Personal Greetings: Personalizing Robot Utterances Based on Novelty of Observed Behavior

Dylan F. Glas, Kanae Wada, Masahiro Shiomi, Takayuki Kanda, *Member, IEEE,* Hiroshi Ishiguro, *Member, IEEE,* Norihiro Hagita*, Senior Member, IEEE*

ABSTRACT

One challenge in creating conversational service robots is how to reproduce the kind of individual recognition and attention that a human can provide. We believe that interactions can be made to seem more warm and humanlike by using sensors to observe a person's behavior or appearance over time, and programming the robot to comment when it observes a novel feature, such as a new hairstyle, or a consistent behavior, such as visiting every afternoon. To create a system capable of recognizing such novelty and typicality, we collected one month of training data from customers in a shopping mall and recorded features of people's visits, such as time of day and group size. We then trained SVM classifiers to identify each feature as novel, typical, or neither, based on the inputs of a human coder, and we trained an additional classifier to choose an appropriate topic for a personalized greeting. An utterance generator was developed to generate text for the robot to speak, based on the selected topic and sensor data. A cross-validation analysis showed that the trained classifiers could accurately reproduce human novelty judgments with 88% accuracy and topic selection with 95% accuracy. We then deployed a teleoperated robot using this system to greet customers in a shopping mall for three weeks, and we present example interactions and results from interviews showing that customers appreciated the robot's personalized greetings and felt a sense of familiarity with the robot.

Keywords: human-robot interaction;social robots;novelty detection;greetings;personalized interaction

1 INTRODUCTION

In today's high-speed world of technology and mass-production, our daily lives are beginning to lose the warmth of personal human interactions. What was once an everyday experience of going to a local shop and being greeted personally – "I like your new haircut," "Is that a new jacket? It looks great," or "Who is this with you today? Your granddaughter?" is fading into a nostalgic memory of a lost era. Perhaps this is inevitable in our modern world - a replaceable cashier working shifts at a busy chain store has little time or inclination to remember hundreds of faces or engage in personal conversations. Yet there is a perception shared by many that something important is being lost.

1.1 Automating Personal Interactions

Ironically, as the roles of people working in such establishments become more "robotic" in nature, service robots performing similar roles might be able to provide some part of that missing "personal touch". By recording sensor observations of the appearance and behavior of individuals over time in a database, a robot could remember thousands of customers and be able to comment when something is unusual or interesting. In this work, we have developed a system to enable a service robot to greet people with such personal comments, in the hopes that technology can contribute to bringing the feeling of personal service back to the experience of daily interactions.

The core challenge here is to reproduce human judgment regarding what constitutes an event worth commenting on. We propose that by training classifiers to identify the novelty or sameness of various attributes detected by sensors over successive interactions, it will be possible to automate the process of generating these personal comments. For example, a robot could say, "I'm surprised to see you here on a Tuesday" to a person

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

who typically comes only on Wednesdays, or even something as simple as, "It's nice to see you every morning" to express that the robot has noticed a person always comes in the morning.

In this report, we present the design and implementation of such a system, and we present a proof-of-concept field trial conducted in a shopping mall where customers were personally greeted by a robot using this mechanism.

1.2 Related Work

1.2.1 Personal relationships in human-robot interaction

Earlier works have reported several cases where people established personal relationships with robots. One of the famous examples is with AIBO, a dog-like robot developed by SONY. It was reported that some people expressed that they attributed mental state to AIBO and perceived social rapport with it [1]. In another study [2], it was found that a majority of children attributed mental state to a human-like robot, personified it, and even established a moral relationship with the robot. For example, they considered it important to treat the robot fairly. In other work, a seal-like robot has successfully been used to comfort elderly people [3], and elderly people have expressed their personal matters and showed attachment with a conversational humanoid robot [4]. It has also been discussed that loneliness could promote people's anthropomorphization toward non-human entities, hence loneliness could facilitate establishment of personal relationship with robots [5].

Furthermore, researchers have studied techniques for screen agents and robots to promote developing personal relationships with people. Bickmore and Picard investigated human communication for personal relationships, and implemented similar 'relational behaviors' used in human communication into a screen agent [6]. The agent greeted users, called them by their first name, used humor, engaged in small talk and empathetic dialog, and performed continuity behaviors (talking about things which happened while being absent). While the agent in this work was bounded within a computer screen, there are some other works that connect robots' sensing and relational behavior. In the Roboceptionist robot study [7], the robot identified users using RFID tags, perceived users' engagement level from their relative locations to the robot, and talked about a series of daily episodes prepared in advance. Leite et al., developed a model of empathy in the context of a chess-playing robot for long-term relationships [8]. Kidd developed a personal robot that conducted long-term interactions with users and tried to establish relationships to help motivate them in a weight loss program [9], and Pan et al. used robots for greeting hotel guests and providing information about the hotel [10].

Non-verbal behaviors were also studied for building rapport with users. Gaze, gestures, and spatial formation have been studied. More specifically, Riek et al, developed a robot that mimic users' head motions [11]. Similarly, it was reported that cooperative non-verbal behaviors provides empathetic impression of a robot [12].

1.2.2 Use of memory or interaction history

Some studies have investigated robot behaviors for long-term or repeated interactions. Sabelli et al. placed a robot in an elderly care center for 3.5 months and identified robot behaviors, including addressing the patients by name, which helped to build rapport over time, resulting in a positive impression of the robot [13]. A robot in an elementary school deployment also called students by name and used some elements of interaction history to determine its behaviors [14], and a study in a shopping mall generated utterances based on conversation history, in which the robot spoke about topics it had discussed previously, such as shops it had recommended in the previous interaction [15]. When the robot showed that it remembered a person's name or information from its previous interactions with them, the people often showed positive responses and reported a feeling of familiarity with the robot.

