an agreement before the given deadline - less than 10 minutes where the deadline is 15 minutes. As a future work, the data requester agents may provide their use-intention and some arguments to convince the data provider. The proposed model can be enhanced to support the needs of more variety of business cases.

Furthermore, this thesis proposes an acceptance strategy model for bilateral negotiations based on deep reinforcement learning. This model can be used as an acceptance strategy module under BOA framework in GENIUS environment. Our approach and generalization method during training could successfully result in a model which can confront various agents in different domains with comparable results in test session. Comparing to other studies done in this domain we can claim that the ability of engaging in negotiation and achieving comparable results regardless of the training opponent and domain is our significance. Experiment results showed that RL acceptance strategy performs at least as well as the AC-next strategy, which is the state of art acceptance strategy in automated negotiation.

As a future work we consider to design models for the bidding strategy based on the knowledge gained in this study. We aim to leverage *transfer learning* to develop a bidding strategy based on knowledge of acceptance strategy. Our motivation is that an agent which consist of strategies developed using same approach may behave more consistent during negotiation. Another idea which we are working on it is the integration of this acceptance policy in a Human-Agent negotiation environment as a module for the opponent agent confronting human users. We are interested in measuring the applicability of such learned behavior in Human-Agent interaction.

Bibliography

[1] M. J. Culnan and P. K. Armstrong, “Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation,” *Organization Science*, vol. 10, no. 1, pp. 104–115, 1999.

[2] M. C. Becker and T. Knudsen, “The role of routines in reducing pervasive uncertainty,” *Journal of Business Research*, vol. 58, no. 6, pp. 746 – 757, 2005.

[3] T. Dinev and P. Hart, “An extended privacy calculus model for e-commerce transactions,” *Info. Sys. Research*, vol. 17, pp. 61–80, Mar. 2006.

[4] R. Mary J. Culnan and J. Bies, “Consumer privacy: Balancing economic and justice considerations,” *Journal of Social Issues*, vol. 59, no. 2, pp. 323–342, 2003.

[5] H. Li, R. Sarathy, and H. Xu, “The role of affect and cognition on online consumers’ decision to disclose personal information to unfamiliar online vendors.,” *Decision Support Systems*, vol. 51, no. 3, pp. 434–445, 2011.

[6] S. Taddei and B. Contena, “Privacy, trust and control: Which relationships with online self-disclosure?,” *Computers in Human Behavior*, vol. 29, no. 3, pp. 821 – 826, 2012.

[7] A. Acquisti and J. Grossklags, “Privacy and rationality in individual decision making,” *IEEE Security and Privacy*, vol. 3, pp. 26–33, jan 2005.

[8] S. Barnes, “A privacy paradox: Social networking in the united states,” *First Monday*, vol. 11, no. 1, 2006.

[9] S. Barth and M. D. de Jong, "The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – a systematic literature review," *Telematics and Informatics*, vol. In press. Available online 28 April, no. 7, pp. 1038–1058, 2017.

[10] R. Aydoğan, P. Øzturk, and Y. Razeghi, "Negotiation for incentive driven privacy-preserving information sharing," in *International Conference on Principles and Practice of Multi-Agent Systems*, pp. 486–494, Springer, 2017.

[11] A. Rubinstein, "Perfect equilibrium in a bargaining model," *Econometrica*, vol. 50, no. 1, pp. 97–109, 1982.

[12] R. Aydoğan, D. Festen, K. V. Hindriks, and C. M. Jonker, "Alternating offers protocols for multilateral negotiation," in *Modern Approaches to Agent-based Complex Automated Negotiation*, pp. 153–167, Springer, 2017.

[13] R. Aydoğan, O. Keskin, and U. Çakan, "Let's negotiate with jennifer! towards a speech-based human-robot negotiation," in *Ito, T., Zhang, M. and Aydoğan (eds), Advances in Automated Negotiations*, 2019.

[14] T. Baarslag, K. Hindriks, and C. Jonker, "Effective acceptance conditions in real-time automated negotiation," *Decision Support Systems*, vol. 60, pp. 68–77, 2014.

[15] R. Lin, S. Kraus, T. Baarslag, D. Tykhonov, K. Hindriks, and C. M. Jonker, "Genius: An integrated environment for supporting the design of generic automated negotiators," *Computational Intelligence*, vol. 30, no. 1, pp. 48–70, 2014.

