A genetic fuzzy system to model pedestrian walking path in a built environment

Mojdeh Nasir * , Chee Peng Lim 1 , Saeid Nahavandi 2, Douglas Creighton 3

Centre for Intelligent Systems Research (CISR), Deakin University, 75 Piddington Road, Waurn Ponds, Vic 3216, Australia

A R T I C L E   I N F O
Article info
Article history:
Received 17 May 2013
Received in revised form 7 January 2014
Accepted 10 March 2014

Keywords:
Walking trajectory prediction
Pedestrian behaviour modelling
Genetic fuzzy system
Pedestrian environmental perception

A B S T R A C T
A study on the pedestrian’s steering behaviour through a built environment in normal circumstances is presented in this paper. The study focuses on the relationship between the environment and the pedestrian’s walking trajectory. Owing to the ambiguity and vagueness of the relationship between the pedestrians and the surrounding environment, a genetic fuzzy system is proposed for modelling and simulation of the pedestrian’s walking trajectory confronting the environmental stimuli. We apply the genetic algorithm to search for the optimum membership function parameters of the fuzzy model. The proposed system receives the pedestrian’s perceived stimuli from the environment as the inputs, and provides the angular change of direction in each step as the output. The environmental stimuli are quantified using the Helbing social force model. Attractive and repulsive forces within the environment represent various environmental stimuli that influence the pedestrian’s walking trajectory at each point of the space. To evaluate the effectiveness of the proposed model, three experiments are conducted. The first experimental results are validated against real walking trajectories of participants within a corridor. The second and third experimental results are validated against simulated walking trajectories collected from the AnyLogic ® software. Analysis and statistical measurement of the results indicate that the genetic fuzzy system with optimised membership functions produces more accurate and stable prediction of heterogeneous pedestrians’ walking trajectories than those from the original fuzzy model.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The field of pedestrian behaviour modelling and simulation has attracted researchers’ attention from multiple disciplines owing to its application in a vast spectrum of domains, including computer graphics, architecture, psychology, urban planning, and robotics [1,2]. The floor layout is one of the key factors that play an important role in a built environment. In terms of designing the indoor areas, operating efficiency of public facilities is a matter of concern that needs to be addressed. Prediction of pedestrians’ walking trajectories is an important requirement for efficient design of urban public areas like

* Corresponding author. Tel.: +61 3 5227 1150; fax: +61 3 5227 1046.
E-mail addresses: mojdeh.nasir@research.deakin.edu.au (M. Nasir), chee.lim@deakin.edu.au (C.P. Lim), saeid.nahavandi@deakin.edu.au (S. Nahavandi), douglas.creighton@deakin.edu.au (D. Creighton).

1 Tel.: +61 3 522 73307.
2 Tel.: +61 3 522 71231.
3 Tel.: +61 3 522 72179.

http://dx.doi.org/10.1016/j.simpat.2014.03.002
1569-190X/© 2014 Elsevier B.V. All rights reserved.
shopping malls and terminals. Nevertheless, the inter-relationship between a pedestrian’s walking trajectory and the pedestrian’s perception towards the environmental design is often neglected in this area of research. As such, the impetus underlying the current study is to develop a useful model that is able to capture the relationship between the environmental design and the pedestrian’s perception, and subsequently to predict the pedestrian’s walking trajectory under normal conditions in a built environment. Moreover, a dominant requirement of the model is its ability to provide natural-looking results for the predicted walking trajectories. To meet these requirements, we treat the pedestrian’s steering behaviours as a complex model that comprises human perception towards the surrounding environment and the possible reactions towards the environmental stimuli. In this regards, a pedestrian’s perception of the surroundings is inherently vague, and is subjective to an individual’s characteristics and intention. To deal with the uncertainty and vagueness issues in pedestrian perception–action modelling, we employ a fuzzy logic approach in this study. Fuzzy logic serves as an appropriate framework to incorporate imprecision and subjective features of the environmental perception into the perceptual–action model. In problems that involve human behaviour modelling, fuzzy logic also has certain advantages over other methods, such as the ability to imitate human thought processes [3].

However, the primary challenge in applying a fuzzy-based approach is the manual and time consuming procedure for rule generation and membership parameter tuning [4]. In this study, the fuzzy rule-based model needs to provide outputs with smooth transitions, rather than sudden changes between states, of the walking trajectory. Moreover, the diverse nature of heterogeneous pedestrians with varying speeds and step-lengths of their walking behaviours requires the model parameters to suit each pedestrian’s characteristics. As such, there is an obvious need to develop a generic and robust model that covers the walking behaviours of heterogeneous pedestrians.

The novelty of this study is the integration of the concepts of environmental stimuli, pedestrians’ subjective and vague perceptions towards the environmental stimuli, and socio-psychological steering forces, to model heterogeneous pedestrians’ walking trajectories. The main contributions are twofold: (i) a representation of diverse and imprecise characteristics of pedestrians’ perceptions towards the surrounding environment by using a fuzzy-based system; and (ii) inclusion of the environmental design and the corresponding socio-psychological stimuli into the proposed fuzzy-based system by employing the Helbing social force model. Specifically, a genetic fuzzy system (GFS) is proposed, whereby the genetic algorithm (GA) is employed to optimise the associated membership function parameters of the fuzzy model. To the best of the authors’ knowledge, this study also constitutes a new application of the GFS in the domain of modelling and prediction of heterogeneous pedestrians’ steering behaviours under normal conditions in built environments.

1.1. Background of pedestrian behaviour models

Over the years, researchers have proposed various approaches to model pedestrian behaviours in either individual or crowd situations. Pedestrian behaviours have been considered from different viewpoints. Amongst them, route (itinerary) choice, steering (navigation), wayfinding (path finding), and crossing intersections [5] are the most dominant behaviours [6]. Moreover, the focus of many investigations is on modelling and simulation of behaviours under panic situations such as crowd evacuation due to fire, natural disasters, or terrorism attacks [7–9]). Hamacher et al. [9] employed a dynamic network flow model in a macroscopic scale and a cellular automaton simulation model in a microscopic scale to study the time bound of evacuation. Manley and Kim [8] addressed emergency evacuation from built environments by employing internal forces to follow the shortest path. In a comprehensive study by Zheng et al. [7], seven methods were presented, which were further categorised with six specific features. The features highlighted a number of important aspects pertaining to modelling of pedestrian behaviours, which include heterogeneity, scale of modelling, condition, space, and time steps.