Similar to these works, our work attempts to create personalized utterances which convey a sense that the robot remembers a person individually. However, our approach is not based on dialog history, but on using the history of information observed from sensors to detect a person's behavior patterns over time, and to generate customized utterances for a specific person based on these patterns. By doing so, our work provides a pioneering demonstration of a robot's capability of replicating human-like daily greeting behaviors. Our study contributes to this body of work by providing an automated way to generate individualized greeting behaviors for a large set of people based on observing changes or tendencies in their behavior over time, in a way that could be used autonomously or in conjunction with a partially or fully teleoperated system.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

1.2.3 Novelty detection

Another aspect of our work is related to novelty, or anomaly, detection, which has been widely studied in machine learning. A survey of anomaly detection techniques can be found in [16]. However, most approaches to anomaly detection, *e.g.* for identifying faults in a system, consist of training a model or classifier with a "normal" data set and subsequently identifying data points which do not conform to that set. In the field of robust statistics, novelty detection techniques are used for rejecting outliers in a dataset in order to avoid corruption of a model fit [17]. In robotics, novelty detection has been used for applications such as environment inspection using mobile robots [18], imitating human gestural behaviors [19], and safety monitoring for assistive elderly care robots [20].

In this study, our objective differs somewhat from typical novelty detection tasks in a few ways. First, novelty detection is typically used for identifying outliers from a pattern or model, based on the assumption that a pattern or model exists in the first place. Such cases are generally represented as a binary classification, where a data point can be described as "normal" or "anomalous", with some confidence measure.

However, although some human behavior conforms to habits and routines, other behavior may simply be random, with no easily perceptible pattern. Part of our goal is to imitate a human's judgment as to which of three mutually-exclusive classes an event should be categorized as: novel (clearly deviating from an established pattern), typical (clearly conforming to an established pattern), or neither (either no pattern exists, or the event is neither clearly typical or clearly novel). For our work, this judgment is based on two questions of human perception: whether a pattern is clearly perceptible, and whether a feature is perceived to be clearly typical or clearly novel to the extent that it could be remarked upon in conversation.

A second distinction between this work and traditional anomaly detection is the relative frequency of novelty in the data set. A significant concern in the detection of anomalies is the fact that they are rare, and it may be difficult or impossible to create a data set exhibiting some anomalous behavior, for purposes of training a detector. However, in our case, we are targeting variations of human behavior which are relatively common, for example, a visitor coming to a shopping mall a bit earlier or later than they typically do. Because the novel situations are not unpredictable outliers or system faults, but merely notable deviations of behavior which occur during normal interactions, we assume that many instances of the novel conditions are included in the training data set.

To avoid confusion with traditional problems of outlier or anomaly detection, we will refer to our technique as "pattern consistency classification" in this paper.

1.3 Research Overview

In this study, we first conducted a data collection of people's behavior in a shopping mall, without using a robot. For this data collection we used laser range finders to track pedestrian motion, and we used facial recognition to identify repeat visitors. We collected several pieces of information about their visits, to be used as candidate features for the robot to potentially comment about if these features were observed to be particularly novel or typical. Specifically, we recorded the following information for each visit: time of day, day of the week, number of days since their last visit, number of times they returned in one day, walking speed, and the number of people visiting together. A description of this system architecture and the details of these features are presented in Sec. 2.

For each visit, we had a human coder evaluate each of these features in comparison with that visitor's history and annotate the data, indicating her judgment of whether a feature was novel enough or typical enough to comment about, as well as which of the six features she judged to be most appropriate for comment. We trained SVM classifiers to reproduce this judgment, and we created a system for generating actual utterances based on the observed and historical data for each feature. The training and evaluation of these classifiers are presented in Sec. 3.

Finally, we conducted a field demonstration, in which a robot interacted with customers in the shopping mall for three weeks. When a visitor returned three or more times, the robot used the proposed system to generate a personalized utterance for that visitor, based on the classifiers and utterance generator we created. We present the details of the field demonstration in Sec. 4, and we present discussion and conclusions from this study in Sec. 5.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

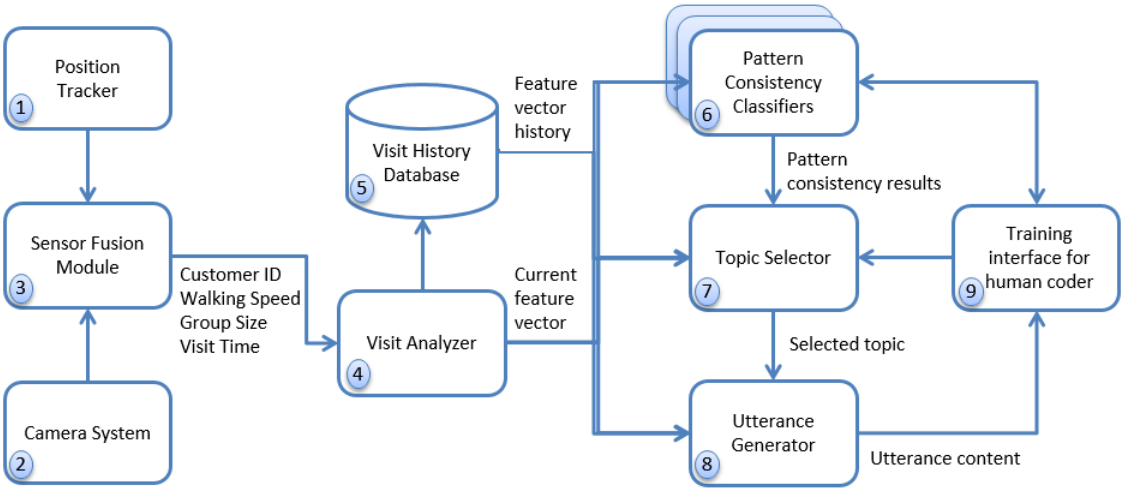

Figure 1. Data flow for training phase. System elements are numbered for reference in the text. Sensor data is used to generate feature vectors which are used as inputs to pattern consistency classifiers and topic selector. A human coder provides training data for the classifiers

2 System Implementation

This section will present the overall architecture and design of the proposed system for generating personalized utterances, including systems for data collection, classifier training, and utterance generation.

2.1 System Architecture

To create a basis for categorizing "typical" and "novel" behavior patterns, we developed the system illustrated in Fig. 1. In this system, a **position tracker** (marked as 1 in the figure) and **networked cameras** (2) using face detection are used to collect data from customers visiting a shopping mall. Next, **sensor fusion** (3) is performed to assign a customer ID from face detection to each observed person's trajectory. Then, data about that person's visit is recorded by a **visit analyzer** (4), which generates statistics about the person's current visit and updates the **visit history database** (5). Current and historical visit data for that person are then used as inputs to **pattern consistency classifiers** (6). The classification results are then used as inputs for **topic selection** (7) and finally **utterance generation** (8) for the robot. In an offline phase, these classifiers are **trained** by a human coder (9) so the robot will be able to reproduce human judgment regarding typical and atypical behaviors and topic appropriateness. In this section, we will present each of these elements.

