Bachelor Thesis
”Cultural awareness in social agents in activity generation on a macro scale.”

Alex Klein
Utrecht University
Supervisors:
Frank Dignum (UU), Steven de Jong (TNO), Ruben Smelik (TNO)

January 19, 2016

Contents

1 Introduction

1.1 Background

(Anecdote about culturally specific situations)

In recent years there has been an increased interest in training through simulation for Defence. Because many modern military missions are performed in other countries, training needs to account for the various challenges that come with such missions. To make realistic and effective virtual training situations, these simulations need to reflect the possible environment that these mission may take place in. This includes the physical environment, the people acting in that environment and the behaviour these people have. The latter comes down to the cultural differences between those who perform the mission and the local population. Although work on modelling the physical environment and the population has been done to some detail, cultural affected behaviour is not been researched much. Although some research has been done on cultural affected agent-based systems, these are normally made for specific situations, with a limited number of agents [11] /citeMascarenhas15.

1.2 Research Objectives

In this thesis a model is proposed for culture based activity generation. The culture is based on Hofstede's dimensional model of culture. The simulation is implemented within the IDSA framework, developed by TNO [6]. Agendas are then created based on need-based activity generation that uses utility maximisation to plan activities. I will show how different cultures can affect the behaviour of the agent, by running the simulation with the cultural values of The Netherlands and Portugal and show that this has an impact on the way the agents lower their needs and how they plan their activities throughout a couple of weeks time.

1.3 Research Questions

The main research question is:

- How can culture be incorporated in activity generation?

Other questions that need to be answered are:

- What study of culture is best suited for modelling?
- How can dynamic activity generation be implemented?
- What aspects of culture can be modelled in activity generation?
- How can we generate social networks that it can match different cultures?
- What kind of differences are expected from two different cultures?

1.4 Outline

In the first part existing research is covered: first culture studies and the reasons why Hofstede's model is used for modelling, as well as a description of Hofstede's model. This is followed by a detailed description on the IDSA system. Its components and the data and algorithms are covered. Current issues and some improvements are discussed as well.

Then, some improvements are discussed, necessary to model culture later on. First an improved social network is described, that will improve the social interactions between agents in the simulations. Then the need-based activity generation algorithm is discussed, where agents plan certain activities based on the needs they have. Special cases like needs for households as a whole, social activities and dynamic activities are also discussed. This followed by modelling of culture and how social interactions and activity generation is influenced by culture.

This is continued by a an experiment where we will compare the behaviour of the agent between Dutch national cultural values and Portuguese cultural national values. Results from the two simulations will show differences in planned activities and changes in needs over several weeks simulated time. Finally we come to the conclusion and possible future work.

2 Culture

Culture can be seen as the pattern of behaviour that emerges from the rules that underlie a group of people. As Hofstede [10] describes it: culture lies between people's primitive needs and their own personality. It could be seen as the personality of a group.

This section starts with a general overview of several cultural studies. This is by no means exhaustive, but it gives an overview on work that has been done in the field. The next section describes Hofstede's multi-dimensional model in more detail. In the last part an overview of the modelling of culture for simulations purposes is shown.

2.1 Cultural Studies

Since the 1960's there has been increasingly sophisticated research on culture. This comes from various fields, including sociology, economics and business. E. T. Hall was one of the first to describe culture in a multi-dimensional model [24].

Later, two major publications on culture were done by Trompenaars and Hofstede [10]. They have similar theories, but different aims. Trompenaars is a business consultant and his work is aimed at that. His work is not peer-reviewed and lack verifiable empirical evidence. Hofstede took a more scientific approach, his work is peer-reviewed and quantifiable. Both take the dimensional approach Hall had, just with different semantics.

In 2009 Solomon and Schell [22] published a modern dimensional approach, where most people from various cultures agree on its intuitive correctness. It is inspired by previous work of Hall, Trompenaars and Hosftede, but they reevaluated the dimensions to describe culture. Their work lacks however a description of their approach and the empirical evidence they base their theory on. Although they have quantified national culture to some extend, where every country has a score between 5 and 25 for every dimension, they place these in five buckets. So we would know all countries that have a value between 5 and 9, which is too limited to base a simulation on.

Because of the scientific background and quantifiability, Hofstede's work is the best work on culture for modelling. Hofstede's model is also been adjusted over the years to take new insights on culture into account.

2.2 Hofstede's Cultural Dimensions

Hofstede describes a total of six cultural dimensions [10]. These are power distance, individualism versus collectivism, masculinity versus femininity, uncertainty avoidance, long term versus short term orientation and indulgence versus restraint. The first four are based on a survey held among the employees of IBM in various countries. Long term versus short term orientation results from a similar survey, which was made by Chinese students, to make decrease the possible Western bias of the first survey. Indulgence versus restraint is mostly adopted from Minkov, who researched the World Value Survey.

In the following paragraphs the six cultural dimensions are explained to more detail. Note that the two extreme ends of the dimensions will be discussed, to give an intuitive notion about the meaning of these dimensions. In real world cultures, they are more likely to lay somewhere in between these extremes. Also note that these are used to describe national cultures, not individual people. Individuals may differ a lot from the national values: the national values are the averages of the values of the people within these societies.

2.2.1 Power Distance (PDI)

Power distance refers to the way people deal with authority and power inequality. When the Power Distance Index (PDI) is high, people are more used to are large power distance. This means people will be more acceptable to the fact that their boss, parents or teacher are higher places than themselves. Contradicting a person that is seen as higher instance than you is not done. A quite strict hierarchical order among people means that they also expect differences in chances and resources between the powerful and powerless.

In low PDI cultures, people seek equality among all the people and power is distributed among the people. If there is a hierarchy, it is mostly formal and when people are working together they threat each other equally, regardless of the set hierarchy. For example, workers can call their boss out on mistakes or give suggestions.

2.2.2 Individualism versus Collectivism (IDV)

IDV is about how people define themselves: as an individual or as part of a group, ”I” or ”we”. In a collectivist culture, when IDV is low, people form strongly bonded in groups, which can be family or another form of in-group. Family, friends and partners are predetermined from birth and unquestionable loyalty is expected within a group. Not following the group can be seen as betrayal and opinions of a person should be conform the opinion of the group.The interest of the group is more important than personal interest.