In this paper, we focus on the interaction of individual pedestrians with the surroundings under normal and non-panic situations. The problem is investigated using a microscopic approach. In this aspect, there are three main microscopic approaches, i.e., the social force model (SFM) [10], Cellular Automata (CA) [11], and agent-based model (ABM) [8,12]. The details are as follows.

The Helbing SFM describes a pedestrian’s with a mathematical model of attractive and repulsive effects from the surrounding environment with respect to the speed and movement direction [10,13,14]. Contradictory socio-psychological forces within the environment motivate the entity to move towards a desired destination, with an ideal speed. The SFM describes the pedestrian’s behaviour in a microscopic level with a continuous deterministic approach. However, capturing complex behavioural rules and behavioural heterogeneity is difficult. Moreover, it is a myopic technique that does not include the vision ability of the pedestrian.

In the second approach, CA expresses the pedestrian flow by a discrete arrangement of the floor into grids of equal cells. According to the CA principles, a pedestrian dynamic model provides a transition matrix that indicates the preferences to move from one cell to the neighbouring cells [11,15]. This method expresses the pedestrian flow in a discrete stochastic framework. Therefore, time and space are discrete elements. This is a drawback for steering behaviour modelling. Another downside of this approach is the weakness in exact calculation of the travel time and distance. Moreover, a cell-based model discretises the floor into cells, which is a limitation for some of the applications that need to consider the topological characteristics of the route.

The recent approach of ABM replicates the pedestrian’s behaviour using an agent with different levels of intelligence, or a set of if-then rules for steering behaviour modelling. The key feature of ABM is the capability of representing heterogeneity in a pedestrian’s behaviour. During the last two decades, ABM has become an increasingly important approach for modelling
the dynamics of complex systems, and many frameworks have been suggested [12,16–18]. As an example, Bonabeau [12]
acknowledged the strong points of this method that captured complex behaviours in different areas of application.
Based on the above discussion, we propose a hybrid model that employs ABM and SFM to predict the walking trajectory of
pedestrians within a built environment. The proposed model takes advantage of the salient features of both ABM and SFM to
develop a microscopic continuous stochastic approach for modelling the steering behaviours of with heterogeneous pedes-
trians. Our goal is to devise an adaptive model that is able to model and predict a realistic walking trajectory influenced by
the floor layout in the built environment by employing engineering techniques. The following section describes the frame-
work that deals with the agent’s (i.e. pedestrian) behaviours.

1.2. Fuzzy logic and application of genetic fuzzy system

Owing to imprecise and uncertain aspects of modelling pedestrians’ perceptions towards the surrounding environment, a
fuzzy logic-based approach has been introduced in our preliminary study that to incorporate the environmental influences in
pedestrian steering behaviours [19]. To fully exploit the advantages of the fuzzy approach and to avoid its drawbacks (i.e.,
manual parameter tuning as explained earlier), we extend our previous work by integrating the GA into the fuzzy model to
form a GFS in this study. The application of the GA to fuzzy systems is well-established [20], whereby the GA provides a flex-
ible and promising approach for optimising fuzzy set parameters and fuzzy rules. In this study, the proposed GFS aims to
predict the pedestrian’s steering behaviours with minimum discrepancy as compared with the real or simulated trajectory
data, as demonstrated in the experimental section of this paper. To the best of our knowledge, this is a new application of the
GA to the domain of pedestrians’ steering behaviour modelling and simulation.

In general, the GA can be used to evolve a fuzzy system in two aspects: genetic tuning and genetic learning. The objective
of genetic tuning is to optimise the parameters of the membership functions (MFs) for a fuzzy rule-based system with a set of
predefined rules. On the other hand, genetic learning aims to generate a set of rules from scratch. So, the genetic learning
problems require an extensive search space in order to generate a set of appropriate and useful rules. Designing a useful
GFS entails a trade-off between the dimension of the search space and the efficiency of the search process. A faster method
often searches through a smaller search space. However, examining search and optimisation problems within a small search
space can lead to suboptimal solutions [20]. As a result, we choose to employ the first option, which is genetic tuning to gen-
erate a GFS with an initial rule set pre-defined by a model developer without domain expertise (i.e. a rough rule base).

We endeavour to develop a useful GFS to predict the pedestrian’s walking trajectory confronting with the environmental
stimuli in this study. The fuzzy system is used to model the walking trajectories of heterogeneous pedestrians with different
personal characteristics. The fuzzy MFs are tuned and optimised by the GA. To do so, we focus on the use of a predefined rule
set and a number of real and simulated trajectory data sets to optimise the MFs. The rules are established based on common
sense knowledge of the model developer. Appendix B shows some of the rules used in the fuzzy system. As an example, Rule
10 (Appendix B) indicates that “If level of stimuli in right position (RP) is high attractive, and front position (FP) is medium
attractive, and left position (LP) is low repulsive, Then, turning angle is to the right (in degrees)”. The details of the GFS and
the modelling study are presented in the subsequent sections of this paper.

This paper is organised as follows. Section 2 provides an overview of related works. Section 3 discusses a basic fuzzy
model for walking trajectory prediction. The GFS architecture is introduced in Section 4. Section 5 presents the details of the
simulation study, which include data collection, experimental results, discussion, and validation. Conclusions and future
work are given in the last section.