2.2 Sensing Framework

We set up our sensor network (elements 1 and 2 in Fig. 1) in the entrance hall of a shopping mall, covering an area of approximately 15m by 15m, as shown in Fig. 2. Pedestrian tracking was performed using the ATRacker[1] human tracking system presented in [21], utilizing 8 laser range finders (LRF's) mounted on portable poles placed around the environment. This system combines range data from multiple sensors to track the trajectories of pedestrians in the space using particle filters, and can provide position data within 6 cm error at a data rate of 37 Hz. Since this data comes only from laser range finders, it is anonymous and cannot provide personal identification.

To add the capability of identifying individuals, we installed two pan-tilt-zoom network cameras in the ceiling, aimed to observe the two sets of sliding doors. These cameras were used for face detection, to identify when a specific visitor entered the shopping mall. Video feeds from these cameras were sent to a server running OKAO

[1] ATRacker is a product of ATR Promotions: http://www.atr-p.com/HumanTracker.html

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

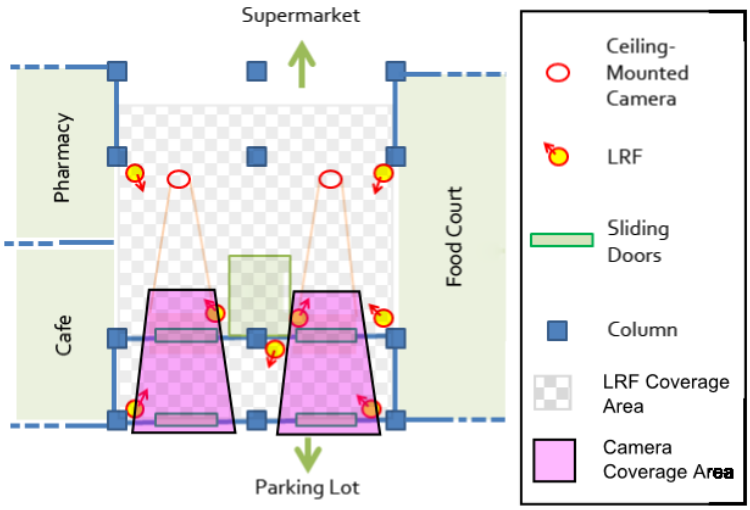

Figure 2. Layout of the data collection environment in the shopping mall entrance

Vision[2] face recognition software. OKAO Vision uses Gabor wavelet transform coefficients as feature values and uses SVM for classification. The performance of face recognition was reported at 89.8% in [22].

The database used for face recognition was trained using data from a three-day subset of the data collection. Faces observed on those days were manually registered in the face recognition software and assigned sequential ID numbers, so that we could identify when a previous visitor returned to the shopping mall. Since the video data recorded by this system contains personally-identifiable information, we obtained special permission from the shopping mall to collect this data on the condition that it be protected and used only for research purposes. A sign was placed in the area to inform visitors that this data was being collected. Furthermore, during the robot experiments, we placed a sign by the robot informing participants that video and audio of people who chose to interact with the robot would be recorded and used in research.

2.3 Sensor Fusion

In the sensor fusion process (element 3 in Fig. 1), anonymous trajectory data from the ATRacker system is combined with the face ID's captured by the face detection system, as illustrated in Fig. 3. The 2D tracking data was converted to a 3D position by assuming an average height of 160 cm for each pedestrian's face. This 3D spatial position in floor coordinates was then projected into the 2D screen coordinate system of the camera. A nearest-neighbor matching was performed between the person's projected position and the screen coordinate positions of all detected faces. Face identification can be slightly unstable from frame to frame, so only the most frequently observed face ID corresponding to a given trajectory was used.

2.4 Visit Analysis

After performing sensor fusion, we mark the trajectory and face detection data as a discrete "visit" and extract several features from the sensor data available for each visit (elements 4 and 5 in Fig. 1). This includes features directly computed from the raw sensor data, as well as statistical descriptors characterizing the history of those features for a specific individual, as described below.

Unfortunately many of the interesting features we would like to use, such as hairstyle or clothing style, are at present quite difficult to detect in a reliable way. We anticipate that systems for detecting these features will

[2] OKAO Vision is a product of OMRON Corporation: http://www.omron.com/r_d/coretech/vision/okao.html

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

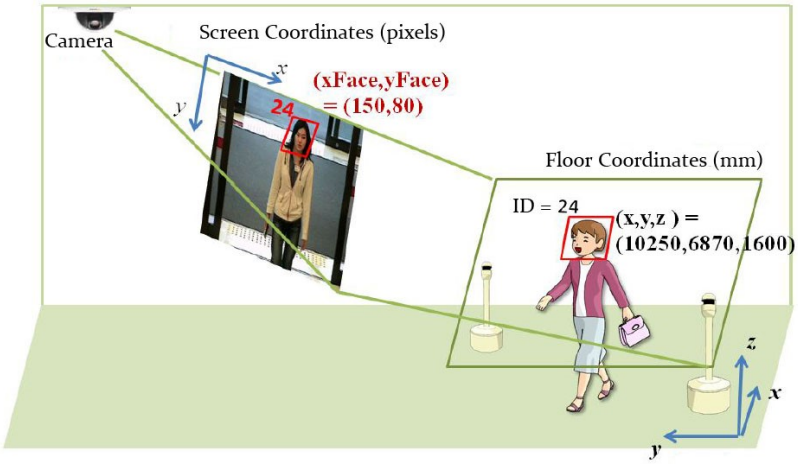

Figure 3. Illustration of sensor fusion between a camera and the position tracking system. The positions of people in the floor coordinate system detected by the position tracker are projected into the screen coordinates of the camera system. Nearest-neighbor matching is then used to associate face ID's with anonymous pedestrians detected by the tracking system

become available in the future, but for the current study we chose features which are easier to identify with available technology.

2.4.1 Feature Definitions

Six features were selected, and each feature was defined as continuous, categorical, or integer type. The entire set of features is shown in Table 1.

TABLE 1. FEATURES USED IN DATA COLLECTION AND CLASSIFIER TRAINING.

