Predicting Human Behavior in Crowds: Cognitive Modeling versus Neural Networks

Mark Hoogendoorn

VU University Amsterdam, Department of Computer Science
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
m.hoogendoorn@vu.nl

Abstract. Being able to make predictions on the behavior of crowds allows for the exploration of the effectiveness of certain measures to control crowds. Taking effective measures might be crucial to avoid severe consequences in case the crowd goes out of control. Recently, a number of simulation models have been developed for crowd behavior and the descriptive capabilities of these models have been shown. In this paper the aim is to judge the predictive capabilities of these complex models based upon real data. Hereby, techniques from the domain of computational intelligence are used to find appropriate parameter settings for the model. Furthermore, a comparison is made with an alternative approach, namely to utilize neural networks for the same purpose.

1 Introduction

As numerous incidents such as the disaster during the Love Parade in Duisberg, Germany in 2010 and the incident on the Dam Square in the Netherlands in that same year have shown: crowds can easily go out of control. Mostly only a small trigger is needed to cause a panic that can end in catastrophe. Therefore, paying special attention to the control of these crows is of utmost importance. This control could involve the strategic positioning of fences, of people, etcetera. However, to decide upon these strategic positions is not a trivial matter as it is not easy to make predictions on the behavior of such crowds.

Recently, within multi-agent systems, models have been developed that are able to simulate crowd behavior, making it possible to explore how well particular measures to control a crowd work. These models are sometimes based on more physics oriented approaches such as the well-known model of Helbing [7], but also cognitive oriented models have been developed, for example the ASCRIBE model [8]. The latter for instance takes the emotions of the individual agents in the crowd into consideration when determining their exact movement. Although validation of these models is essential, only few have been rigorously validated using real data. In [1; 9], actual human movement data obtained from video footage of a panicked crowd has been used to show that the model is able to reproduce such movements. These experiments have however been more targeted towards showing that the models are able to *describe* the data well, and not on their *predictive* capabilities, which is of course crucial to obtain a model that can be used for the purposes listed above. As a consequence, a lot of parameters of the individuals are highly tailored (in fact, each individual has a

M. Ali et al. (Eds.): IEA/AIE 2013, LNAI 7906, pp. 73–82, 2013.

set of unique parameters), making it hard to utilize for predictions of behavior of individuals with different characteristics. Secondly, the cognitive models specified are quite complex and the question is whether techniques from computational intelligence such as neural networks are not able to do the job just as well, or perhaps even better.

In this paper, the main purpose is: (1) to see whether a generic set of parameters can be found for an existing cognitive model to make robust predictions on crowd movement, and (2) to see how well a neural network would describe and predict crowd behavior compared to that tailored cognitive model. In order to do so, learning techniques from the domain of evolutionary algorithms have been utilized to tune the parameters of the existing cognitive model based upon training data, and a dedicated neural network is designed that is trained using a standard back propagation algorithm. Tests are performed using an existing dataset (cf. [1]).

This paper is organized as follows. The existing cognitive model used is described in Section 2, followed by the neural network design in Section 3. Section 4 presents the learning algorithm for the cognitive model, and Section 5 discusses the experimental setup and results. Finally, Section 6 is a discussion.

2 Existing Model

As mentioned in the introduction, one of the purposes of this paper is to judge the predictive capabilities of an existing model, namely the ASCRIBE model. This section discusses the model on a high level. Note that the precise mathematical details of the model have been omitted for the sake of brevity, see [8] for more details. In Figure 1, the main concepts underlying the ascribe model are shown.

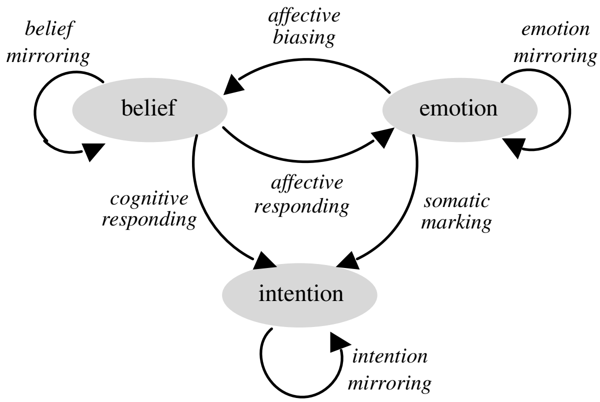


Fig. 1. Concepts and their relationship in ASCRIBE (cf. [8])