2. Related works

There is a wide range of GFS applications in different areas, such as transportation [21,22], industry [23,24], control
[25,26], pattern classification [27–29], virtual environment and computer games [4,30]. In the area of modelling and simu-
lation of steering behaviours, the GFS-based approaches have been used for optimisation tasks in a few domains such as ani-
minated steering agents in a 3D virtual environment [4], bipedal animated character behaviours in the environment [31],
vehicle decision systems for safe driving [22], and terrain mobile robot path planning and obstacle avoidance [32].

A dominant area of the GFS applications is intelligent agents for computer animation and games. In Gerdelan and O’Sul-
vian [4], the GFS was used for modelling the steering behaviours of animated moving agents in a 3D virtual environment.
The focus of their study was on providing an automatic rule calibration process in real time. The aim was to provide an adap-
tive agent to tune the rules based on the dynamic environment. The authors compared different features of the GFS with
other approaches in agent-based modelling, such as Genetic Neural Network (GNN) [33], the rtNeat algorithm [30], and the
basic fuzzy model. While GNN and rtNeat are black box approaches, the basic fuzzy system and GFS are able to provide a
set of comprehensible rules [4].

Allen and Faloutsos [31] developed an evolutionary algorithm to evolve a neural network framework as the controller for
use with bipedal human-like characters. The simulation results showed a reliable and smooth walking motion of the human-
scale characters. The GFS has also been widely used for other mobile robot path planning tasks. As an example, Mohamma-
dian and Stonier [32] described an integration of the GA and a fuzzy logic controller to adapt a set of fuzzy rules to control the
robot steering behaviours. The inputs to the system comprised the Cartesian position and angle of movement, and the output
was the steering control signal. Their work imitated the human-like controller and generated a set of rules using a fuzzy amalgamation process.

The main focus of the aforementioned works is to provide an automatic mechanism for tuning parameters or calibrating rules in fuzzy-based systems. Similarly, we apply the GA to tune and optimise the proposed GFS for modelling the pedestrian steering behaviours in this study. The aim is to have an adaptive GFS that allows its parameters or rules to be automatically optimised based on data samples; hence saving time and effort in developing the fuzzy system. How the proposed fuzzy-based model can be used for trajectory prediction of pedestrians within a built environment is explained in the next section.

3. A fuzzy-based system for modelling and simulation of walking trajectory

The proposed fuzzy-based model shows how fuzzy logic is used to capture the environmental effects in steering behaviour modelling and simulation. The first step of modelling is discretisation of the floor to represent the environment. There are three methods of spatial representation of the environment, i.e., grid-based, network-based, and radial-based [34]. In this study, we form the terrain within the pedestrian’s field of view using the radial-based discretisation. The spatial representation of the environment is shown in Fig. 1. Three radial positions located within two metres from the current location with 45 degrees to the right or left define the possible positions of the next step. These three future positions are denoted as Left Position (LP), Front Position (FP), and Right Position (RP). The effects induced by the surroundings to these three positions are represented by the input fuzzy sets.

The future positions are collected in set $S$, i.e.,

$$ S = \{ \text{Front Position (FP)}, \text{Right Position (RP)}, \text{Left Position (LP)} \}. $$

Two main variables in the steering activity are the speed and direction of movement. So, each pedestrian is characterised with two state variables, i.e., $P_i(t)$ and $S_i(t)$, and a personal variable, i.e., $L_i$. As such,

$$ P_i(t) = (X_{position_i}(t), Y_{position_i}(t)) \in \mathbb{R}^2, $$

where $P_i(t)$ indicates the position of pedestrian $i$ at time $t$. Besides, $S_i(t) \in \mathbb{R}$ indicates the pedestrian’s speed at time $t$, while $L_i$ indicates the pedestrian’s step-length. The step-length used in this simulated environment is set in accordance with the settings recommended in Weidmann [35], i.e., the range of [0.6 m, 0.7 m]. We denote $L$ as the family of indicators that represents the variables engaged in the fuzzy model of pedestrian steering behaviours with inclusion of environmental effects, as follows.

$$ I = (P_i(t), S_i(t), L_i, P_{FP}, P_{RP}, P_{LP}, R, T_{angle}), $$

With this definition, three variables, i.e., $P_{FP}$, $P_{RP}$, $P_{LP}$ denote the pedestrian’s perception of the environmental effects in three possible future positions. $R$ is a set of inference rules, and $T_{angle}$ is the turning angle, which is the output of the fuzzy model. Fig. 2 depicts the architecture of the proposed fuzzy model, which is used to infer the degree of turning angle for the next step according to the environmental perceptions of three future positions as its inputs. A detailed description of the model has been presented in our previous work [19].

The fuzzy rule-based model is described as follows.

$$ Y = F(x_i), \quad i = 1, 2, 3, $$

Rule $R_j$: IF $x_1$ is $A_{1j}$ and $x_2$ is $A_{2j}$ and $x_3$ is $A_{3j}$, THEN $T_{angle}$ is $Y$, $j = 1, 2, \ldots, 216$,

where variable $x_i$, $i = 1, 2, 3$, are the three inputs, i.e., $P_{FP}$, $P_{RP}$, $P_{LP}$, and $Y$ indicates the output, which is the angular displacement of the next step (denoted by $T_{angle}$), and $A_{nj}$, $n = 1, 2, 3$, represent the antecedent fuzzy sets.