Feature	**Variable Name**	**Type**	**Source**
Time of day	t_{visit}	Continuous	Face detector
Day of week	$weekday$	Categorical	Face detector
Days since last visit	*dslv*	Integer	Face detector
Same day visit count	*sdvc*	Integer	Face detector
Walking speed	v_{walk}	Continuous	Pedestrian tracking system
Group size	n_{group}	Integer	Pedestrian tracking system

Although the face detection system produces many kinds of data, such as gender and age estimation, these are not variables that change in ways that typically influence interactions on the time scale of weeks. We did not consider emotion estimation to be reliable enough for use, so we used the time of detection as the primary feature of interest. Specifically, we used the time of the visit in four ways: time of day, day of week, days since the customer's last detected visit, and "same day visit count," that is, the number of unique times that customer has been detected entering the shopping mall on the same day.

2.4.2 Time-based features

Time of day: The time of the person's visit (t_{visit}) was defined in seconds, by subtracting the timestamp corresponding to the previous midnight from the timestamp of the first detection of a person entering the door. Values ranged from 10am (36000) to 7pm (68400).

Day of week: This was our only categorical feature. Although it could arguably be represented as an ordinal quantity, our intuitive rationale was that socially, days of the week are often treated as independent entities. If a person's typical behavior is to visit on a Wednesday, it would be no more unusual for them to come on a Friday than for them to come on a Thursday. Data was only collected on weekdays. Thus, we defined a feature $weekday \in \{Mon, Tue, Wed, Thu, Fri\}$.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

Days since last visit: The integer number of days since the person's last visit was computed. Note that data was only collected on weekdays, so for purposes of computing "days since last visit," we did not count weekend days. Hence, a visit on Friday followed by a visit the following Monday was given a "days since last visit" value of 1. We defined this feature as *dslv*, with integer values. In total, 36% of unique visits had a *dslv* value of 1, and the values ranged from 1 to 22, with an overall average of 3.6 days and a standard deviation of 3.5 days.

Same day visit count: We observed that sometimes customers would return multiple times on the same day. We separated this phenomenon from the "days since last visit" metric because these would represent socially different situations and hence result in different utterances from the robot, and thus we recorded the number of visits detected on the same day as a feature, designated *sdvc*, with integer values. To prevent these multiple-visit situations from corrupting the statistics of other features, we updated the "days since last visit" field only on the first visit of each day, and for the statistical descriptors like mean and standard deviation described below, we considered only the first visit of each day in our computations for "days since last visit" and "day of week".

2.4.3 Tracking-based features

We additionally used the position tracking system to identify two features:

Walking speed: Although it is a very simple measurement, we considered that it could be possible to make a rough judgment of a person's mood based on their walking speed. A person walking quickly could be in a hurry, whereas a person walking slowly could be tired. Thus if a person shows a walking speed which is very different from usual, it might be a feature worthy of comment. For each visit, we measured the visitor's walking speed v_{walk} based on the tracking system's output and stored this value in m/s. The mean value of v_{walk} in our data set was 0.96 m/s, with a standard deviation of 0.24 m/s.

Group size: Another feature which we thought would be interesting for inclusion in this system was the number of people visiting the shopping mall together. For example, if a person who usually visited the mall alone came with a companion one day, it might be interesting for the robot to comment on this fact. Using the data from the tracking system, we used a group detection algorithm to identify social groups [23]. We designated this feature as n_{group}. Overall, 90% of registered customers were shopping alone, 9% were in pairs, and less than 0.5% had group sizes of 3 or 4.

2.4.4 Statistical Descriptors

To characterize the pattern of a given feature over the history of an individual's visits, we compute statistical descriptors for each feature, depending on its type. These statistical descriptors are included in the feature vector used by the classifiers.

For continuous and integer quantities, such as time of day, we calculate the mean and standard deviation for the feature. Intuitively, a feature with a low standard deviation will be more likely to represent a consistent behavior pattern. In the feature vector, we included the difference between the current value and the average value, as well as the standard deviation.

For categorical quantities, we compute the mode (the most frequently observed category) and the percentage of observations from that visitor's history which agree with the mode. Intuitively, a high percentage of observations in agreement with the mode will indicate a consistent behavior pattern. "Day of week" was the only categorical feature, so we calculated the most frequent day of the week from the visitor's history, and for the feature vector we calculated the percentage of visits which occurred on that day of the week as well as a Boolean field indicating whether or not the current visit occurred on that day of the week.

For "same day visit count", the statistical descriptors were defined differently. Unlike the other features, the value for this feature did not consist of independent events, but it incremented at each visit on the same day. Thus, it is not meaningful to calculate the mean of this value *per visit*, and instead we considered the maximum value *per day* and calculated the mean of that value over all days. For this classifier we used the current value of *sdvc* and the mean value of *sdvc* per day in the visit history as the statistical descriptors.

2.4.5 Feature vectors

The entire set of features included in the training vectors for each Stage 1 classifier is summarized in Table 2.

TABLE 2. FEATURES USED FOR THE STAGE I CLASSIFIERS.

Classifier	**Features**	
Time of day	$t_{visit} - \overline{t_{visit}}$	$\sigma^2_{t_{visit}}$
Day of Week	$\%mode\ (weekday)$	$weekday == mode(weekday)?$
DaysSinceLastVisit	$dslv - \overline{dslv}$	σ^2_{dslv}
SameDayVisitCount	$sdvc$	$\overline{sdvc}\ (per\ day)$
WalkingSpeed	$v_{walk} - \overline{v_{walk}}$	$\sigma^2_{v_{walk}}$
GroupSize	$n_{group} - \overline{n_{group}}$	$\sigma^2_{n_{group}}$

The Stage 2 classifier used as inputs all of the above features, as well as the "typical", "neither", or "unusual" classification output from each of the Stage 1 classifiers.

2.5 Pattern Consistency Classification

Once features have been collected, it is possible to conduct **pattern consistency classification**. In this stage (element 6 in Fig. 1), the system predicts the likelihood that an observed feature is *typical* (follows an established trend), *unusual* (deviates from an established trend), or *neither* (either a trend has not been established, or the observation cannot easily be classified as either typical or unusual).

For each feature, a multi-class SVM classifier was trained to output probabilities of these three classes. For example, if a customer came between 10am and 11am for several days in a row, but at 6pm on another day, the pattern consistency classifier for the "time of day" feature would be expected to classify the visit as "unusual".