The states within the agent comprise of beliefs, emotions, and intentions. In the mathematical model, each state has a value between 0 and 1. Internally, the beliefs influence the emotion (i.e. an emotional response to a belief) and the intentions of the agent (i.e. responding to a certain belief by intending a certain action). The emotions in turn influence the way in which beliefs are perceived by means of an affective bias, and influence the value of intentions by associating a feeling with a certain option (referred to as Somatic Marking, following Damasio (cf. [4])). Besides these internal influences, the agent is also influenced by external factors, namely the agents in their neighborhood, resulting in (to a certain degree) mirroring the beliefs, emotions, and intentions of the others.

The model is highly generic in the sense that it describes social contagion in general, but an instance of the model has been specified that involves movements in crowds and panic situations (cf. [1]). Hereby, specific beliefs, intentions, and emotions have been inserted. More precisely, these states now concern beliefs, intentions and emotions associated with the options that the agents has. The options are the wind directions the agent can move in (N, NE, etcetera). In addition, a belief is present on the seriousness of the current situation, and an extra emotion, namely a general emotion of fear. The agent moves into the wind direction with the highest intention value (in case this direction is not blocked, otherwise the second highest intention is taken) and the speed of movement depends on the strength of the intention, ranging from a certain minimum speed to a set maximum speed.

This given instantiation of the model has a number of parameters that can be set on an individual basis. In previous research, the following parameters were found to work better in case they were to be set individually: (1) the distance within which agents influence each other (i.e. when mirroring takes place); (2) the way in which the initial belief about the positivity of the situation is obtained (by means of a parameter expressing from how far an agent is able to obtain information based on sight); (3) the minimum travel speed of an agent, and (4) the maximum travel speed. Therefore, only these parameters will be considered in the remainder of this paper, and the other values will remain at the value they have been given during previous work.

3 Neural Network Design

Next to the approach to express a complete and highly complex cognitive model, an alternative approach would be to utilize a neural network (representing the class of pure learning-based approach) for the learning of the behavior of agents within crowds. In this section, a dedicated neural network design is introduced to control agents in a crowd. The architecture is shown in Figure 2. Essentially, the input of the neural network comprises of different groups of nodes: (1) information about the agent's own prior movements; (2) information about possible obstacles in the direction the agent is currently heading at, (3) information about the movements and locations of other agents, and (4) information about the current time point. The information about the movement of other agents contains information on the i closest agents as well as the source of a certain panic. For the former, input is received on the distance of the other agent compared to the agent itself, and the direction of movement of the other agent (x and y movement). For the source of the panic, the same input is provided except now the direction of movement does not concern the movement of the panic source, but the direction compared to the current position of the agent itself. All the values provided to these nodes are scaled to the values [-1,1] for directions and [0,1] for obstacles and time. As output, the neural network provides the x-and y-movement. The hidden neurons have a hyperbolic tangent sigmoid function, whereas the output nodes contain a linear function. The number of hidden neurons and layers is left variable; different values will be tried when running experiments.