The inputs are scalar quantities computed using the SFM. They represent the environmental effects induced by the surroundings [13]. In the SFM, the socio-psychological interactions with the environment are categorised with two indicative forces, namely attractive and repulsive stimuli. The attractive or repulsive stimuli of any objects within the surrounding

![Fig. 1. The geometrical configuration of the terrain in front of the pedestrian.](image-url)
environment vary in each point of the space. The dynamic variations of these forces are considered in the modelling process. These scalar quantities are fuzzified using six fuzzy MFs, \( \mu_f \), i.e. high attractive, medium attractive, low attractive, low repulsive, medium repulsive, and high repulsive, as follows.

\[
\mu_f(x_i) = \begin{cases} 
\text{High Attractive (HA)}, & \text{Medium Attractive (MA)}, \\
\text{Low Attractive (LA)}, & \text{Low Repulsive (LR)}, \\
\text{Medium Repulsive (MR)}, & \text{High Repulsive (HR)}.
\end{cases}
\]

In addition to the fuzzy MFs, different speeds and step lengths of different pedestrians are used as the inputs to the model. Each fuzzy MF, \( \mu_f \), implies the degree that a crisp value \( x_i \) belongs to a fuzzy set, i.e.,

\[
\mu_f : x_i \rightarrow [-1, 1], \quad i = 1, 2, 3.
\]

The fuzzy MFs can assume different shapes, which include triangle, trapezoidal, and Gaussian. In this study, both the input and output linguistic variables are represented using the Gaussian MFs, which is in agreement with the exponential decay (i.e. Gaussian function-like influence) of the environmental effects on the surroundings [36]. The Gaussian MF parameters are defined with a 2-tuple \((\mu, \sigma)\) element comprising the mean and standard deviation of the Gaussian function, i.e.,

\[
\mu_f(x_i) = \exp \left( \frac{x_i - \mu}{\sigma} \right)^2,
\]

where \( \mu \) and \( \sigma \) are the mean and standard deviation that form the shape of the MF. These two parameters play an important role in the fuzzification and defuzzification procedure. Thus, the aim of this study is to find the optimum values of these parameters in order to produce a reliable and useful fuzzy predictive model.

Fig. 3 shows six MFs of an input depicting the perception in one of the future points (front position). The same input MFs are applied to the other two positions, i.e., the right and left positions.

Once the perception from the surrounding environment has been classified in three human-like descriptions, we need to define the output fuzzy set and the rules to infer a decision for the turning angle of the next step. The output that deduces the next step displacement is

![Fig. 3. Fuzzy membership functions of the input that indicates the pedestrian's perception from the environmental effects in possible future positions.](image-url)
A fuzzy model with three fuzzy inputs, each includes six MFs each, leads to a total of 216 (or $6^3$) if-then rules. These rules offer a human-like process of thought for confronting the three criteria and inferring the possible turning angle of the next step direction. Each rule consists of three antecedents and one consequent. An example of the rule is as follows.

$$R_{10} : \text{IF } P_{RP} \text{ is } HA \text{ and } P_{FP} \text{ is } LR \text{ and } P_{LP} \text{ is } LATHENT \text{ angle is move to the right},$$

where $P_{RP}, P_{FP}, P_{LP}$ are linguistic variables and $HA, LR, LA$ are linguistic values.

In this study, the fuzzy inference engine consists of 216 rules that transfer the inputs of the system to the output. These rules guide the pedestrian to walk towards the attractive side and avoid the repulsive stimulation. According to McCrae and Terracciano [37], a pedestrian tends to have a smooth and regular angular displacement, rather than a sudden change of direction. Moreover, these rules are established in a way that in case of having equal stimuli in all three future positions, a pedestrian inclines to walk forward.

**4. The GFS architecture for walking trajectory modelling and simulation**

In general, the knowledge base of a fuzzy rule-based system comprises two sets of information that are related to the variables and parameters of the fuzzy system, known as the data base and the rule base. The data base considers the shapes and parameters of the input and output MFs, while the rule base determines the combination of the inputs and outputs in rule construction to provide a valid result. The GFS combines the GA as an optimisation tool to help develop a more accurate and feasible fuzzy model with respect to the fuzzy set parameters and rules [20]. The GFS implements either separate optimisation of the MF parameters or rules, or simultaneous optimisation of both rules and parameters. The latter is a complex process.

In the proposed fuzzy model, a set of pre-defined MFs is employed initially. The aim is to apply the GA to fine-tune the initial MF parameters and develop optimised MFs according to the trajectory data sets. To achieve this, the MF parameters are adapted according to a fitness function. Fig. 4 depicts an overview of the proposed GFS for optimising the MF parameters.

**4.1. Genetic algorithm**

Evolutionary search algorithms including the GA are considered as one of the most robust search methods in recent years [38–40]. Specifically, the GA has been theoretically and empirically shown to provide a reliable and effective search
procedure in complex search spaces [41]. The GA possesses a number of salient properties as compared with traditional search and optimisation techniques, such as mathematical programming [38,42,43]. Firstly, the GA performs a stochastic search process with probabilistic transition rules, rather than a deterministic search, which is efficient in finding the optimum solution in many complex search spaces. Secondly, the GA takes into account different points in the search space simultaneously, rather than single point search in one iteration as in traditional gradient descent optimisation techniques. This increases the chance of finding the global optimum solution, instead of being stuck at a local one. Thirdly, the GA does not rely on the information about the structure of the parameters or other auxiliary knowledge of the problem. This is owing to the use of the chromosome representation to encode the solution; hence it is not problem dependent [25]. So, the effectiveness and robustness of the GA provides a useful optimisation tool for parameter tuning of the input and output fuzzy sets, as shown in this study. The following section explains the GA parameters used in the experiments.

4.2. Genetic algorithm parameters

In the experiments, the GA with the following parameters is implemented.

- **Population size** ($N_{pop}$): This parameter is fixed to 20 chromosomes, i.e., $c_1^i, c_2^i, \ldots, c_R^i$, for $R = 20$, where each solution $c_i^j$ for $i = 1, \ldots, R$ indicates chromosome $i$ in iteration $I$.
- **Chromosome structure and length**: The chromosome structure is the genetic representation of the solution. Note that the current problem involves three input sets, each with six MFs, and one output set with three MFs. Each MF has two parameters, $(\mu, \sigma)$. As such a fixed-length binary string is used to define a chromosome with 42 genes, i.e., $c_j^i = b_1^i, b_2^i, \ldots, b_m^i$, for $m = 42$. Each gene represents one MF parameter (either $\mu$ or $\sigma$) in sequence, as follows.