To preserve as much information as possible, each classifier outputs not only a classification result, but also the raw likelihood scores for each class prediction, enabling distinctions to be made between strong and weak judgments of novelty or sameness. This is because the decision process for selecting topics is unknown. It is possible that some topics are consistently selected over others, but it is also possible that the degree of novelty also affects this selection.

2.6 Topic Selection

The second stage is **topic selection** (element 7 in Fig. 1). Given the outputs of all the pattern consistency classifiers as well as the feature data, an SVM classifier is trained to identify which topic is appropriate for the robot to speak about.

For example, if the "time of day" pattern consistency classifier predicted likelihoods of (50%, 25%, 25%) for "typical" "unusual", and "neither", while the classifier for "group size" predicted likelihoods of (5%, 85%, 10%), then it seems possible that it would be more appropriate for the robot to comment on the unusual group size – perhaps the customer brought her child shopping, although she usually comes alone.

However, topic selection may depend on more than just the degree of novelty/sameness of the features. For example, some topics, such as group size, might be fundamentally more interesting to comment on than others, such as walking speed. Since we do not possess an analytical model of how people select topics based on this combination of variables, we trained an SVM to reproduce the coder's preferences.

2.7 Utterance Generation

The final stage is **utterance generation** (element 8 in Fig. 1). This stage uses both the feature data of the selected feature and the result of its pattern consistency classification to create a specific utterance. This stage was not trained through machine learning, but rather hand-coded to generate appropriate utterances for each topic. For example, if the selected topic is "day of week", and the current visit has a value of "Tuesday", which is classified as "unusual", the generator would output the utterance "Hi, it's good to see you again. It's unusual to see you here on a Tuesday".

Note that a feature does not need to be novel to generate an utterance – it is possible to comment on consistency as well. For example, if the selected topic is "time of day", and the current visit has a value of "10:20am", which is classified as "typical", the generator would output the utterance, "Hi, it's good to see you again. You always come in the morning, don't you?"

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

TABLE 3. UTTERANCE TEMPLATES FOR EACH FEATURE.

Feature	Condition	Utterance templates
Time of day	Unusual, $t_{visit} - \overline{t_{visit}} > 3\ hours$	It's unusual to see you here [right after opening, in the morning, around noon, in the early afternoon, in the evening, so close to closing time]
	Unusual, $t_{visit} - \overline{t_{visit}} < 3\ hours$	Today you came much [earlier, later] than usual.
	Typical	You always come [in the morning, around noon, in the afternoon, in the evening], don't you?
Day of week	Unusual	It's unusual to see you here on a [Monday, Tuesday, Wednesday, Thursday, Friday].
	Typical	You often come on [Mondays, Tuesdays, Wednesdays, Thursdays, Fridays], don't you?
Days since last visit	Unusual, $dslv < \overline{dslv}$	Long time, no see! I was lonely because I didn't see you for a while.
	Unusual, $dslv > \overline{dslv}$	I'm surprised to see you so soon. Usually I don't see you so often.
	Typical, $dslv = 1$	I'm happy to see you again. You come every day, don't you?
	Typical, $dslv > 1$	I noticed that you come here every [$dslv$] days, so I was looking forward to seeing you today!
Same day visit count	Unusual, $sdvc = 2$	Hey, I just talked to you, didn't I? It's nice to see you twice on the same day.
	Unusual, $sdvc > 2$	Wow, you've come to see me many times today. I'm flattered that you like me so much!
Walking speed	Unusual $v_{walk} - \overline{v_{walk}} > 0$	You're walking very quickly today. Is it because you wanted to meet me? I'm happy!
	Unusual $v_{walk} - \overline{v_{walk}} < 0$	Are you a little tired today? You seem to be walking more slowly than usual.
Group size	Unusual, $n_{group} = 1$	I see you came shopping by yourself today.
	Unusual, $n_{group} > 1$	I'm happy that [n_{group}] of you came to see me today.
	Typical, $n_{group} = 1$	I always see you shopping by yourself. Your bags must be heavy.
	Typical, $n_{group} = 2$	I always see the two of you shopping together. It seems like fun!
No topic		I remember you! Thanks for coming back to see me again.

The complete set of utterance templates is listed in Table 3. Some situations in which it would not be interesting to make a comment are not covered, such as when a person's walking speed is always the same.

3 TRAINING AND CLASSIFIER EVALUATION

3.1 Data Collection

To collect data for training the system, we recorded video from 10am to 6pm, Monday through Friday, for 23 days during the months of June and July. From this period, a subset of three days were arbitrarily chosen, and every face observed by either camera during each of those days was registered in the face detection database with an anonymous ID number. Shopping mall staff and experimenters were excluded from registration, resulting in 1786 faces registered in the system.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

Next, we identified when these 1786 individuals visited the shopping mall throughout the 23-day period. All of the video data from the 23 days was run through the face detection software to generate a time series of faces detected at each second of video for each camera. Whenever a consistent ID was observed for more than 5 frames within a 60-second interval, it was counted as a visit by that person. Subsequent detections of that individual within 10 minutes were considered to be a part of the same visit. A total of 5002 visits by registered individuals were detected in the total data set. The distribution of visits per person approximately followed a power-law distribution. The mean number of visits was 3.6, with a standard deviation of 4.2. 57% of visitors came only 1 or 2 times, and 82% of visitors came 5 or fewer times.

Since features require a minimum of three visits in order to estimate pattern consistency (two to establish a pattern and one to compare with the pattern), we considered only the third and subsequent visits from each visitor in the process of classifier training. Although the first two visits for each visitor were not annotated, they were included in the history database, and thus they were used for generating the feature vectors for the subsequent annotated visits. This final dataset contained a total of 755 visits.

3.2 Classifier Training

Training data for the system was then created by a human coder (element 9 in Fig. 1). There were two steps to this process: pattern consistency classification and topic selection. Coding was performed for each of the 755 visits in the dataset.

Pattern consistency classification: For each visit in the training set, the coder was presented with one feature at a time. She was shown the most recent value for that feature alongside the values from the preceding sequence of visits in that person's history, and she was asked to classify each feature as "typical", "unusual", or "neither", according to the definitions presented in Sec. 2.5.

Topic selection: After all of the individual features were classified, the data were input to the utterance generator and an utterance was generated for each possible topic (7 topics were presented, one for each of the six features, and one "none" topic which produced a generic greeting). The coder then chose which topic was most appropriate for that situation.