When IDV is high, people show the same amounts of respect to all other people. Time and effort should be spent on friendships to maintain them and throughout people’s lives these friendships may change a lot. People are less connected in an individualist society than in collectivist societies.

2.2.3 Masculinity versus Femininity (MAS)

Masculine societies favour achievement, heroism and material reward. Stereotypes about the roles of men and women in society are strong: men should be macho and women should be caring. People tend to assertive.

In feminine culture people are more tender and caring. Society should take care of people who have it worse. Cooperation is the standard to solve problems and people are looking for consensus and comprises.

2.2.4 Uncertainty Avoidance (UAI)

UAI describes how people handle uncertainty and ambiguity. It also describes how people look on fate: can one control their own life, or does one follow their fate?

When UAI is high, people avoid others who are different to them and situations they are uncomfortable or unfamiliar with.

In low UAI cultures people are more relaxed and more likely to do seek out new experiences and take a risk. Skills are more important than talent and can be improved by practice.

2.2.5 Long Term versus Short Term Orientation (LTO)

LTO describes how people perceive time. In long term oriented cultures people want result in the far future, even if that means that their will not be much gain on the short run. They also make more use of knowledge from the past to solve modern problems and traditions are seen as essential.

In short term societies, actions should have results fast, otherwise they are seen as not working. Financial profits are expected within a short period of time and problems should be solved in a quick fashion, even if that implies a similar problem will occur later on.

2.2.6 Indulgence versus Restraint (IND)

IND describes how people deal with their (primitive) needs. In indulgence cultures people will be comfortable fulfilling their needs. They tend to be more out going and sexuality is more accepted. The aim of life is on having fun and enjoying life.

In restraint cultures there are strict rules on what can and cannot be done to deal with one's needs and strict social norms about what behaviour is accepted.

2.3 Agent-based Culture Models

There has been some research in modelling of cultural aware agents in the recent years, using Hofstede's model.

Hofstede et al. [11] modelled negotiating agents that make different decisions based on their national culture. They used utility functions for the decision process, which was influenced by the cultural values. They showed how the decision making of the cultural aware agents was in line with the expected behaviour based on Hofstede's multi-dimensional model.

Mascarenhas et al. [14] showed how agents with different cultural values treat each other differently, using the Belief-Desire-Intent model to represent knowledge of the agents. They showed how these agents from different cultures react to someone they do not consider part of their "in-group": the people they consider part of their group, as part of their "we".

Schram [20] also used the the BDI model to model different reactions and effects to various actions among different cultures. The build a simulation where agents' beliefs are updated in reaction to actions of others, like leaving food on a plate. Within this simulation, different cultures have different reactions to such actions.

3 The IDSA System

IDSA stands for Intend Driven Scenario Authoring [6] [13] and was developed at TNO, at the Modelling, Simulation and Gaming group. IDSA had several goals:

- Model a world and population based on data;
- All agents should have an agenda for a day;
- Non-scripted, dynamic events can be placed by the user. Agents will react to these events and will return to their normal behaviour, when the event is over.

IDSA is a multi-agent simulation where people will do their every day life, like going to work, bringing their children to work and doing groceries. This everyday life can be interrupted by events that are placed by the user. The system will construct these events at the moment the user places it: they are not scripted and should be different every time. Only one day is simulated each time, so an agenda is planned for one day only. For the demo the Dutch municipality of Rijswijk was modelled, which will also be the test-case for this thesis.

The IDSA can be seen as three parts: the simulation, the pre-processor and the event planner. The simulation takes care the behaviour of the agents at run-time, the visualisation and user input. The pre-processor consists of the world data model, the population generator and agenda planner. The world data model will construct the world from data, the population generator will model the population and their properties based on data and the agenda planner will give all the agents their agenda for that day. The event planner translates requests for events from the user to actual behaviour of the agents. It does this in two steps: first a planner makes a sequence of actions that need to be performed for the event. Second, a sampler looks for agents that need participates in the event.

The following sections will the describe the IDSA system in more detail. The first three sections cover the pre-processor: the world data model, population generator and agenda planner, respectively. Then the event planner is discussed in more detail and finally some improvements that have been added to make it more suitable for future modelling.

3.1 World Data Model

The environment model, which contains buildings, infrastructure, and outdoor areas such as parks and water, is derived from geospecific open data sources, such as OpenStreetMap, and includes a functional description of each building or public area, which is represented in a hierarchy that features multiple inheritance. For example, a shop is inside, is a place where people work, and is a place where people can buy products.

Each building or area also includes relevant environmental properties such as housing capacity, taken from the Dutch building register. In cases where data is insufficiently detailed, incomplete, or inconsistent, it is estimated based on heuristics (such as the size of a building), or, in some cases, manually provided.

3.1.1 Representation

The world is represented as a graph, where edges are roads where the agents can walk, and nodes connect these edges. Some nodes may be linked within areas, which can be any place that is of interest, like a house, shop, park or workplace. Most areas will only have one node, like houses and workplaces, but some have multiple nodes in their area, like parks where people can walk around. These areas will have a list of functions, as one area can have multiple functions, like a shop can be a workplace and a place to get groceries, but an area can also have multiple of the same function, because one area covers for example six apartments. These functions will have some properties, like surface area and capacity, which reflect the maximum number of inhabitants or workers.

3.1.2 Data

The world is based on data from OpenStreetMap and BAG (Basisadministratie Adressen en Gebouwen) data, which is the Dutch building register. We use OSM to construct the graph, where the edges are the streets. Then OSM and BAG are combined to load the areas and these areas are then linked to the vertices of the graph. Then functions are added to the areas based on a combination of BAG and OSM. The primary source is BAG, but sometimes BAG lacks a proper labels, so OSM is used to complement BAG data where necessary and possible. When some buildings are still have no function attached to them, these are completed by making an educated guess what the function could be, by looking at the surface area or the surrounding buildings, because if a building is surrounded by only houses in a quiet neighbourhood, it is very likely the unlabeled building is also an house.