[16] C. M. Jonker, V. Robu, and J. Treur, "An agent architecture for multi-attribute negotiation using incomplete preference information," *Autonomous Agents and Multi-Agent Systems*, vol. 15, no. 2, pp. 221–252, 2007.

[17] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, M. J. Wooldridge, and C. Sierra, "Automated negotiation: prospects, methods and challenges," *Group Decision and Negotiation*, vol. 10, no. 2, pp. 199–215, 2001.

[18] T. Baarslag, E. H. Gerding, R. Aydogan, and M. Schraefel, "Optimal negotiation decision functions in time-sensitive domains," in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, vol. 2, pp. 190–197, IEEE, 2015.

[19] O. Tunalı, R. Aydoğan, and V. Sanchez-Anguix, "Rethinking frequency opponent modeling in automated negotiation," in *International Conference on Principles and Practice of Multi-Agent Systems*, pp. 263–279, Springer, 2017.

[20] R. Aydogan and P. Yolum, "Ontology-based learning for negotiation," in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, vol. 2, pp. 177–184, IEEE, 2009.

[21] T. Baarslag, M. J. Hendrikx, K. V. Hindriks, and C. M. Jonker, "Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques," *Autonomous Agents and Multi-Agent Systems*, vol. 30, no. 5, pp. 849–898, 2016.

[22] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[23] M. Minsky, "Steps toward artificial intelligence," *Proceedings of the IRE*, vol. 49, no. 1, pp. 8–30, 1961.

[24] K. Hindriks, C. M. Jonker, S. Kraus, R. Lin, and D. Tykhonov, "Genius: negotiation environment for heterogeneous agents," in *Proceedings of the 8th International*

Conference on Autonomous Agents and Multiagent Systems-Volume 2, pp. 1397–1398, International Foundation for Autonomous Agents and Multiagent Systems, 2009.

[25] T. Baarslag, K. Hindriks, M. Hendrikx, A. Dirkzwager, and C. Jonker, "Decoupling negotiating agents to explore the space of negotiation strategies," in *Novel Insights in Agent-based Complex Automated Negotiation*, pp. 61–83, Springer, 2014.

[26] H. Raiffa, *The art and science of negotiation.* Harvard University Press, 1982.

[27] M. B. Adar, N. Sofy, and A. Elimelech, "Gahboninho: Strategy for balancing pressure and compromise in automated negotiation," in *Complex Automated Negotiations: Theories, Models, and Software Competitions*, pp. 205–208, Springer, 2013.

[28] S. Kawaguchi, K. Fujita, and T. Ito, "Agentk: Compromising strategy based on estimated maximum utility for automated negotiating agents," in *New Trends in Agent-Based Complex Automated Negotiations*, pp. 137–144, Springer, 2012.

[29] T. Baarslag, *What to bid and when to stop.* PhD thesis, Delft University of Technology, 2014.

[30] N. van Galen Last, "Agent smith: Opponent model estimation in bilateral multi-issue negotiation," in *New Trends in Agent-based Complex Automated Negotiations*, pp. 167–174, Springer, 2012.

[31] T. Baarslag, K. Hindriks, C. Jonker, S. Kraus, and R. Lin, "The first automated negotiating agents competition (anac 2010)," in *New Trends in agent-based complex automated negotiations*, pp. 113–135, Springer, 2012.

[32] B. An and V. Lesser, "Yushu: a heuristic-based agent for automated negotiating competition," in *New Trends in Agent-Based Complex Automated Negotiations*, pp. 145–149, Springer, 2012.

[33] T. Ito, M. Zhang, V. Robu, S. Fatima, and T. Matsuo, *New trends in agent-based complex automated negotiations*, vol. 383. Springer, 2011.

[34] C. R. Williams, V. Robu, E. H. Gerding, and N. R. Jennings, "Iamhaggler: A negotiation agent for complex environments," in *New Trends in Agent-based Complex Automated Negotiations*, pp. 151–158, Springer, 2012.

[35] Z. Khosravimehr and F. Nassiri-Mofakham, "Pars agent: Hybrid time-dependent, random and frequency-based bidding and acceptance strategies in multilateral negotiations," in *Modern Approaches to Agent-based Complex Automated Negotiation*, pp. 175–183, Springer, 2017.

[36] C. M. Jonker, R. Aydogan, T. Baarslag, K. Fujita, T. Ito, and K. V. Hindriks, "Automated negotiating agents competition (anac).," in *AAAI*, pp. 5070–5072, 2017.