Table 4 shows an example of data a coder might see in the training phase. The rows in the table show information for a customer's current (fourth) visit and all previous visits (three, in this case). In this data, the customer's previous visits show no clear pattern regarding "time of day," so the coder classifies it as "neither" – that is, as there is no established pattern, the current visit cannot be said to be either typical or unusual. Next, the "day of week" and "days since last visit" columns show that the current visit clearly deviates from a well-established pattern, so the coder marks those features as "unusual". The remaining columns show a clear pattern in the visit history and the current visit also agrees with that pattern, so the coder marks them as "typical".

TABLE 4. EXAMPLE OF INPUTS AND CODING RESULTS FOR TRAINING PHASE

	Time of day	**Day of week**	**Days since last visit (weekdays)**	**Same Day Visit Count (# visits)**	**Walking Speed (m/s)**	**Group Size (# people)**
Visit N-3	10:20	Thursday	5	1	1.10	1
Visit N-2	15:45	Thursday	5	1	1.08	1
Visit N-1	13:10	Thursday	5	1	0.96	1
Current Visit	11:03	Wednesday	4	1	1.02	1
Pattern Consistency Judgment	Neither	Unusual	Unusual	Typical	Typical	Typical

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

The coder must then select a topic which would be appropriate to comment on for the current visit. The system generates a sentence for each of the features based on whether they were coded as "typical", "unusual", or "neither". In this case, the coder selects the sentence based on the "day of week" feature: "Hi, it's good to see you again. I don't usually see you here on Wednesdays." Finally, the coder's judgments of "typical", "unusual", or "neither" for each of the six features, and the coder's chosen topic (day of week) are recorded in a database for use as training data for the classifiers.

The ratings for training our classifier were generated by a single coder. To confirm the generalizability of the coder's decisions, we had a second person perform coding with 10% of the data set, and we calculated Cohen's kappa to evaluate interrater reliability. Kappa scores were 0.49 for "time of day", 0.55 for "day of week", 0.54 for "days since last visit", 0.92 for "same day visit count", 0.46 for "walking speed", 0.82 for "group size", and 0.68 for topic selection. We interpret these scores to indicate moderate to substantial agreement between the raters, especially in the final stage of topic selection.

3.3 Classifier performance

The classifiers were trained and then evaluated using a 10-fold cross-validation with the training data. The implementation of these classifiers used the Java version of LibSVM [24]. C_SVC regularized support vector classification was used, with a radial basis function (RBF) kernel. Cross-validation accuracy results for each of the classifiers are presented in Table 5.

TABLE 5. CROSS-VALIDATION ACCURACY OF PATTERN CONSISTENCY CLASSIFICATION AND TOPIC SELECTION.

Feature	**Type**	**Prediction accuracy**
Time of day	Continuous	75.4% (569/755)
Day of week	Categorical	94.3% (712/755)
Days since last visit	Integer	79.3% (599/755)
Same day visit count	Integer	98.5% (744/755)
Walking speed	Continuous	83.0% (627/755)
Group size	Integer	98.9% (747/755)
Selected Topic		95.1% (718/755)

In these results, predictions for "same day visit count" and "group size" were nearly 100% accurate, which is to be expected given that each value was nearly always 1, which can easily be categorized as "typical". At times when either feature had a value of 2 or larger, it would almost certainly result in a categorization of "unusual". For continuous values such as time of day or walking speed, it is less clear what constitutes "typical" vs. "unusual" behavior, and prediction accuracy was correspondingly lower for these features. However, prediction accuracy was still quite high despite this greater ambiguity, demonstrating that the system was successfully able to reproduce the judgment of the human.

To illustrate the distribution of prediction results in more detail, the confusion matrices are shown in Table 6. Overall, prediction accuracy was reasonably high, and we were satisfied with these results.

4 FIELD DEMONSTRATION

To showcase the ability of the system to perform in real situations, we conducted a field trial in which a robot greeted customers at a shopping mall using the trained classifiers to generate personalized utterances. In this trial, the actual visit data from the first data collection was not used, as it would be socially inconsistent behavior for a robot to "remember" a person's activities when the robot was not there. Thus, for this experiment we created a new visit database, registering only customers who directly participated in conversational interactions with the robot.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

TABLE 6. CONFUSION MATRICES FOR PATTERN CONSISTENCY CLASSIFICATION AND TOPIC SELECTION

Time of day

		Predicted		
		Typical	Neither	Unusual
Actual	Typical	65	93	3
	Neither	15	453	19
	Unusual	6	50	51

Day of week

		Predicted		
		Typical	Neither	Unusual
Actual	Typical	9	35	0
	Neither	2	703	0
	Unusual	0	6	0

Days since last visit

		Predicted		
		Typical	Neither	Unusual
Actual	Typical	149	12	0
	Neither	21	338	35
	Unusual	2	86	112

Same day visit count

		Predicted		
		Typical	Neither	Unusual
Actual	Typical	202	0	0
	Neither	11	312	0
	Unusual	0	0	230

Walking speed

		Predicted		
		Typical	Neither	Unusual
Actual	Typical	0	73	0
	Neither	0	625	3
	Unusual	0	52	2

Group size

		Predicted		
		Typical	Neither	Unusual
Actual	Typical	318	0	0
	Neither	0	385	8
	Unusual	0	0	44

Topic

		Predicted						
		Time of day	Day of week	Days since last visit	Same day visit count	Walking Speed	Group Size	No topic
Actual	Time of day	129	0	18	3	0	0	0
	Day of week	2	6	0	0	0	0	0
	Days since last visit	4	0	196	0	0	0	1
	Same day visit count	0	0	0	226	0	0	0
	Walking speed	0	0	0	0	6	1	2
	Group size	0	0	0	0	0	14	1
	No Topic	0	0	2	1	0	2	141

4.1 Procedure

The robot was deployed in the same location each day from 10am to 5pm for a period of 14 weekdays. During this time, one operator monitored the robot from a control room while an assistant stood near the robot for safety.

Customers interacted with the robot, and the operator controlled the conversation using a scripted sequence of utterances. Our focus in this study was on the greetings only, so the interaction was kept very short. In the interactions, the robot explained that it was studying to be a greeter for a shop, so it was practicing to remember people and give polite greetings. The operator actuated the utterance timings and handled off-topic questions from the customers such as "can we take a picture together?"