$$c = \{\sigma_1, \mu_1, \sigma_2, \mu_2, \ldots, \sigma_{21}, \mu_{21}\}.$$

- **Initial population**: To start the GA, an initial population, $P(0)$, is acquired from the pre-defined MF parameters. It complies with the lower bound, upper bound, and linear constraints of the problem.
- **Evaluation function**: A fitness function, $J(c_i)$, which calculates the Mean Square Error (MSE) between the predicted walking trajectory and the real or simulated trajectory data, is formulated. The aim is to optimise the MF parameters such that the GFS is able to model the steering behaviours of heterogeneous pedestrians with different personal characteristics within a built environment accurately. For performance evaluation, the mean MSE measured from each pedestrian is computed, as follows.

$$J(c_i) = \text{Mean } \langle \text{MSE } (|T_{algorithm} - T_{dataset}|) \rangle, \text{ Training data set } = 1, 2, \ldots, \text{Number of training data set},$$

where $T$ is the turning angle, $T_{algorithm}$ corresponds to the values produced by the GFS, and $T_{dataset}$ are the values measured from real or simulated trajectory data. Based on the fitness function, the GA is used to perform search in a high-dimensional space and find a solution that caters for heterogeneous pedestrians with different characteristics. The MSE is a useful performance indicator, and has been used in other GA-related applications, e.g. tuning the optimal PID control parameters using the GA by using the MSE as the objective function [44].

- **Genetic operators**: Three commonly used operators are adopted, i.e., reproduction, crossover, and mutation. Reproduction produces a new generation from the individual solutions by performing an elitist selection. The selection criterion specifies the parents to generate the next generation. The solutions that satisfy the criterion survive and move to the $(I + 1)$ iteration as the parent chromosomes to produce offspring, while the rest of the chromosomes are deleted from the population; hence the survival-of-the-fittest principle. In this study, the uniform selection function is used. The number of elite solutions ($N_{elite}$) is set to 2. The crossover operator applies the crossover fraction ($p_c$) to combine two good chromosomes (i.e. parents) and produce two offspring. This operator provides an opportunity to exchange information through genes. The crossover probability ($p_c$) is set to 0.6 in this study. Mutation is another commonly used operator that provides an opportunity to generate new genetic structure in the population. Bit mutation means flipping a bit from 1 to 0, or vice versa. Here, the mutation function in Matlab™ [45] provides genes that are feasible in terms of complying with the constraints and boundaries.

- **Termination test**: the number of iterations is represented by $I$, which is limited by a specified value as the stopping criterion. In this study, the search process is said to converge to the optimum value if the average change in the fitness value is smaller than 1e-6.

In short, the GA parameters used are summarised as follows.

- **Number of variables**: $N_{var} = 42$.
- **Population size**: $N_{pop} = 20$.
- **Reproduction strategy**: $N_{elite} = 2$.
- **Crossover probability**: $p_c = 0.6$.
- **Termination criteria**: the average change in the fitness value is smaller than 1e-6.
4.3. Genetic algorithm operation

In the experimental study, we implemented the proposed GFS to predict the pedestrian’s walking trajectory. The MF parameters were fine-tuned by performing optimisation with the following GA fitness function and constraints.

\[ J = \min (\text{Mean} (\text{MSE} (|T_{\text{algorithm}} - T_{\text{training data set}}|))), \text{training data set} = 1, 2, \ldots, \text{Number of training data set}, \]

Subject to,

\[ \mu_i < \mu_{i+1}, \quad \text{for } i = 1, \ldots, 5 \]
\[ \mu_i < \mu_{i+1}, \quad \text{for } i = 7, \ldots, 11 \]
\[ \mu_i < \mu_{i+1}, \quad \text{for } i = 13, \ldots, 17 \]
\[ \mu_i < \mu_{i+1}, \quad \text{for } i = 19, 20 \]
\[ \sigma_i > 0, \quad \text{for } i = 1, \ldots, 21 \]

The GA was used to find the optimal \( \mu \) and \( \sigma \) settings of the input and output MFs. Based on the MF parameters, the fuzzy rule-based model was applied to predict the pedestrian’s walking trajectory. In each fitness function call during the optimisation process, the average MSE was computed for the entire training data set. As such, each individual was tagged with a scalar value from the fitness function. The minimum value signified the best individual of the population.

5. Modelling and simulation study

Fig. 5 shows the entire model execution process, which encompassed two main loops, i.e., prediction of the walking trajectory using the fuzzy system model, and genetic fuzzy system for optimisation of the fuzzy MF parameters using the GA. The GA fitness function takes the deviation between the data samples and the predicted walking trajectories into consideration. As shown in Fig. 5, the algorithm starts by development of the physical environment. Parameters of the social force model and MFs are initialised. The model process is performed the step-by-step trajectory prediction using the social force model and the fuzzy inference engine within the loop encompassing \( t = t + 1 \). This process leads to a complete walking trajectory prediction at the end of the loop. The GA training procedure, as indicated by the loop encompassing Individual \( i = i + 1 \), is then triggered. Note that the loop of the walking trajectory prediction is part of the GA training loop. Within the GA training loop, the fitness function measures the deviation between predicted trajectory points and those from the

Fig. 5. Flow chart of the model execution process for the pedestrian walking trajectory prediction. (a) Genetic fuzzy system that includes fuzzy system model that applies for training, (b) indicates the fuzzy system model that is used within genetic fuzzy system for training, also for path prediction simulation.
data samples. The whole process continues generation by generation until the GA termination criterion is met. The tuned MF parameters are now ready for use in fuzzy system model to predict the walking trajectory for heterogeneous pedestrians using the test samples that are not involved in the training process; hence ascertaining the usefulness of the GFS.