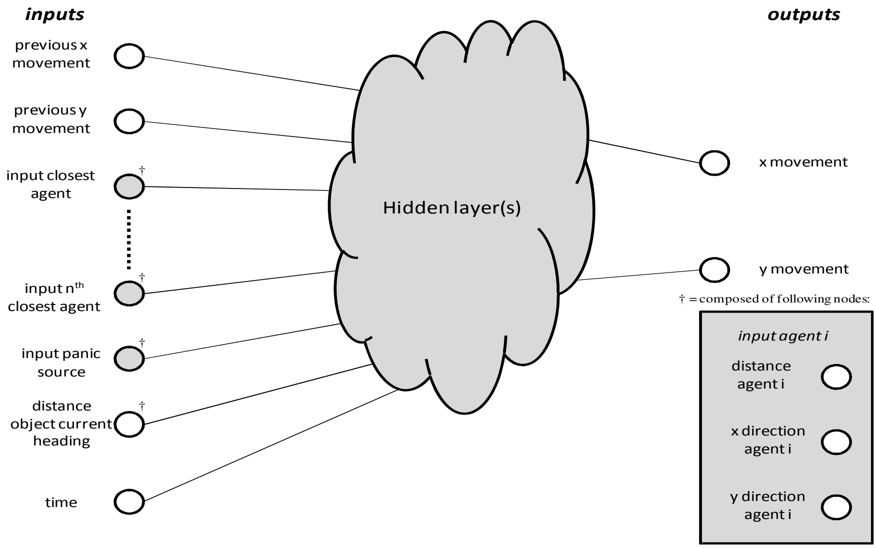


Fig. 2. Neural network architecture

4 Learning Algorithms

It is assumed that a dataset is available which expresses the coordinates of agents, sources of panic, and obstacles at each time point of a crowd panic. Using this information, the parameters of the two models can be learned. For the neural network this learning is relatively straightforward: the dataset is translated into the corresponding input and desired output of the neural network and a time point in combination with a single agent is considered a training sample for the Levenberg-Marquardt back propagation algorithm [3; 11]. For the learning part of the ASCRIBE model things are a bit more complicated. In order to tune the parameters of ASCRIBE, an evolutionary approach is used. Hereby, parameters are not tuned on an individual agent level, but parameters for the population of agents as a whole are determined. This facilitates prediction of movement of the agents. A "standard" genetic algorithm is used:

- *Individual representation.* The population is composed of individuals that are represented by a binary string which represents real values (i.e. the parameters of the ASCRIBE model) with a certain precision.
- *Population initialization.* The population is initialized randomly.
- *Selection.* The selection of individuals is performed by first ranking the individuals using linear ranking and then selecting individuals based upon stochastic universal sampling.
- *Mutation.* The mutation operator used is straightforward: each bit is simply mutated with a certain probability.
- *Crossover.* A single point crossover function is used to combine the individuals.

Crucial for the evolutionary algorithm is the specification of the fitness function. In order to determine the fitness, an entire simulation is performed using the model with the set of parameters represented by the individual subject to evaluation. Hereby, the agents are placed at their initial positions and follow the path dictated by the model with the selected parameter settings. As the main goal is to reproduce the movements of the people involved in the scenario, it was decided to take the average (Euclidean) distance (over all agents and time points) between the actual and simulated location:

$$\varepsilon = \sum_{agents\ a} \sum_{timepoints\ t} \frac{\sqrt{(x(a,t,sim)-x(a,t,data))^2+(y(a,t,sim)-y(a,t,data))^2}}{\#agents \cdot \#timepoints}$$

Here, *x(a, t, sim)* is the x-coordinate of agent *a* at time point *t* in the simulation, and *x(a, t, data)* the same in the real data (similarly the y-coordinates). Both are in meters. This fitness measure complies with the measure used in [1].

5 Case Study: 4 May 2010

This section presents the results for a case study using a dataset from an incident on the 4th of May in Amsterdam, the Netherlands. First, the dataset is described, followed by the experimental setup and results.

5.1 Dataset Description

As a case study, data from an incident which took place on the Dam Square in Amsterdam, the Netherlands on May 4th 2010 has been used. The incident involves a person that starts shouting during a two minute period of silence during the national remembrance of the dead. As a result of the shouting person, people start to panic and run away[1]. A dataset has been composed that describes the coordinates of movement of 35 people during the incident (see [1]). Next to these coordinates, the coordinates of the source of the panic (i.e. the shouting man) as well as the objects on the square have been logged as well. This data can be used for training the various models and judging the outcome of the models. In order to train the neural network a slight translation of the data needs to take place, but this translation is quite trivial (translating coordinated over time into movement and direction, etcetera) and therefore not further described here.

5.2 Experimental Setup and Results

This section describes the experimental setup used, and presents and discusses the results obtained in detail.

Experimental Setup

During the experiments, comparisons will be made on the descriptive performance (i.e. the performance on the training set, which contains the full set of individuals) of the various approaches as well as the prescriptive performance. For judging the

[1] See `http://www.youtube.com/watch?v=1AEXcxwHJVw` for a movie clip

predictive performance a 5-fold cross validation approach is used, whereby each fold consists of 1/5th of all agents. The performance is averaged over the performance on the five configurations of the folds. Based upon this approach, the following models will be compared:

1. *ANN.* An artificial neural network as detailed in Section 3 and 4.
2. *ASCRIBE generic.* The ASCRIBE model with a single set of parameters for all agents, whereby the learning takes place using evolution (see Section 2 and 4).
3. *ASCRIBE individual.* The ASCRIBE model with individual parameter per agent. This approach is used as a benchmark, it can be seen as the optimal way of describing the individual agents with ASCRIBE. Note that using such individual parameters is not realistic for judgment of predictive capabilities, but it is used nonetheless.
4. *No movement.* Assumes that all agents remain at their position.