3.2 Population Generator

The population generator creates the population based on data and places them in the world. First section will give a description of the agent, then the data that is used for the population is discussed. Then we continue with the algorithm used to model the population and finally the social network is discussed.

3.2.1 Agents

The agents in the IDSA system have a couple properties, being: gender, age, household type and household role. On top of this they have knowledge of who is part of their household and the location of the house. All of these properties are given to them by the population generator. They also have a model stack, which is used to give the agents their behaviour. A behavioural model takes control over an agent and make the agent act out their behaviour. This can be simple behaviour, like just standing still or walking from one place to another, or more complex, e.g. following another agent at a distance. The model stack keeps track of the order the models need to be run. The model at the top of the model stack is the one that is currently running or the one that needs to be started. When a model has finished it is removed from the stack and the next model is started.

Models can be pushed on the stack from two places: an activity in the agenda or an event that needs the agent. Every activity knows which model is needed and an activity is allowed to push its models onto the stack if the agent's model stack is empty. The agent will then do a look up in the agenda to see which activity is next and that activity is then allowed to push its model. Event on the other hand take control regardless of the models already on the model stack. As soon as the agent is sampled for the activity, new models are pushed on to the model stack and the event's models are therefor started to make the agent act in the event. When the models of the event are finished and popped from the model stack, the old models are checked for relevance. If the model can no longer be acted out, for example when the activity they belong to is already over or the agent cannot reach the activity on time, it is popped from the model stack.

3.2.2 Data

To model the population of Rijswijk, data from the Dutch Central Bureau of Statistics is used. This includes multiple data sources from the CBS. First, data about the local population is used. These are data tables about neighbourhoods, which covers up to a couple of thousand people. This includes fields for, but is not limited to, the number of males and females in neighbourhood, percentages for age categories and average household size. However, it lacks detailed conditional information, like how many men are in their twenties. We do know P(gender = male) and $P(15 < age < 25)$, but $P(15 < age < 25|gender = male)$ is unknown. This information is not available for the neighbourhoods, out of privacy considerations. Another issue with this data is that some values are rounded. For example, because of rounding of the values, the number of males plus the number of females might not be equal to the number of inhabitants.

To add this information we will assume that the national data about this, which is available for use, is sufficient to bridge this gap in the data. Now we do know the conditional probabilities, but at the cost of maybe having information that is not completely in line with the real world. We will try to make this error as small as possible, as we will discuss later.

The national data contains two datasets that we will use together to make the households. The first dataset contains numbers for all combinations of gender, age category (in groups of ten years), household type and civil status. Household type can be Pair if they have children, In a relationship if they do not have children, Single parent or Single. The second dataset contains the number of children given the age and household type of the reference person. If used together, we can construct one person from the first dataset and if necessary draw how many children this agent has and use the first dataset again to draw their partner and children.

3.2.3 Algorithm

There are two main algorithms found in the literature for population synthesis: Synthetic Reconstruction (SR) and Combinatorial Optimisation (CO).

SR will first construct one large data table that combines all the available data and estimates a probability distribution from that data table. The unknown conditional data is estimated using Iterative Proportion Fitting. It does this by iteratively estimating unknown fields between two data tables, which will convert to stable values.

Co first uses the the data on the largest scale, e.g. national, and constructs a population based on that large scale data, by doing Monte Carlo draws from the data. CO then uses the more local data to optimise the results from more general data to the local data.

We have chosen to use CO, because it gives more realistic results in general, at the cost of more computational time. Our population of 30 thousand people is relatively small, so the computational cost was not much of an issue. In our simulation we first use the national data to draw households, so we have agents with an age, gender, household type (e.g. Pair, Single Parent or Single) and their household role (mother, father, child or single). We make five times more agents than we want to fit in the world, so also keep a pool pf households that we can use for optimisation. Then all the houses get filled with households that fit within the capacity of the house. The next step is to optimise the population to the local data of the neighbourhoods. Per neighbourhood we keep track of the total absolute error (TAE), which takes into account differences in number of inhabitants, males, females and five age categories. A random household is chosen within the neighbourhood and one within the household pool and we swap these households if it lowers the TAE. After a set number of iterations (1,000,000 per neighbourhood) we stop optimising. This gives us optimal results for all the neighbourhoods, although it does not gives us TAE of zero everywhere, because of the rounding that has been done in the neighbourhood data.

3.2.4 Social Network

Between the agents is a social network, which indicates which agents are friends and could do activities together. The social network is constructed based on the work of Toivonen et al. [23]. The algorithm uses five steps:

1. Make an initial network of N_0 agents;
2. Find for an agent that is not yet in the network an average of m_r initial contacts, where $m_r \geq 1$;
3. Then find it to on average m_s secondary contacts. Secondary contacts are contacts of the initial contacts, where $m_s \geq 0$;
4. Connect the agent to the initial and secondary contacts and connect the contacts to the agent;
5. Repeat step 2 tot step 4 until all agents are in the network.

Agents are explicitly connected to their household members. This approach has nice properties, because we can easily adjust the average number of contacts and the clustering by adjusting m_r and m_s respectively.

However, this algorithm to generate the social network leaves out too much detail and does not take into account semantics in our world model. People do not know their neighbours or their coworkers (at least, it is very unlikely) and no information about the agents themselves. In the real world, similar people are more likely to become friends, where in this case we could use age and gender for example. Later we will discuss improvements made to the social network to include these semantics.

3.3 Agenda Planner

When the world has been constructed and the population synthesised, all agents need to get an agenda for one day. These activities should not overlap and the agent should have enough time to walk from activity to activity.

First we will cover what activities are and how they are planned. We continue with the constraint satisfaction that makes sure activities do not overlap. Third, the multipliers, which control how frequently a certain activity gets planned, are covered. Then the use of the agendas during the run-time simulation is discussed and finally the issues with the current system and what could be improved.

3.3.1 Planning Activities

An agenda is an ordered list of activities, where the activities are ordered by starting time (or ending time, because we assumed that activities do not overlap in time). An activity is a tuple a, l, p, s, e, where a is the name of the activity, l is the location of the activity, p is a group of agent that participate in the activity and s is the start time, e is the end time. The tuple also indicates the order in which the variables are given a value.