[37] K. Fujita, R. Aydoğan, T. Baarslag, K. Hindriks, T. Ito, and C. Jonker, "The sixth automated negotiating agents competition (anac 2015)," in *Modern Approaches to Agent-based Complex Automated Negotiation*, pp. 139–151, Springer, 2017.

[38] K. El-Khatib, "A privacy negotiation protocol for web services," in *Workshop on Collaboration Agents: Autonomous Agents for Collaborative Environments*, pp. 85–92, Halifax, 2003.

[39] S. Preibusch, "Implementing privacy negotiation techniques in e-commerce," in *E-Commerce Technology, 2005. CEC 2005. Seventh IEEE International Conference on*, pp. 387–390, IEEE, 2005.

[40] T. Baarslag, A. T. Alan, R. Gomer, M. Alam, C. Perera, E. H. Gerding, *et al.*, "An automated negotiation agent for permission management," in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 380–390, 2017.

[41] N. Kokciyan and P. Yolum, "PRIGUARD : A semantic approach to detect privacy violations in online social networks," *IEEE Transactions on Knowledge and Data Engineering*, 2016. In press.

[42] Ö. Kafalı, A. Günay, and P. Yolum, "Protoss: A run time tool for detecting privacy violations in online social networks," in *ASONAM*, pp. 429–433, 2012.

[43] L. Chen, H. Dong, Q. Han, and G. Cui, "Bilateral multi-issue parallel negotiation model based on reinforcement learning," in *International Conference on Intelligent Data Engineering and Automated Learning*, pp. 40–48, Springer, 2013.

[44] A. Papangelis and K. Georgila, "Reinforcement learning of multi-issue negotiation dialogue policies," in *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 154–158, 2015.

[45] A. Rudnicky and W. Xu, "An agenda-based dialog management architecture for spoken language systems," in *IEEE Automatic Speech Recognition and Understanding Workshop*, vol. 13, 1999.

[46] Y. Zou, W. Zhan, and Y. Shao, "Evolution with reinforcement learning in negotiation," *PLOS one*, vol. 9, no. 7, p. e102840, 2014.

[47] D. Kröhling, F. Hernández, E. Martínez, and O. J. A. Chiotti, "The importance of context-dependent learning in negotiation agents," in *XIX Simposio Argentino de Inteligencia Artificial (ASAI)-JAIIO 47 (CABA, 2018)*, 2018.

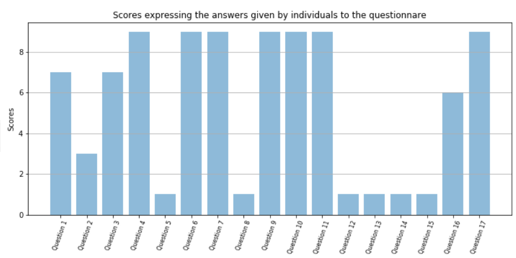
[48] J. Rodriguez-Fernandez, T. Pinto, F. Silva, I. Praça, Z. Vale, and J. Corchado, "Context aware q-learning-based model for decision support in the negotiation of energy contracts," *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 489–501, 2019.

[49] V. Sunder, L. Vig, A. Chatterjee, and G. Shroff, "Prosocial or selfish? agents with different behaviors for contract negotiation using reinforcement learning," *arXiv preprint arXiv:1809.07066*, 2018.

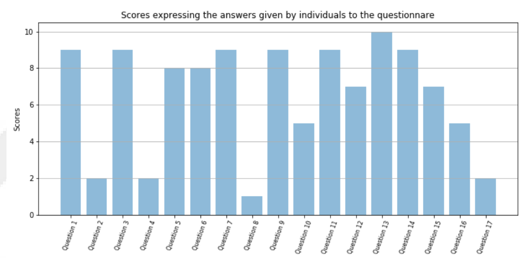
[50] J. Bakker, A. Hammond, T. Baarslag, and D. Bloembergen, "Rlboa: A modular reinforcement learning framework for autonomous negotiating agents," in *Proceedings of the 2019 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '19, (Richland, SC), International Foundation for Autonomous Agents and Multiagent Systems, 2019.