For the neural network based approach various configurations of the neural network have been used:

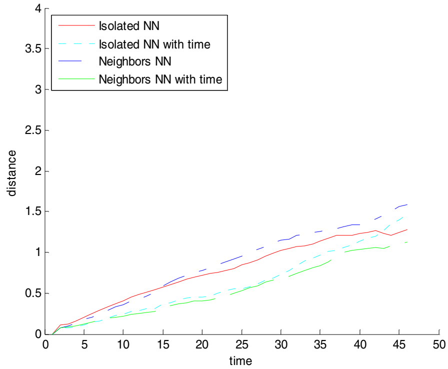
1.1 *Isolated NN.* Using no information of movement of surrounding agents;
1.2 *Isolated NN with time.* Using time and no additional information of movement of surrounding agent;
1.3 *Neighbors NN.* Using information of movement of the closest two surrounding agents;
1.4 *Neighbors NN with time.* Using information of movement of the closest two surrounding agents and time.

For each of the learning algorithms 10 runs have been performed for each setting (the choice for 10 runs was made due to the costly fitness function of the evolutionary algorithm, making it very time consuming to run for the predictive case as 50 runs are required). For the neural network a run consisted of 1000 epochs of the Levenberg-Marquardt back propagation algorithm with an initial μ set at 0.001, the μ increase factor is set to 10 and the decrease factor to 0.1. Furthermore, in experiments not reported here for the sake of brevity different settings for the hidden layers in the neural network have been used, thereby showing that a single hidden layers with three hidden neurons worked best for the first two networks, and two hidden layers with five and three neurons in the respective layers for the third and fourth neural network configurations. For the evolutionary approach to learn the parameters of the ASCRIBE model (setting 2) a population size of 50 was used, and 100 generations. A mutation rate of 0.01 has been used and a crossover rate of 0.5. These parameters are based on the "standard" settings as proposed by DeJong [5], but some modifications were made (the process converges earlier resulting in a need for fewer generations, and the crossover and mutation rate have been altered based on test runs).

Results

The results are detailed in this section. First, the descriptive capabilities are shown, followed by the predictive capabilities.

Performance Training Set. Figure 3a shows an overview of the average performance of the different neural network configurations on the training set. Here, the x-axis shows the time of the incident, whereas the y-axis shows the average difference between the simulated and actual position of the agents in the crowd (cf. Section 4, except the value is not averaged over all time points). It can be seen that the Neighbors NN performs worst, whereas the other network configurations perform very similar. The averages over time points and standard deviations are shown in Table 1. The differences are not statistically significant (using a t-test with 5% significance level).



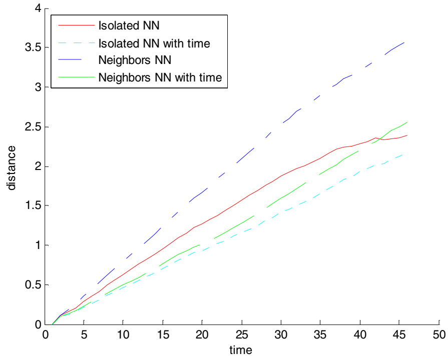


Fig. 3. (a) Comparison average training set performance different NN configurations; (b) Comparison best runs on training set of NN's

Figure 3b shows the runs during which the performance was best. Here, the performances are again very close to each other. Of course, this comparison is interesting, but the most interesting is to see how well the neural networks perform compared to the benchmark, and also how well the ASCRIBE model with a single set of parameters for all agents is able to describe the behavior. Figure 4a compares the average performances of the different approaches (the precise results are again shown in Table 1). It can be seen that the ASCRIBE model with individual parameter performs best, which is obvious, it is highly tailored to the case at hand. The ASCRIBE model with evolution performs somewhat worse, whereas the average performance of the best neural network configuration is even worse than predicting no movement (whereby the latter gives a good impression of the average movement of the crowd as the error found at the end is equal to the average final distance traveled per agent).

When comparing the best runs, the outcome is suddenly completely different (see Figure 4b). The neural network performs even better than the ASCRIBE model with individual parameters (note that in this case the neural network which considers the movement of neighboring agents and time is best). Hence, the neural network based approach has huge fluctuations in outcome, from a very bad performance, to excellent performance.