The agenda planner has an order in which activities are planned. First mandatory activities are planned, which are, in order: parental duties, work and school, and dinner. After these activities have been planned, all the other optional activities are planned, like taking a walk, doing groceries and visit friends. Parental duties were introduced to cope with the fact that very young children cannot yet do activities on their own and need to be accompanied by some form of supervision. So parents will bring young children to school and pick then up again later. This is planned before even work is plan to ensure that at least one of the parents is available to take care of the parental duties. Then work and school is planned. We will refer to school and work as work, because in the system they are planned in a similar fashion. Some agents will not have these planned, because they are retired or just have a day off. Then dinner is planned, if the agent does not already have plans around dinner time. Then all the other activities are planned around these mandatory activities.

To plan a specific activity, all the variables in the tuple a, l, p, s, e are instantiated. We pick a certain activity to plan. How these are picked will be covered under the multipliers. So we already know a. Then a location l is selected. Normally a search through the graph is done to find nearby locations that fit the description given by the activity we are planning, e.g. a shop when an agent is planning to do groceries or a workplace when planning work. A timeframe that is still available is picked (timeframes will be covered under constraint satisfaction) for further planning. From the list of possible locations one is picked that is close enough to the location of the previous activity. If no previous activity has been planned yet, the house of the agent is used as the previous location. The same holds for the activity that comes after the activity that is being planned. If more agents should take part in the activity the list of contacts of the agent is used to search for available people, otherwise one agent is the only participant. Lastly the start and end time are picked within the selected timeframe, such that the agent can still make it to the activity and to the next activity.

Then the activity is inserted in the right place in the agenda of the agent.

3.3.2 Constraint Satisfaction

When planning activities, each activity that can still be added to the agenda keeps track of the times it can be planned. This done by using constraint satisfaction based on Forward Checking.

Every activity has a list of timeframes in which it can be planned. A timeframe is a tuple ¡s, e¿ where s is the start of the timeframe and e the end. So for example, doing groceries is can be planned as long as the grocery shop is still open, which might be between $8 : 00$ and $18 : 00$, which means that the initial timeframe for grocery shopping is $< 8 : 00, 18 : 00 >$. If another activity is planned on a time that overlaps with the timeframe of the activity, the timeframe is adjusted. This way it is no longer possible to plan two activities at the same time. So, continuing our example, if another activity is planned that starts at $7 : 00$ and ends at $10 : 00$, the timeframe for grocery shopping is shortened to $< 10 : 00, 18 : 00 >$. A special case is when an activity is planned within the the timeframe. This makes that the timeframe has to split. So if an activity is planned from $12 : 00$ to $14 : 00$, the timeframe will split and the grocery shopping will keep track of both timeframes: $\{< 8 : 00, 12 : 00 >, < 14 : 00, 18 : 00 >\}$. Either of those timeframes can be used for planning, but it is not possible to plan any activity outside of its timeframes.

Every activity also has minimum and maximum duration, d_{min} and d_{max}. if a timeframe becomes shorter than $d_{min}$the timeframe is removed from the list of timeframes. If the list of timeframes becomes empty, the activity can no longer be planned.

3.3.3 Multipliers

When selecting activities for planning, there is normally a chance it will be planned (except for parental duties, which must be done by one of the parents). So if an activity is selected and if a generated random number between 0 and 1 is lower than the chance, the activity gets planned. Otherwise the activity is no longer available for planning.

Because of the difference age and gender among the agents, should some agents be more likely and some less likely to plan certain activities. For example, elderly people are less likely to go to work, but might be more likely to go for a walk in the park that day. For this we use multipliers. We can adjust these chances with these multipliers to make agents with certain characteristics more likely or less likely to plan certain activities. A special case is the multiplier of 0, which makes an activity no longer available for planning. This can be used to make sure young children cannot go to work, but have to go to school.

These multipliers can also be used in a global sense. For example, summer will make it globally more likely for agents to go to the park and less likely in the winter.

This is season based. The alternative is more culturally based, like fewer working people and fewer children going to school on holidays.

3.3.4 Use of Agendas

During the simulation, agents will do look ups in their agendas. When an agent's model stack becomes empty, when the agent has finished the previous activity and is not involved in an event, the agent will look for the next activity. The model from the activity will be pushed onto the agent's model stack, which will start and end at the time as stated by the activity, and a walking model to the location of the next activity is pushed on top of that. The agent will then start by walking to the next location and when they arrive they will start the next activity.

The same hold for the first activity of the day. If there is no next activity, i.e. the agent just performed the last activity of their agenda, the agent will push a walking model to their house, where they will stay until the end of the day.

3.3.5 Issues

There were a few issues with this system. The first being the multipliers, because the resulting behaviour was not directly predictable when changing the multipliers. Intuitively, halving the multiplier should half the number of planned activities. But because of other multipliers, for example from the season, and normalising probabilities later among the activities, this was not the case and iteratively doing trial and error to find the right parameters was the best solution.

Another problem with this system is that it is not suitable to support multi-day activity generation. It will be able to generate activities for a next day, but the activities planned will be inconsistent. For example, when work is planned, a location for a job is found semi-randomly. So when we plan activities for the next day, another location will most likely be selected. This problem will in part be addressed in to next section.

Lastly, an issue that comes from the fact that agenda planner was developed specifically in the knowledge that Rijswijk was the use case. This means that the agenda planner works in a very Dutch or Western fashion. Because of this it will give unrealistic scenarios if the IDSA system needs to model different parts of the world. The main focus of this thesis is to lower this bias and make the IDSA system more suitable to also model other areas, populations and cultures than the Dutch.

3.4 Improvements

Since the initial development of IDSA, I have made some improvements to the system to make it better scalable for further expansions. These improvements will be necessary for the additions later described. I will discuss two improvements: first, a dynamic spatial data structure and, second, work allocation.