A few changes were made to the system configuration for the field trial. The robot's on-board cameras were used instead of the ceiling-mounted cameras, because the interaction history for the field trial was based on actual interactions with the robot. For example, it would be strange for the robot to remember somebody's actions based on ceiling camera data on a day when the robot was not deployed.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

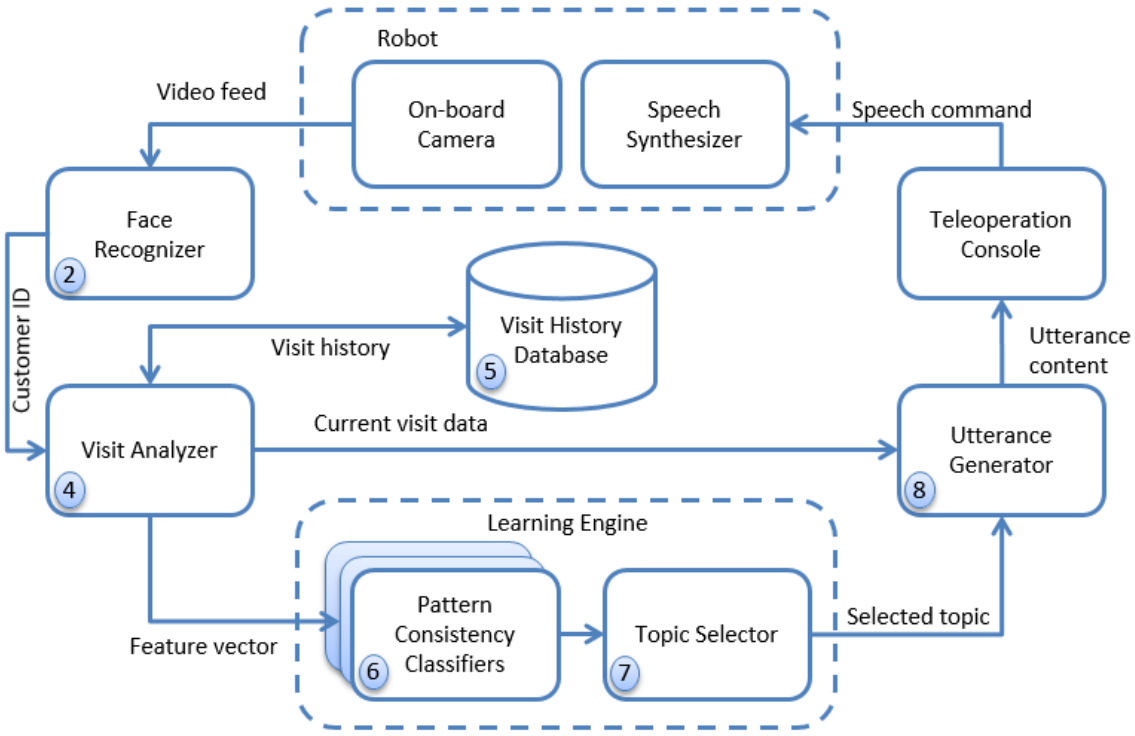

Figure 4. Architecture of the greeting system used in the field trial. Numbered elements correspond to elements presented in Fig. 1. Elements 1, 3, and 9 are not present here because the tracking system, sensor fusion, and training interface were not used in the field trial

Also, the LRF-based tracking system was not used for the robot experiment. This was a practical decision, which we considered to be justifiable because the two features based on the LRF system, group size and walking speed, had been found to be of only marginal utility, as they comprised only 3.2% of utterance topics coded in the original dataset. Removing the LRF system eliminated robot localization as a possible source of error in data association, and it also greatly reduced the amount of effort required to set up and maintain the robot system.

4.2 System

In this trial, we used a Robovie II communication robot placed in a stationary position near an entrance to the shopping mall. The system configuration for the field demonstration is shown in Fig. 4.

Starting from the top of the diagram, the data flow proceeds as follows. First, a video feed from one of the **robot**'s cameras is sent to a face recognition system using Okao Vision software. If the **face recognizer** (2) identifies the customer, it sends the customer's ID to the **visit analyzer** (4). The visit analyzer generates a new visit for that customer and retrieves historical visit data for that customer from the **visit history database** (5). In this trial, only the following features were used: **day of week**, **days since last visit**, **time of day**, and **same day visit count**.

Using the customer's current and historical visit data, the visit analyzer generates an input vector for the learning engine, including statistical descriptors as described in Sec. 2.4.4. The **pattern consistency classifiers** (6) predict "typical", "unusual", or "neither" values for each of the features in the input vector, and the **topic selector** (7) predicts the optimal topic for the robot to speak about. This selected topic is sent to the **utterance generator** (8), which uses data from the current visit to generate an utterance for the robot. This utterance content is sent to a **teleoperation console** used by an operator to supervise and control the robot, and when the operator presses a button, the speech command is sent to the robot.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