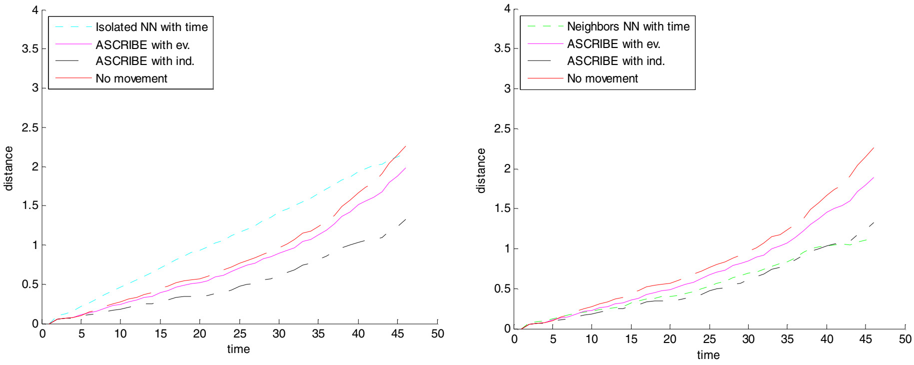


Fig. 4. (a) Comparison averages performances, and (b) comparison best performances on training set NNs, ASCRIBE, and no movement

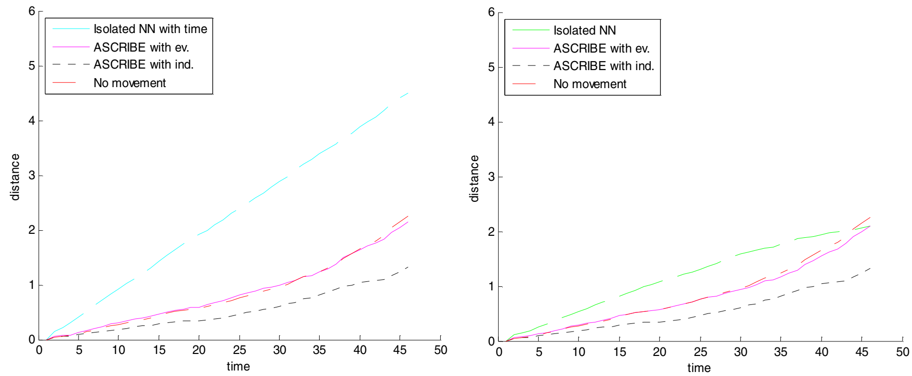


Fig. 5. (a) Comparison averages performance, and (b) comparison best runs performances on test set NNs, ASCRIBE, and no movement

Performance Test Set. As said, in this paper the predictive capabilities of the various approaches are also subject of investigation as this would facilitate the utilization of models for simulating hypothetical scenarios. Figure 5a shows the predictive capabilities of the average performance of the best neural network configuration with respect to the predictive capabilities (in this case the isolated neural network with time). It can be seen that the performance of ASCRIBE with generic parameters is the best of the more complex predictive models (although the performance is not too good, it is close to no movement), but for the neural network approach the performance is very bad. In this case, there is a statistically significant difference between the best neural network approach and the ASCRIBE model with generic parameters (p-value of 0.0001).

When comparing the best runs again, the predictive capabilities of some neural networks trained (in this case with neighbors and time considered as that provides the best predictive run) are a bit better (see Figure 5b), but still performing worst in comparison with the other approaches.

Table 1. Detailed results various algorithms

Approach	*Training set*			*Test set*		
	Avg. perf.	*Std. dev.*	*Best perf.*	*Avg. perf.*	*Std. dev.*	*Best perf.*
Isolated NN	1.4002	0.9678	0.7777	2.4158	0.5574	1.1893
Isolated NN with time	1.1051	1.0956	0.619	2.2519	0.5422	1.6082
Neighbors NN	1.9134	1.7985	0.8531	2.6535	0.5124	1.9511
Neighbors NN with time	1.2477	1.2195	0.551	2.6269	0.8239	1.3894
ASCRIBE generic	0.7584	0.0507	0.7228	0.8482	0.0234	0.8082
ASCRIBE individual	0.5231	0	0.5231	0.5231	0	0.5231
No movement	0.8394	0	0.8394	0.8394	0	0.8394