3.4.1 Dynamic Spatial Data Structure

Within the world data model, there already was a static spatial data structure. This static grid made it fast to look up where nearby building were, which was used for visualisation: agent that entered a building would no longer be drawn, until they left that building. This was a grid with 20m by 20m tiles and in every tile the buildings that are (partially) located in that tile are saved. Later we want agents to be able to meet their contacts in the street, even if it was not planned. So if two contacts are within a certain radius of each other, some form of social interaction will occur.

I have extended the grid, such that the old functionality for the static objects is maintained, and the location of the agents can be dynamically updated. Now every tile has information about both the static objects and the agents that are on that tile. Every time an agent moves, it will send an update to the grid and the grid will remove the agent from one tile and place it to the next tile, of that is necessary. Every agent can also send a request to find all agents that are within a certain range. The agent can then use that to look for the contacts within the agents that are close.

3.4.2 Work Allocation

In the IDSA system, workplaces and schools were selected for an agent when they started to plan a work or school activity. This was good for the IDSA system, because it only simulated one day, so continuity about the workplace or school was not necessary. But if the system is extended to support multi-day simulation, workplace consistency have to be considered. Because of this workplace allocation is moved to the population generator and the workplace/school is added to the properties of the agent. After the agents have their houses allocated, workplace allocation is started. For every household, we give all the adults a workplace and, if there are children, allocate all children to one and the same school. Elderly people have a smaller chance of having work, which becomes smaller as the agent is older than the retirement age.

The heuristic for workplace allocation is:

$$h_{a,w} = h_{a,w}^{dist} + h_{a,w}^{size}, \text{ where}$$
$$h_{a,w}^{dist} = \frac{1}{2+EuclidDist(house_a, w)} \text{ and}$$
$$h_{a,w}^{size} = \frac{cap_w}{\sum_{x \in W} cap_x}.$$

Where a is an agent, w is a workplace, W is the set of all workplaces and cap_w is the capacity of w. For school, only the distance heuristic is used.

After work allocation, all agent have a workplace, a school or are otherwise unemployed/retired and their workplace is therefor null. Because the workplace/school is stored by the agent, it can be used for consistency of workplaces between multiple days.

3.5 Agents and Cultural Values

Agent in the simulation need to have their own set of cultural values. However, we cannot simply assign the national values to the agent as their cultural values, because then we would ignore the fact that a national culture is based on the cultural values of its individuals and the national cultural values are only the averages of these of the individuals.

We will assume that each cultural value is normal distributed, where the mean is equal to the national value. The decision for a normal distribution is that most people will be conform the national cultural values and a minority has values far from the national ”standard”. Each agent will then get a normal distributed random cultural value assigned, where the standard deviation is set to an arbitrary chosen value of 20.

4 Social Network

National culture is for a great part reflected in social interactions between the member of the society. Dignum et al. [7] explains the importance of social connection and interaction between realistic social agents. A realistic social network is essential for realistic cultural-aware agents and the behaviour that is expected of them.

As stated before, the social network model used in the IDSA system, based on the work of Toivonen et al. [23], does not take into account the semantics of the agents, like similar agents being more likely to know each other, or world, like neighbourhoods and colleagues knowing each other. Social network models should, as a network, hold the follow properties citeSnijders01:

- x contacts per agent on average;
- c amount of clustering between the agents.

The exact values for x and c are dependent on what kind of social network we are modelling and which relationships we consider to include in to social network. In most social networks, the clustering is significantly higher that in random networks. Even though Toivonen's model holds nice properties for the resulting social network (it can be specified to reach specific values for x and c), it does this randomly among the agents, without any information other than the agents themselves.

Edmonds [8] showed the importance to link the physical, topological world to the social networks. This link to the physical world, together with agent similarity, can be found in a model Arentze et al. [2]. They use utility maximisation to construct the social network. The utilities will take agent similarity and topology into account and threshold values that will be used as criteria to connect agents. These threshold values need to be estimated, because they are not known a priori.

In the rest of this chapter the social network generation process of Arentze et al. is discussed. First the generation itself and its utility functions, then the process of parameter estimation and the evaluation functions used for that.

4.1 Social Network Generation

The model for social networks assumes that the probability of friendship between two persons depends on the evaluation of the utility the two individual expect to gain from such a relationship. This depends on three elements:

1. Homophily;
2. Geographic distance;
3. Presence of common friends, or transitivity.

Homophily, or similarity between agents, will make it more likely that agents that are similar, e.g. having similar interests, become friends. The probability will decrease when agents life far apart from each other: neighbours and colleagues are more likely to know each other. When geographic distance increases, the probability to be friends should decrease, therefor the utility should decrease. When two persons have common friends they are more likely to know each other, because they could have been introduced or all had some sort of common history together. When common friends are present, the utility should increase. This results in the following utility function:

$$U_{ij} = U_{ij}^{H} + U_{ij}^{D} + U_{ij}^{C} + \epsilon_{ij}$$

Where U_{ij}^{H}, U_{ij}^{D} and U_{ij}^{C} are the utility components for homophily, geographic distance and the presence of common friends, respectively. To account for the non-observable properties that do exist in the real-world, but are not present in the model, the error term ϵ_{ij} is introduced. For now, we will assume that the utility between two agents is symmetric, i.e. $U_{ij} = U_{ji}$. Note that this can be adjusted to model different perceptions on the relationship.

A person can be friends with a limited number of people, because maintaining relationships takes time and effort. To decide who will become friends, there is a threshold on the utility value. When the utility for both people are above the threshold of both people, they are connected:

$$C_{ij} = \begin{cases} 1 & if\ U_{ij} \leq U_i^0\ and\ U_{ji} \leq U_j^0 \\ 0 & otherwise \end{cases}$$

Where C_{ij} is the connection between agent i and j and U_i^0 is the threshold of agent i. The threshold represents both the ”cost” to become friends, but also the opportunity that a person has to has a meeting with another person.

Then, for every combination of two agents the utility is computed and the threshold is applied. The resulting network is the social network.