5.1. Data collection

A challenging issue in validating the pedestrian steering behaviour model is data collection [46]. Brogan and Johnson [47] indicated that animated walking trajectories were hardly validated against real trajectories. Indeed, microscopic pedestrian data, especially trajectory data, are rare owing to complexity of data collection and lack of requirement. In addition, the manual process of data refinement for extracting useful information constitutes another issue. However, recent technologies such as laser scanner, micro-electro-mechanical systems, and automatic tracking algorithms require less effort and offer higher accuracy in capturing and processing trajectory data, as compared with manual extraction of pedestrian trajectory from video footage. As an example, Soyguder [48] proposed a vision-based mobile robot tracking system that employed wavelet decomposition and artificial neural network to track human walking trajectories. In this study, two case studies are conducted. In the first case study, we use a motion capture device to collect real pedestrian walking trajectory data within the capturing stage. For the second case study, two experiments with simulated trajectory data collected from the AnyLogic® software [49] are carried out. AnyLogic® has some built-in libraries such as the pedestrian library that is useful for simulating pedestrian steering behaviours. The following subsections describe the experimental setup for the two case studies and the validation process.

5.2. Experimental procedure

5.2.1. Case Study 1

In the first case study, the data set used for model validation was collected using a motion capture device, namely the OptiTrack™ system [50]. OptiTrack™ captured the position of moving pedestrians at a rate of 30 frames per second. The experiment was conducted in a hallway of a laboratory, which was rectangular in shape with 7 m (length) × 2 m (width) including an obstacle 70 cm (length) × 60 cm (width). A total of 25 participants walked several times from a specific origin to the destination in a controlled experiment. In each frame, the position and orientation angles of the pedestrian were extracted. A total of 74 samples were collected from heterogeneous pedestrians during the experiment.

Note that the proposed GFS comprises a large (i.e. 42) unknown parameters, while the training data set contains only a small (i.e. 74) samples of pedestrian walking trajectories. From the machine learning perspective, if the model contains a high degree of freedom while a low number of training data is available, then the probability of over-fitting is high [51,52]. To solve the issue associated with over-fitting with a small training data set, the n-fold cross validation technique was used [53]. With this technique, it is possible to achieve a more accurate validation using limited training and test data samples [1,46]. In this study, however, the reverse cross validation technique [54] was adopted owing to the long training time of the GA optimisation process (as reported in the section 5.3). As such, in Experiment 1, the real data samples were divided into 15 subsets, and one subset was used for training while the remaining was used for test. This reverse cross-validation technique was repeated 15 times, each time with a different data subset for training.

5.2.2. Case Study 2

In the second case study, two experiments, i.e. Experiments 2 and 3, were conducted. Two simulated trajectory data sets were collected using the AnyLogic® software. In Experiment 2, a hexagonal-shaped environment with two obstacles was simulated. In Experiment 3, the same environment but with one obstacle was simulated. A total of 30 data samples (divided into 6 subsets, each with 5 trajectory samples) were collected for training purposes. For performance evaluation, a total of 100 and 125 new, previously unseen trajectory data were used as the test samples in Experiments 2 and 3, respectively. Similar to Experiment 1, the reverse cross validation technique was applied for performance evaluation.

5.3. Results and discussion

5.3.1. Case Study 1: real data (Experiment 1)

Figs. 6 and 7 show the overall results of 15 different simulation runs. The fitness value (measured in square metre) indicates the MSE between the pedestrian position in each time step and the corresponding point from the real trajectory. As shown in Fig. 6, all the fitness values obtained were smaller than 0.05, except the eighth run (close to 0.1). A detailed analysis revealed that this discrepancy was caused by outliers, i.e., an unexpected behaviour of a participant during the experiment. Fig. 7 shows the training time of 15 subsets. As explained in Sub-section 4.2., the GA process terminated when the improvement in the fitness value was smaller than the tolerance. During the optimisation process, the GA search process converged after 51 iterations. The computational time varied from one run to another. Note that $\text{MSE}_{\text{test original MF}}$ and $\text{MSE}_{\text{test tuned MF}}$ indicates the MSE of the test data set when the optimised and original MF parameters were used for prediction, respectively. Fig. 8 depicts these two MSE scores, which signify the difference between the original model and the optimised one by using the GA.
Figs. 9 and 10 show the best individuals as well as the best, worst, and mean values of the fitness function in each generation for the first run. Fig. 9 indicates the value of optimised MF parameters obtained from the first run. As can be seen in Fig. 10, the GA search process converged to the best fitness function score after 50 iterations, when the stopping criterion was satisfied. A discussion on the improvement of the results and its statistical analysis are as follows.

Fig. 11 depicts the hallway with the obstacle as well as all 74 real walking trajectories collected within the environment for Experiment 1, as explained in Section 5.2.1.

The real data samples were divided into 15 subsets, one subset was used for training while the remaining was used for test. This procedure was repeated 15 times, each time with a different data subset for training, i.e. the reverse cross-validation procedure. The test data set was used to evaluate the fuzzy model using both pre- and post-optimised MFs. As shown in Table 1, the MSEs obtained from the optimised MFs are smaller than those from the original MFs. In other words, the GFS performed better than the initial fuzzy model.

The bootstrap analysis was performed to measure stability of the results by resampling the MSE scores from both optimised and original MFs. In the case of having a small sample size with an unknown distribution, the bootstrap method is useful for reconstructing a large number of samples that follow a similar distribution as that of the original data [55]. The bootstrap mean and standard deviation of the optimised MFs are 0.0214 and 0.0011, respectively. These values are close to equivalent statistics calculated, as shown in Table 1. The last column of Table 1 indicates the improvement in percentage when the optimised MFs are applied. Except for two subsets (i.e. 5 and 8), all the results show positive improvements with the optimised MFs. The maximum improvement is 24.3% with subset 11.
Fig. 12 shows the predicted walking trajectories of a participant. The red line indicates the real trajectory from the collected data, the blue dots show the trajectory generated by the fuzzy model with original MFs, and the black stars depict the trajectory produced by the GFS with optimised MFs. As can be clearly seen, the GFS generated a trajectory that closely matched the real one. To sum up, the results from the fuzzy model tuned by GA were better than those from the initial fuzzy model.