6 Discussion

In this paper, two approaches have been used to both describe as well as predict behavior of humans within crowds: a neural network based approach and an approach based on theories from the domain of social psychology called ASCRIBE [8]. To make it possible for ASCRIBE to predict behavior of individuals an evolutionary approach has been developed that tunes the parameters of the model based on a training set, and uses those parameters settings for making predictions. Such parameters are generic, and not specifically tailored towards each individual agent. In order to evaluate the approaches on their suitability they have been tested against a real-world dataset (cf. [1]). The results showed that both for the descriptive as well as the prescriptive behavior the performance of the ASCRIBE model with learning of generic parameters showed reasonable performance, although for the case of prediction it does not move a lot beyond predicting no movements. The neural network based approach performed significantly worse especially for the predictive case, and showed a large variation in performance. One can thus conclude that it is quite difficult to find a generic set of parameters that is able to predict human movement in crowds well due to the large individual differences.

Of course, more work has been done in this area. As said, Helbing [7] has proposed a model based upon particle physics that is able to describe panic in crowds, the model is able to generate realistic crowd behavior, but has not been evaluated for its predictive capabilities. In [1] a comparison has been made regarding the descriptive capabilities of the models, showing that ASCRIBE is able to describe this behavior best. In [9] a comparison between three models is made (ASCRIBE, an epidemiological-based Durupinar model [6], and the ESCAPE model) by evaluation of the descriptive capabilities, showing that ASCRIBE perform superior. Again, no evaluation based on the prescriptive capabilities is made, which is the main focus of this paper, nor have the parameters been rigorously tunes as done here.

References

1. Bosse, T., Hoogendoorn, M., Klein, M.C.A., Treur, J., van der Wal, C.N.: Agent-Based Analysis of Patterns in Crowd Behaviour Involving Contagion of Mental States. In: Mehrotra, K.G., Mohan, C.K., Oh, J.C., Varshney, P.K., Ali, M. (eds.) IEA/AIE 2011, Part II. LNCS (LNAI), vol. 6704, pp. 566–577. Springer, Heidelberg (2011)
2. Braun, A., Musse, S.R., de Oliveira, L.P.L., Bodmann, B.E.J.: Modeling Individual Behaviors in Crowd Simulation. In: The 16th International Conference on Computer Animation and Social Agents, CASA 2003, pp. 143–147. IEEE Press, New Jersey (2003)
3. Marquardt, D.: An algorithm for least squares estimation of non-linear parameters. J. Ind. Appl. Math., 431–441 (1963)
4. Damasio, A.: The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex. Phil. Transactions of the Royal Society: Biological Sciences 351, 1413–1420 (1996)
5. DeJong, K.A., Spears, W.M.: An Analysis of the Interacting Roles of Population Size and Crossover in Genetic Algorithms. In: Schwefel, H.-P., Männer, R. (eds.) PPSN 1990. LNCS, vol. 496, pp. 38–47. Springer, Heidelberg (1991)
6. Durupinar, F.: From Audiences to Mobs: Crowd Simulation with Psychological Factors. PhD dissertation, Bilkent University, Dept. Comp. Eng. (July 2010)
7. Helbing, D., Farkas, I., Vicsek, T.: Simulating Dynamical Features of Escape Panic. Nature 407(6803), 487–490 (2000)
8. Hoogendoorn, M., Treur, J., van der Wal, C.N., van Wissen, A.: Modelling the Interplay of Emotions, Beliefs and Intentions within Collective Decision Making Based on Insights from Social Neuroscience. In: Wong, K.W., Mendis, B.S.U., Bouzerdoum, A. (eds.) ICONIP 2010, Part I. LNCS (LNAI), vol. 6443, pp. 196–206. Springer, Heidelberg (2010)
9. Tsai, J., Bowring, E., Marsella, S., Tambe, M.: Empirical evaluation of computational emotional contagion models. In: Vilhjálmsson, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS (LNAI), vol. 6895, pp. 384–397. Springer, Heidelberg (2011)
10. Tsai, J., Fridman, N., Bowring, E., Brown, M., Epstein, S., Kaminka, G., Marsella, S., Ogden, A., Rika, I., Sheel, A., Taylor, M.E., Wang, X., Zilka, A., Tambe, M.: ESCAPES - Evacuation Simulation with Children, Authorities, Parents, Emotions, and Social comparison. In: Tumer, K., Yolum, P., Sonenberg, L., Stone, P. (eds.) Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2011 (2011), Innovative Applications Track
11. Levenberg, K.: A method for the solution of certain problems in least squares. Quart. Appl. Math. 2, 164–168 (1944)