4.1.1 Clustering

Because social networks have a higher clustering than normal networks, a special mechanism in needed to make this clustering happen. Clustering is defined as:

$$C = \frac{3x\#triangles}{\#triples}$$

A triangle are three nodes that are all connected to each other. A triple is two edges connected by one node. Clustering can be modelled by using a clustering utility component U_{ij}^{C}. However, the mutual friends are not yet known completely during the generation of the network, so U_{ij}^{C} cannot be used during generation of the network. However, clustering can also be modelled in the threshold value of the agent. So a second round is added to the generation algorithm and this time the threshold is defined as:

$$U_i^1 = \begin{cases} U_i^1 - \theta & if\ C_{ij} = 1 \\ U_i^1 & otherwise \end{cases}$$

Where θ is a threshold lowering constant that favours clustering. The first round can also be seen the same as the second round, but with $\theta = 0$. Because of this, the second round only changes utilities for pairs of agents who have mutual friends, so only these cases have to be considered. It is possible to do more rounds, similar to the second round. This will not be done in the simulation discussed here, because a high enough clustering can easily be reached with one additional round.

4.1.2 Utility Functions

De daadwerkelijke implementatie can de utility components.

4.1.3 Computational Demands

With the simulation there are $30,000$ agents, which need to be connected in a social network. The first phase requires $\frac{N(N-1)}{2}$ steps, where N is the number of agents. This is quite demanding for the system, even when parallelised. However, to reach x and c it is not necessary to evaluate all pairs of agents: considering only a part of the agents can be sufficient. So Arentze et al. introduces a step size, such that for every agent i, i only considers agent j such that:

$$j = i + 1, i + 1 + y, i + 1 + 2y...$$

This will limit the number of evaluations in the first phase to $\frac{N(N-1)}{2y}$. However, this also implies that it is now impossible to consider any other agent than the agents selected by the steps size. This could be too limiting, because a large portion of the population can no longer become connected. This why the step size will not be a set value, but a random integer r such that $1 \leq r \leq 2 * stepsize$ for every step. This way all agent combinations are still possible, yet the expected speed up holds, because the expected value for the step size is still the previously set step size, therefor resulting in an expected number of evaluations of $\frac{N(N-1)}{2y}$.

The second phase is less demanding, because only a limited set of the agents have to be evaluated. Unless the average number of contacts is very high, the second step should behave linear in complexity, thus making the first phase the bottleneck.

4.2 Parameter estimation

When the social network generation is started, the threshold value and theta are not known, because they cannot be analytically derived. They can be estimated however. Let x^* be the expected average number of contacts of the agents, c^* the expected amount of clustering. We can run the network generation and compute the actual x and c, and update the threshold u^0 and θ accordingly. This way it is possible to iteratively estimate the right values for u^0 and θ. This can be done with the following procedure:

1. Initialise the thresholds as $u^0_1 = u^0_0$ and $\theta_1 = \theta_0$ and $t = 1$;
2. Run the social network generation using u^0_t and θ_t;

3. Compute x_t and c_t;

4. If $x_t < x^* - d_1$ then set $u^0_{t+1} = u^0_t - s_1$
   else if $x_t > x^* + d_1$ then set $u^0_{t+1} = u^0_t + s_1$;

5. If $c_t < c^* - d_2$ then set $\theta_{t+1} = \theta_t - s_2$
   else if $c_t > c^* + d_2$ then set $\theta_{t+1} = \theta_t + s_2$;

6. If $u^0_{t+1} \neq u^0_t$ or $\theta_{t+1} \neq \theta_t$ then set $t := t + 1$ and repeat from 2.

Where d_1 and d_2 are, respectively, how far off the x and c can be from the expected values. s_1 and s_2 are the adjustment parameters, i.e. how much the threshold and θ are adjusted. In the actual implementation s_1 and s_2 are starting off high and every iteration are divided by 1.5 until a minimum value is reached. This is to speed up the search for the right value, by narrowing down the possible values. So during the parameter estimation, the threshold and θ will first "jump" around. Because of the influence of the threshold and θ on each other, minimum values for s_1 and s_2 are used, to make sure the values can "walk" to the right values, if the right values were not yet found. This process is similar to simulated annealing.

4.2.1 Evaluation Functions

[19]

4.2.2 Sampling

Because doing the parameter estimation on the whole population is too demanding, it is possible to use only part of the population and use only these agents to estimate the parameters with. So part of the population is sampled randomly and a network only for that part is constructed and evaluated. When the parameters have approximated the right values, the threshold value and θ is used or the final social network for the entire population.

5 Need-Based Activity-Generation

All agents need an agenda for the whole day. The agenda planner used in IDSA could generate agendas for one day, on a statistical basis. Because culture needs to be modelled, a different approach is needed to model culture. The statistical approach does not model much of a cognition of the agents and its outcomes are too opaque. To model culture realistically, it is necessary that agents have a decent level of cognition to model decision making of the agent.

Research done about activity generation is mostly done in the field of travel prediction and transportation. Although the simulation purpose discussed in this thesis, i.e. simulation for training, is different from transportation, the approaches used for these simulations are useful for training purposes as well. It gives a framework for pattern of life behaviour of the agents on a macro-scale. This means that the agents will go through their daily-life, where agents will create patterns, even though non of the agents have any knowledge about them.

The previous agenda planner was based on the work of Kitamura [12] and was based on statistics. The model discussed in this section is developed by Arentze et al [3] [1] [4] , together with the work of Nijland[15]. Another method for activity generation is developed by Bhat et al. Arentze's model is specifically useful for culture modelling, because of its use of needs and utilities at its core. Using these needs and utilities agent decide which activities to conduct and which are better to postpone to another day. How agents perceive these needs and how these utility functions are defined can be used to model cultural aware agents. These utilities can be seen as a model for the thought process of a person who has to decide what activities to plan the next day.

In the first part the activity generation process is discussed. This includes the needs, the growth functions and the utility functions. Second, the activities themselves are cover: how they are defined and some special cases of activities. Finally, the sequencing process, which will schedule the activities into the agent's agenda.