5.3.2. Case Study 2: synthetic data (Experiments 2 and 3)

The second case study was conducted in a hexagonal environment of 16 m (length) × 14 m (width), as shown in Fig. 13. The environment contained two obstacles, i.e., a larger obstacle of 3 m (length) × 9 m (width) and a smaller obstacle of 5 m (length) × 2 m (width) (see Fig. 13). A total of 100 simulated trajectory data were collected within this environment using the AnyLogic® software. All the walking trajectories of the collected samples within the environment are plotted in Fig. 13.

In comparison with the first case study, the second case study involved a larger environment that correspondingly consumed longer computational time for training. Table 2 indicates the training time for 6 different training subsets, each with 5 trajectory data samples. The optimised MF parameters from these 6 training subsets were used for evaluation with new, previously unseen test trajectory data samples.

Fig. 14 depicts an example of three walking trajectories from Experiment 2. The red line represents the walking trajectory generated by the AnyLogic® software, the blue dots indicates the prediction using the original MF parameters, while the black stars show the prediction using the optimised MF parameters. As can be seen, the predicted trajectory using the
optimised MF parameters, is closer to the generated trajectory from the AnyLogic® software; hence ascertaining the usefulness of the GFS.

In Experiment 3, the same environment as in Experiment 2, but with the smaller obstacle removed, was used. The rationale was to evaluate the effect of a change in the environment towards the trained GFS. As such, a total of 125 new trajectory data were generated using the AnyLogic® software for evaluation purposes. Fig. 15 shows all the trajectory data.

Fig. 16 indicates the fitness function scores of 6 runs in Experiment 3.

To assess the usefulness of the GFS in Experiment 3, the predicted trajectories with the original and optimised MF parameters together with the generated trajectory from the simulation AnyLogic® software are depicted in Fig. 17. The red straight line indicates the walking trajectory generated by the AnyLogic® software. The blue dots and black stars represent the predicted trajectory using the original and optimised MF parameters, respectively. It can be clearly seen that the GFS was able to produce a more accurate prediction for the walking trajectory as compared with the original fuzzy model.

Table 1
Prediction results of experiment 1 in 15 runs.

<table>
<thead>
<tr>
<th>Run</th>
<th>Training</th>
<th>MSE training (m²)</th>
<th>MSE test with optimised MFs (m²)</th>
<th>MSE test with original MFs (m²)</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Subset 1</td>
<td>0.031</td>
<td>0.0233</td>
<td>0.0235</td>
<td>0.85</td>
</tr>
<tr>
<td>2</td>
<td>Subset 2</td>
<td>0.0286</td>
<td>0.0209</td>
<td>0.0216</td>
<td>3.24</td>
</tr>
<tr>
<td>3</td>
<td>Subset 3</td>
<td>0.0174</td>
<td>0.0214</td>
<td>0.0229</td>
<td>6.55</td>
</tr>
<tr>
<td>4</td>
<td>Subset 4</td>
<td>0.0097</td>
<td>0.0229</td>
<td>0.023</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>Subset 5</td>
<td>0.045</td>
<td>0.0224</td>
<td>0.0223</td>
<td>-0.45</td>
</tr>
<tr>
<td>6</td>
<td>Subset 6</td>
<td>0.0282</td>
<td>0.0211</td>
<td>0.0229</td>
<td>7.86</td>
</tr>
<tr>
<td>7</td>
<td>Subset 7</td>
<td>0.0457</td>
<td>0.0198</td>
<td>0.0213</td>
<td>7.04</td>
</tr>
<tr>
<td>8</td>
<td>Subset 8</td>
<td>0.0944</td>
<td>0.0192</td>
<td>0.0191</td>
<td>-0.52</td>
</tr>
<tr>
<td>9</td>
<td>Subset 9</td>
<td>0.0464</td>
<td>0.0199</td>
<td>0.0204</td>
<td>2.45</td>
</tr>
<tr>
<td>10</td>
<td>Subset 10</td>
<td>0.0085</td>
<td>0.0207</td>
<td>0.0224</td>
<td>7.59</td>
</tr>
<tr>
<td>11</td>
<td>Subset 11</td>
<td>0.0283</td>
<td>0.0218</td>
<td>0.0288</td>
<td>24.31</td>
</tr>
<tr>
<td>12</td>
<td>Subset 12</td>
<td>0.0228</td>
<td>0.0218</td>
<td>0.0270</td>
<td>19.26</td>
</tr>
<tr>
<td>13</td>
<td>Subset 13</td>
<td>0.0086</td>
<td>0.0220</td>
<td>0.0225</td>
<td>2.22</td>
</tr>
<tr>
<td>14</td>
<td>Subset 14</td>
<td>0.0316</td>
<td>0.0198</td>
<td>0.0215</td>
<td>7.91</td>
</tr>
<tr>
<td>15</td>
<td>Subset 15</td>
<td>0.0085</td>
<td>0.0225</td>
<td>0.0264</td>
<td>14.77</td>
</tr>
</tbody>
</table>

Mean 0.0303 0.0213 0.0230
SD 0.0215 0.0012 0.0025
Mean-bootstrap 0.0301 0.0214 0.0229
SD-bootstrap 0.0201 0.0011 0.0024
### Table 2
The overall results of training in 6 different runs in the second case study.

<table>
<thead>
<tr>
<th>Run</th>
<th>Training</th>
<th>No of data samples</th>
<th>Training time (hr)</th>
<th>MSE training (m$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Subset 1</td>
<td>5</td>
<td>3.4</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>Subset 2</td>
<td>5</td>
<td>3.5</td>
<td>0.55</td>
</tr>
<tr>
<td>3</td>
<td>Subset 3</td>
<td>5</td>
<td>3.8</td>
<td>0.35</td>
</tr>
<tr>
<td>4</td>
<td>Subset 4</td>
<td>5</td>
<td>3.45</td>
<td>0.29</td>
</tr>
<tr>
<td>5</td>
<td>Subset 5</td>
<td>5</td>
<td>3.45</td>
<td>0.31</td>
</tr>
<tr>
<td>6</td>
<td>Subset 6</td>
<td>5</td>
<td>3.42</td>
<td>0.27</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td>0.12</td>
</tr>
</tbody>
</table>

**Fig. 14.** Comparison of three walking trajectories generated by the AnyLogic© software, predicted using the original and optimised MF parameters of Experiment 2, respectively.