Appendix A

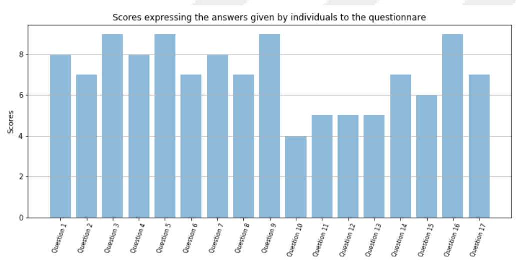
QUESTIONNAIRE



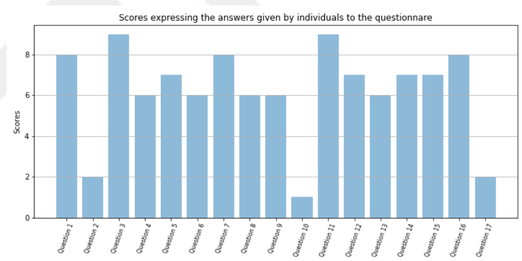
((a))



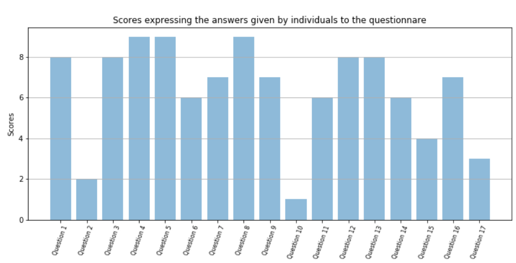
((b))



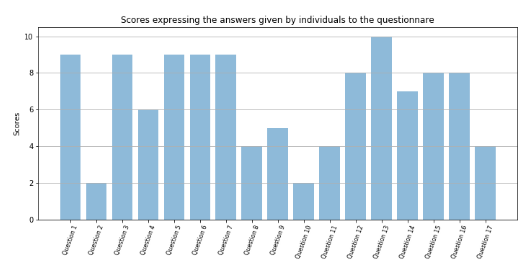
((c))



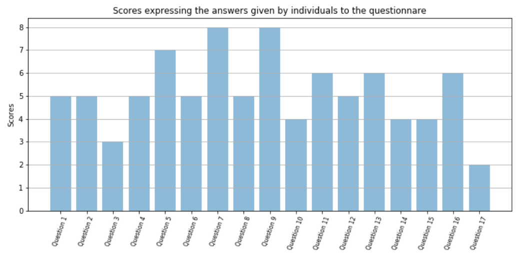
((d))



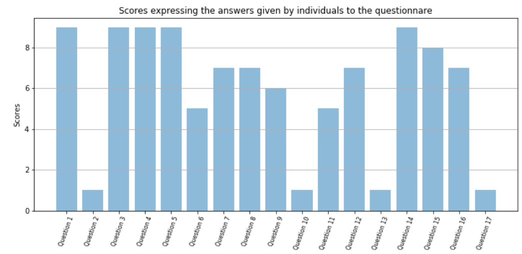
((e))



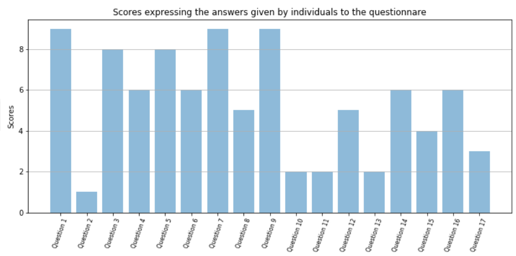
((f))


Scores expressing the answers given by individuals to the questionnare
Scores

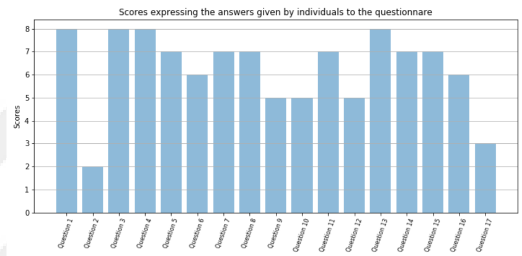
((g))


Scores expressing the answers given by individuals to the questionnare
Scores

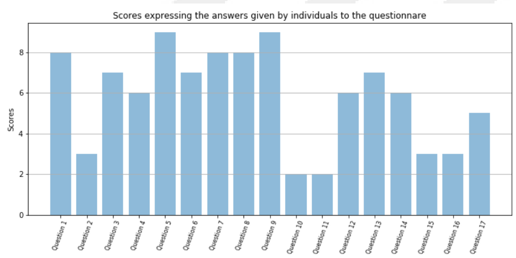
((h))