5.1 Activity Generation

The big assumption in need-based activity generation is that people conduct activities to fulfill certain needs. An agent has a set of needs and a set of activities it can choose from and every activity has an influence on the needs of the agent. An activity generates utility based on the influence it has on the agent's needs. This utility is used by the agent to decide which activity conduct. The utility for an activity a is defined as:

$$U_a^t = U_a^0 + \sum_i \Delta B_{ai}^t$$

Where U_a^0 is the need independent utility for activity a, ΔB_{ai} is the difference a makes to need i and t is the start time of the activity. U_a^0 can incorporate preferences to conduct an activity on certain days. ΔB_{ai} can be defined as:

$$\Delta B_{ai} = \begin{cases} b_{ai} B_i^t & if\ b_{ai} \leq 0 \\ b_{ai}(B_i^{max} - B_i^t) & if\ b_{ai} > 0 \end{cases}$$

Where b_{ai} is the potential of activity a, B_{ai}^t the need i on time t and B_{ih}^{max} is the maximum value need i can take. Because the needs can work on any arbitrary scale, B_{ih}^{max} can be set to any arbitrary value. In this thesis a $B_{ih}^{max} = 100$ is assumed for every need. The potential $-1 \leq b \leq 1$ of an activity on a need i is the influence it has. So if $b > 0$ it will increase the need and $b < 0$ will decrease the need, which is a property of the activity. The potential b is defined as

$$b_{ai} = (\frac{b_a^0}{1+exp(\beta_a[\alpha_a - D_a])})$$

Where D_a is the duration of a and b_a^0, α_a and β_a are activity specific constant. α_a gives an indication about the normal duration of activity a. The function for b_{ai} describes a s-shaped function, which inflection point is defined by α_a and β_a indicates the marginal potential at the inflection point. b_a^0 is the base potential of the activity and are pre-defined.

5.1.1 Needs

All agents have a set of needs, which grow automatically over time and can be affected by the influence of conducted activities. The growth function G for a need i is defined as:

$$G_i(B_i^t, D) = \frac{B_i^{max}}{1+(\frac{B_i^{max}}{B_i^t}-1)exp(-\gamma_i D)}$$

Where D is the duration since the last update and γ_i is the growth rate of need i. When an activity is conducted, it will update the needs according to the ΔB_{ai} defined for it. If for a need i there is no change made by the activity, then i will increase according to the growth function. The update function after the activity is conducted then:

$$B_i^{t+D_a} = \begin{cases} B_i^t + \Delta B_{ai} & if\ \Delta B_{ai} \neq 0 \\ G_i(B_i^t, D_a) & if\ \Delta B_{ai} = 0 \end{cases}$$

Nijland found that the needs the motivate people to conduct activities can be reduced to six core needs:

- Rest;
- Social contact;
- Physical exercise;
- Being outdoors;
- Entertainment;
- New experiences.

These six needs are added to our agents. These six needs only apply to individual persons. Agents will also share needs among their household members. In addition to the personal needs, three household activities will be added to this:

- Daily goods;
- Housekeeping;
- Non-daily goods;

Household needs are defined per agent, and not household-wide, because possible difference in perception among the household members. When an activity is conducted by one of the agents, all agent in the household can benefit from this. Agents can also decide that another agent should perform the household activity. This means that agents within a household should find consensus on who will do the household activities.

5.1.2 Utilities

Before activities can be planned, a few more notions need to made. First is the utility of time (UoT), which is the utility per time unit of an activity. This is the utility of the activity divided by the duration of the activity plus any travel time:

$$UoT_a = \frac{U_a^t}{D_a + D_a^{trav}}$$

$$U_a^{patt} =$$

Give the formulas given above, we can now define when to conduct an activity. A threshold on the UoT, UoT^*, is defined for every agent-day combination. The threshold represents the scarcity of time on a day for an agent. The higher the time pressure, the higher the threshold. When $UoT_a > UoT^*$, the activity can be selected for that day. From the activities that reach the threshold, the one with the highest pattern utility is selected. Then, the utility, utility of time and pattern utility is computed again for the other activities and this is done until no activity is left that reached the threshold.

5.2 Activities

5.2.1 Order of Planning

The planning of activities is done in several steps.

First all mandatory activities are added to the agenda of the agent. This includes work and school. The duration and starting time of these activities are pre-defined in a contract. That contact is created when work is allocated and dictates which days and how long the agent is working and what time they have to start.

Second all household activities are planned, either to the agent themselves or to another household member. Because inconsistencies may arise, an exchange phase is done to reach consensus.

Lastly all personal activities are planned. Every time an activity is added to the agenda, the needs of the agent are updated.

5.2.2 Household Activities

5.2.3 Social Activities

5.2.4 Dynamic Activities

5.3 Threshold Estimation

The threshold for all agent-say pairs are not known a priori, because of the influence the threshold have on each other between days. Like in the social network generator, the thresholds can be estimated by running the algorithm and adjust the thresholds iteratively.

5.4 Sequencing

When all activities have been selected, they have not yet been ordered and inserted into the agenda of the agent. This is done be the sequencing phase.

6 Modelling Cultural Agents

The modelling will be limited to two of Hostede's dimensions: individualism versus collectivism (IND) and uncertainty avoidance (UAI). By applying changes to the social network and the activity generation, both can reflect the cultural values of the individual agents. Between two cultures the individual choices might not seem very different, but the overall pattern will show that one culture behaves differently to another on a macro-scale.

Note that Hofstede's dimensions only apply to the values of people, not the norms in their society. The values are what underlies the decision making on the cultural level. The norms that a culture follows may be explained by the cultural values, but cannot be prescribed by them. This will place limits on the modelling capabilities of culture in the multi-agent system discussed in this thesis.

6.1 Social Network

Because culture is shown in great part in how people interact with each other and how they are connected is the social network. Therefor culture has an influence on the social network generation.

6.1.1 Individualism versus Collectivism

Individualist cultures are known to more loosely connected and collectivist cultures form more tight groups. These tight groups are the clusters in the social network. To model this difference between these two ends of this cultural dimension, a multiplier on the θ is used, to influence clustering.

Adjusting the θ for the agents will give different clustering among the agents, but the clustering that has been set in the social network generation process, c^*, will not change. This means that no change will emerge between cultures. Therefor, c^* needs to be adjusted with the culture as well. For this, a collectivist bonus (or individualist penalty) can be used to also adjust the c^*. this adjustment should be based on the national cultural values, because the change in c^* applies to the entire population as a whole.