**Fig. 15.** The walking trajectories of 125 pedestrians within the simulated environment of Experiment 3.

**Fig. 16.** The fitness function scores from 6 different runs in Experiment 3.
Table 3 shows the MSE rates computed throughout the evaluation process of both Experiments 2 and 3. The percentages of improvement between the original and optimised MF parameters for both experiments are also shown in Table 3. The percentages of improvement range from 8% to 17% and from 6% to 12% in Experiments 2 and 3, respectively. As can be seen, amongst the 6 training subsets in both experiments, the lowest MSE with the optimised MF parameters was produced by the first training subset.

6. Conclusions

A GFS that is able to model and predict the walking trajectories of heterogeneous pedestrians within a built environment has been developed. The GFS entails a representation of diverse and uncertain aspects of pedestrian perceptions from the surrounding environment by the fuzzy model. Inclusion of the environmental design and the corresponding socio-psychological stimuli into the fuzzy model is accomplished by employing the Helbing SFM approach. Besides that, optimisation of the MF parameters is performed using the GA. To evaluate the effectiveness of the GFS, two case studies have been conducted. In case study 1, real pedestrian trajectory data collected using a motion capture device have been used. In case study 2, two experiments using simulated trajectory data generated by the AnyLogic software have been employed. The MSE of the fitness function scores indicate that the GA is able to minimise the deviation between the predicted walking trajectories and the real/simulated trajectory data from heterogeneous pedestrians. To better utilise the data samples for training and testing purposes, the reverse cross validation method has been implemented. In most of the evaluation runs with different test samples, the experimental outcomes record improvements by using the GA-optimised MF parameters as compared with the original MF parameters; hence ascertaining the usefulness of the GFS in modelling and prediction of pedestrian’s walking trajectory in normal conditions in built environments.

The prediction of walking trajectory in relation to the environmental design and floor layout is beneficial to improve the quality of indoor spaces. Generating realistic human walking trajectory is difficult, laborious, and time-consuming [1]. Realistic results have been obtained by using the proposed GFS in this study. As such, it would be interesting to further investigate the application of the GFS to other problems, which include personal geo-positioning and navigation, efficient design of urban public areas like shopping and terminals. On the other hand, it is plausible to extend the GFS to a macroscopic scale by incorporating the dynamic network flow [9,56] approach for evacuation scenarios. As such, complex tasks that are related to transportation and pedestrian safety issues can be investigated, whereby the GFS can be used to help design more efficient public areas for crowd evacuation in emergency situations.
On the other hand, another direction of research is to investigate the scenario that involves multiple pedestrians at the same time in the built environment. Modelling moving objects in the environment requires the consideration of interactive forces between two or more moving elements, as well as their interactions with the environmental stimuli. In this case, it would be useful to examine the feasibility of employing the GFS for modelling and simulation of multiple pedestrians in a built environment.

**Acknowledgements**

This research is fully supported by Australian Research Council Linkage Grant LP0776826 and Centre for Intelligent Systems Research (CISR) at Deakin University.

**Appendix A**

List of key abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABM</td>
<td>Agent-based model</td>
</tr>
<tr>
<td>CA</td>
<td>Cellular Automata</td>
</tr>
<tr>
<td>FP</td>
<td>Level of environmental stimuli in the Front Position</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic algorithm</td>
</tr>
<tr>
<td>GFS</td>
<td>Genetic fuzzy system</td>
</tr>
<tr>
<td>GNN</td>
<td>Genetic Neural Network</td>
</tr>
<tr>
<td>HA</td>
<td>High attractive</td>
</tr>
<tr>
<td>HR</td>
<td>High repulsive</td>
</tr>
<tr>
<td>LP</td>
<td>Level of environmental stimuli in the left position</td>
</tr>
<tr>
<td>LA</td>
<td>Low attractive</td>
</tr>
<tr>
<td>LR</td>
<td>Low repulsive</td>
</tr>
<tr>
<td>MFS</td>
<td>Membership functions</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>MSE_tuned MF_test</td>
<td>Mean Square Error score for test data set with tuned membership function</td>
</tr>
<tr>
<td>MSE_original MF_test</td>
<td>Mean Square Error score for test data set with original membership function</td>
</tr>
<tr>
<td>MA</td>
<td>Medium attractive</td>
</tr>
<tr>
<td>MR</td>
<td>Medium repulsive</td>
</tr>
<tr>
<td>RP</td>
<td>Level of environmental stimuli in the Right Position</td>
</tr>
<tr>
<td>SFM</td>
<td>Social force model</td>
</tr>
</tbody>
</table>

**Appendix B**

See Table A2.1.

**Table A2.1**

Fuzzy logic inference rules for the movement direction of pedestrian steering behaviour.

<table>
<thead>
<tr>
<th>Rule</th>
<th>IF-THEN statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 1</td>
<td>IF RP is High-attr AND FP is High-attr AND LP is Medium-attr THEN Turn angle is Not left</td>
</tr>
<tr>
<td>No. 10</td>
<td>IF RP is High-attr AND FP is Medium-attr AND LP is Low-repul THEN Turn angle is Right</td>
</tr>
<tr>
<td>No. 55</td>
<td>IF RP is Medium-attr AND FP is Low-repul AND LP is High-attr THEN Turn angle is Left</td>
</tr>
<tr>
<td>No. 110</td>
<td>IF RP is Low-repul AND FP is High-attr AND LP is Medium-attr THEN Turn angle is Forward</td>
</tr>
<tr>
<td>No. 200</td>
<td>IF RP is High-repul AND FP is Low-repul AND LP is Medium-attr THEN Turn angle is Left</td>
</tr>
<tr>
<td>No. 216</td>
<td>IF RP is High-repul AND FP is High-repul AND LP is High-repul THEN Turn angle is Forward</td>
</tr>
</tbody>
</table>