Scores expressing the answers given by individuals to the questionnare
Scores

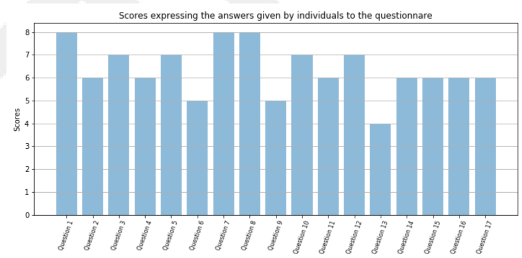
((i))


Scores expressing the answers given by individuals to the questionnare
Scores

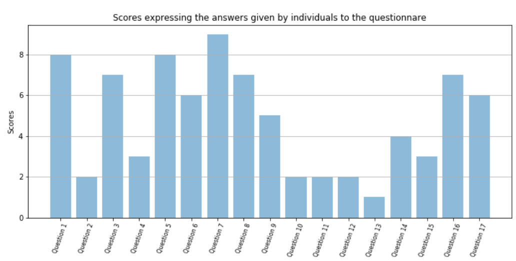
((j))


Scores expressing the answers given by individuals to the questionnare
Scores

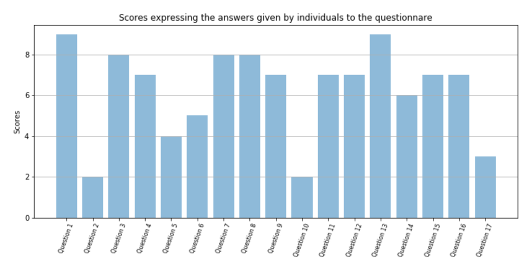
((k))


Scores expressing the answers given by individuals to the questionnare
Scores

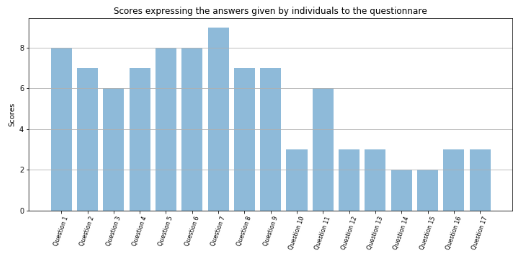
((l))


Scores expressing the answers given by individuals to the questionnare
Scores

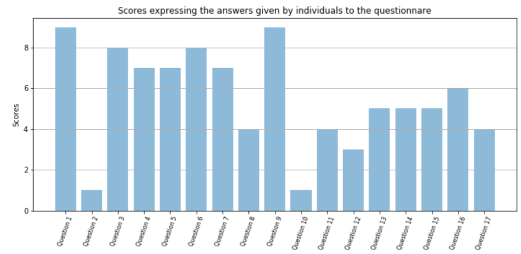
((m))


Scores expressing the answers given by individuals to the questionnare
Scores

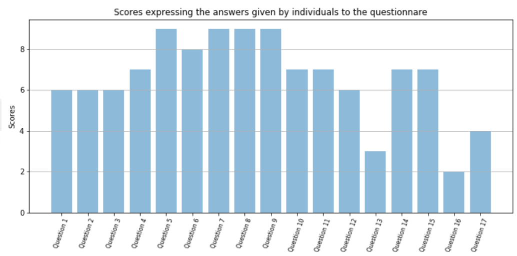
((n))


Scores expressing the answers given by individuals to the questionnare
Scores

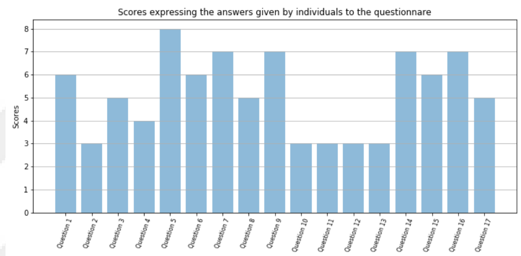
((o))


Scores expressing the answers given by individuals to the questionnare
Scores

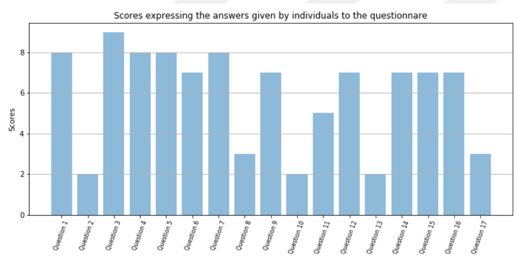
((p))