6.1.2 Uncertainty Avoidance

When UAI is high, people tend to avoid risk and ambiguous situations. This also applies to their social life. People with similar ideas are more likely to know each other than people who are very different. This is because people with different ideas are kept at a distance. This means that the higher the UAI, the more people want to avoid other who are different to them.

In a culture with high UAI, homophily is high: homophily is the similarity between agents. So when the utility between agents is computed, the effect of homophily compared to the distance utility will become greater. When UAI decreases, this utility bonus can be relaxed.

6.2 Activity Generation

Activity generation is the main source of behaviour in the simulation: although the behaviour models could be implemented in such a way that more detailed behaviour is performed by the agents, but this is beyond the scope of this thesis. Because most of the modelled cognition of the agents resides in the activity generation process and its utility functions, this is the best place to model culture.

6.2.1 Individualism versus Collectivism

In collectivist cultures people are more grouped together. In most cases, this coherence is especially strong within families. Families are represented in the simulation as the households. Because in collectivist cultures the family, here household, is more important, implies that a greater emphasis should lay on the household needs and thus household activities. This balance in importance between personal and household needs can be adjusted during the household activity planning phase.

6.2.2 Uncertainty Avoidance

A change of plans becomes less likely when an agent has a higher UAI. So when the agent meets another agent and the dynamic activity protocol is initiated, the probability of a change of plan for the agent depends on the UAI of the agent. If UAI it is high, it becomes less likely the agent will change their already planned activities.

7 Experiments and Results

To determine the effect of the different cultures on the behaviour on the population as a whole, the simulation is ran twice with different national cultural values: once with Dutch and once with Portuguese national cultural values. These two cultures differ mostly on IND and UAI.

During the simulation, the overall need of the agents and the number of planned activities for each sort of activity will be plotted. In order to show that the current system works, the results have to be in line with what Hofstede's model of culture describes to be different between these two cultures. The expected results are described in the previous section. Throughout the simulation, every 10 simulated minutes the system records for every need the summed value of all the agents. At the end of every day the total amount of activities planned that day is also written away.

During the experiment, all other variables will remain the same. Although there might be differences in tokens of activities that people can plan between the two cultures, these are left out of the experiment, to show how the model works. In a realistic setting, different activities are considered in different cultures.

8 Discussion

9 Conclusion

10 References

[1] T. Arentze, D. Ettema, and H. Timmermans. Estimating a model of dynamic activity generation based on one-day observations: method and results. *Transportation Research*, Part B: Methodological(45(2)):447–460, 2011.

[2] T. Arentze, M. Kowald, and K. Axhausen. An agent-based random-utility-maximization model to generate social networks with transitivity in geographic space. *Social Networks*, 35(3):451 – 459, January 2013.

[3] T. Arentze and H. Timmermans. A need-based model of multi-day, multi-person activity generation. *Transportation Research*, Part B: Methodological(43(2)):251–265, 2009.

[4] T. Arentze and H. Timmermans. A dynamic model of time-budget and activity generation: development and empirical derivation. *Transportation Research*, Part C: Emerging Technologies(19(2)):242–253, 2011.

[5] R. L. Axtell. Effects of interaction topology and activation regime in several multi-agent systems, July 2000. Working paper no. 12.

[6] S. de Jong, A. Klein, R. Smelik, and F. van Wermeskerken. Integrating run-time incidents in a large-scale simulated urban environment. 2015.

[7] F. Dignum, G. J. Hofstede, and R. Prada. From authistic to social agents. *AAMAS 2014*, January 2014.

[8] B. Edmonds. How are physical and social spaces related? cognitive agents as the necessary glue. Technical report, Centre for Policy Modelling, 2003.

[9] L. Hamill and N. Gilbert. Social circles: A simple structure for agent-based social network models. *Journal of Artificial Societies and Social Simulation*, 12(2):3, 2009.

[10] G. Hofstede, G. jan Hostede, and M. Minkov. *Cultures and Organizations: Software of the Mind*. McGraw Hill Education, 3rd edition edition, 2010.

[11] G. J. Hofstede, C. M. Jonker, and T. Verwaart. Cultural differentiation of negotiating agents. *Springer*, February 2010.

[12] R. Kitamura, C. Chen, R. M. Pendyala, and R. Narayanan. Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation*, 27:25–51, 2000.

[13] A. Klein. Internship report tno population modelling: Project intent driven scenario authoring, November 2015. Utrecht University and TNO Modelling, Simulation and Gaming.

[14] S. Mascarenhas, N. Degens, A. Paiva, R. Prada, G. J. Hofstede, A. Beulens, and R. Aylett. Modeling culture in intelligent virtual agents: From theory to implementation. *AAMAS 2015*, October 2015.

[15] L. Nijland. *A need-based approach to dynamic activity generation*. PhD thesis, Eindhoven University of Technology, January 2011.

[16] A. S. Rao and M. P. Georgeff. Bdi agents: From theory to practice. *Technical Note 56*, April 1995.

[17] N. Ronald, V. Dignum, C. Jonkers, T. Arentze, and H. Timmermans. On the engineering of agent-based simulations of social activities with social networks. *Information and Software Technologies*, 53:625 – 638, January 2012.

[18] S. Russell and P. Norvig. Artificial intelligence: A modern approach. *Third Edition*.

[19] T. Schank and D. Wagner. Finding, counting and listing all triangles in large graphs, an experimental study. University of Kalrsruhe, Germany.

[20] J. Schram. Culturally affected behaviour in belief-desire-intention agents. diploma thesis, Tilburg University, 2011.

[21] T. A. Snijders. The statistical evaluation of social network dynamics. *Sociological Methodology*, January 2001.

[22] C. Solomon and M. S. Schell. *Managing Across Cultures: The Seven Keys to Doing Business with a Global Mindset*. McGraw Hill Education, May 2009.

[23] R. Toivonen, J.-P. Onnela, J. Saramki, J. Hyvnen, and K. Kaski. A model for social networks. *Physica*, (371):851 – 860, 2006. Helsinki University of Technology.

[24] P. van Nispen. Culture encyclopedia.