Scores expressing the answers given by individuals to the questionnare
Scores

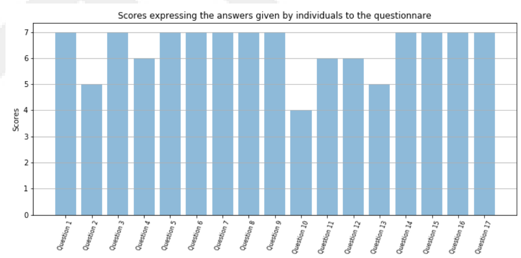
((q))


Scores expressing the answers given by individuals to the questionnare
Scores

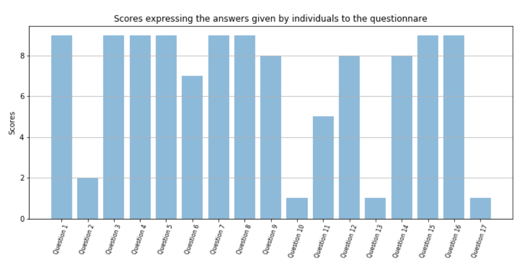
((r))


Scores expressing the answers given by individuals to the questionnare
Scores

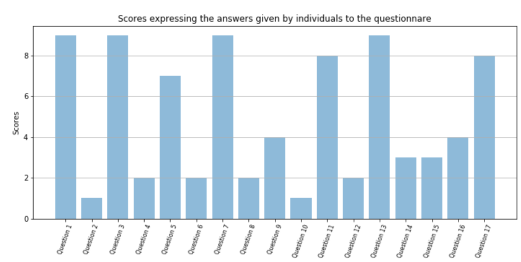
((s))


Scores expressing the answers given by individuals to the questionnare
Scores

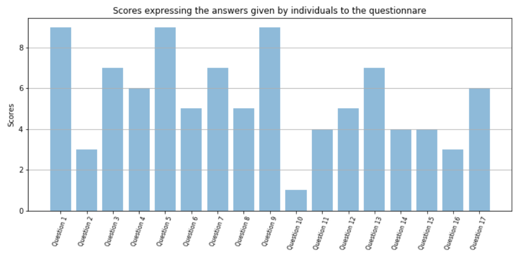
((t))


Scores expressing the answers given by individuals to the questionnare
Scores

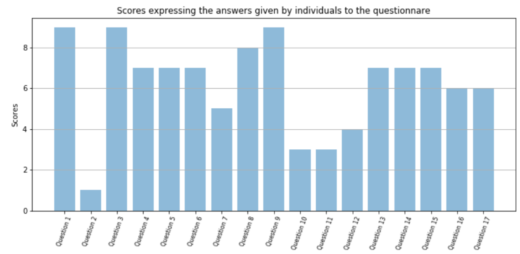
((u))


Scores expressing the answers given by individuals to the questionnare
Scores

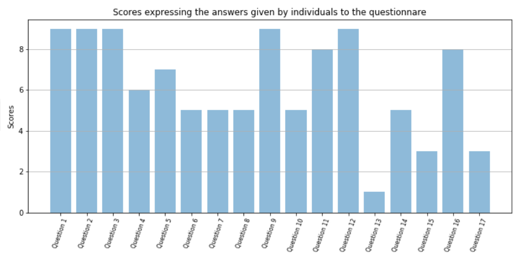
((v))


Scores expressing the answers given by individuals to the questionnare
Scores
0
2
4
6
8
Question 1
Question 2
Question 3
Question 4
Question 5
Question 6
Question 7
Question 8
Question 9
Question 10
Question 11
Question 12
Question 13
Question 14
Question 15
Question 16
Question 17

((w))


Scores expressing the answers given by individuals to the questionnare
Scores
0
2
4
6
8
Question 1
Question 2
Question 3
Question 4
Question 5
Question 6
Question 7
Question 8
Question 9
Question 10
Question 11
Question 12
Question 13
Question 14
Question 15
Question 16
Question 17

((x))


Scores expressing the answers given by individuals to the questionnare
Scores
0
2
4
6
8
Question 1
Question 2
Question 3
Question 4
Question 5
Question 6
Question 7
Question 8
Question 9
Question 10
Question 11
Question 12
Question 13
Question 14
Question 15
Question 16
Question 17

((y))