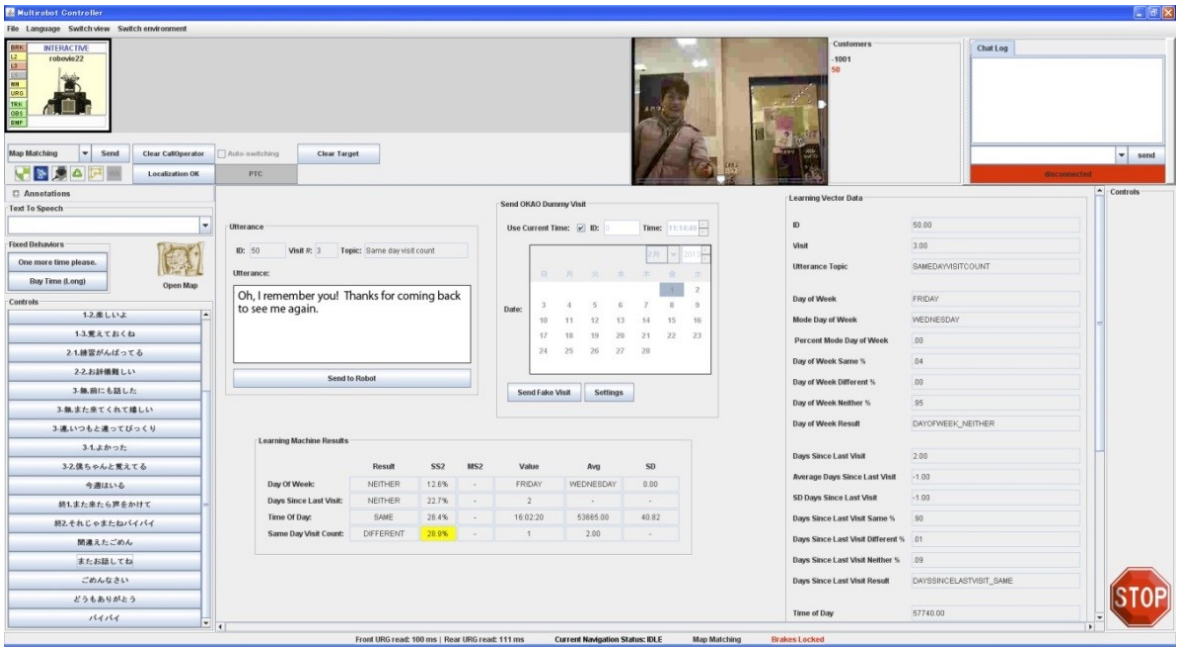

Figure 5. Graphical interface used by the operator. Video is shown on the upper center panel, buttons to trigger general utterances are presented on the left, and the panel in the center displays the greeting utterance generated by the system, which is sent to the robot when the operator clicks the button below it. The remaining displays show internal variables for the prediction system, which were only used for testing and debugging

The operator was placed in a separate room to monitor the robot and trigger its utterances, using the teleoperation interface shown in Fig. 5. The utterances of the robot were scripted for the first two interactions, and the interface showed a list of buttons corresponding to each of the phrases in the script. Beginning with the third interaction, a script was used which included a placeholder for the greeting phrase generated by the system. The utterances in each script were triggered by the operator in order to ensure proper timing of the interaction. Additional utterance buttons were also provided for the operator to use in case the customers had off-topic questions, such as "May I take a picture with you?" and text-to-speech capability was also provided to handle unanticipated situations. When the robot was unable to respond quickly to something a person said, automated conversation fillers were used to buy time for the robot to respond [25].

4.3 Results

The purpose of this field deployment was to demonstrate that our system could be put to practical use in the field. The actual performance of the system is difficult to measure, as the appropriateness of the robot's utterances depends heavily on the variability of behavior exhibited by the individual customers. If a customer exhibits none of the behavior patterns for which the system is designed, then the system's performance cannot be evaluated. Thus, it was impractical to conduct a controlled comparison experiment using real customers. Instead, we will present some examples of visitors' interactions with the robot. We also interviewed customers to obtain their qualitative impressions of the robot.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

First interaction with a boy and his family.

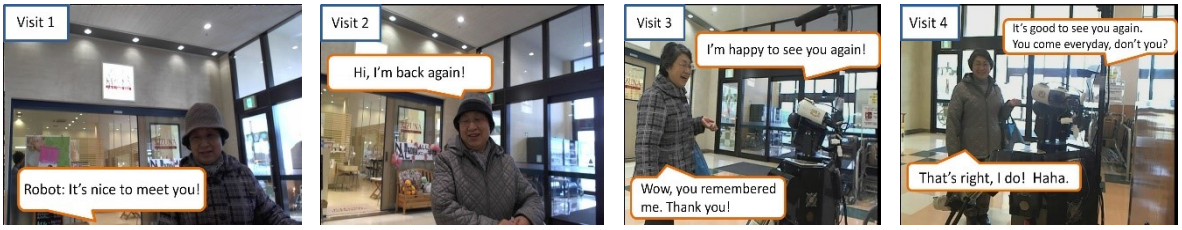

Figure 7. Four interactions with a woman shopping alone. On visit 3, the robot comments on "no topic", and on visit 4, the robot comments on "days since last visit"

4.3.1 Overall statistics

Customers interacted with the robot a total of 368 times during this study. Within these interactions, 54 customers interacted more than once, and 8 interacted with the robot three or more times. As the robot's utterances were scripted for the first two interactions, only the third and subsequent visits were of interest.

Of the 8 customers who came 3 or more times, three were adult women, two were elderly men, two were young boys, and one was a mother accompanied by a young child. A total of 17 interactions occurred in the third visit or later.

4.3.2 Example interactions

Some examples of the robot's interactions are shown in Figs. 6-8. Fig. 6 shows a first-time interaction with a boy and his family. Since there is no visit history yet, the robot cannot generate a personal greeting, but it tells the visitor that it will try to remember him the next time he comes.

Fig. 7 shows a woman who visits several times. On the third visit, no topic was selected for a personal greeting, so the robot simply said that it remembered her and was happy to see her again. On the fourth visit, the topic "days since last visit" was selected, and the robot commented that it noticed she came every day. The woman laughed and said that the robot was correct.

Figure 8. Scenes of a mother and her son visiting the robot. The center frame shows her reaction when the robot said that it noticed she came every three days

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

Fig. 8 shows a woman visiting the shopping mall with her son. The little boy was shy and did not talk to the robot, but the mother did. The transcript, translated from Japanese, follows (C represents the customer, and R represents the robot.)

C: Hello.
R: Hello. Welcome back to the mall. I'm happy that you came back to talk to me!
C: (Laughs)
R: I noticed that you come here every three days, so I was looking forward to seeing you today!
C: (To son) Wow, that's amazing, isn't it!
R: I've been talking with a lot of people to practice my greetings. I wonder if I'm improving?
C: (To son) It has, don't you think? (To robot) Yes, you have!
R: This is fun. I think I'll really be ready to work here soon. Please come see me sometime.
C: (To robot) Ok! (To son) The robot wants us to come visit again!
R: Thanks, I'll see you next time.
C: See you later!
R: Bye bye
C: (To robot) Bye bye. (To son) Isn't that cool? It really remembered us!

4.3.3 Details of interactions

As Table 7 shows, the robot system was able to conduct 11 successful interactions based on personal interaction history. In these interactions, three topics were used to generate utterances as follows. Note that these phrases are translated from Japanese, so some of the nuances may be lost in translation.

TABLE 7. SUMMARY OF INTERACTIONS ON THIRD VISIT OR LATER.

Selected topic	**Number of visits**	
	Typical	Unusual
Time of day	4	0
Days since last visit	4	0
Same day visit count	0	3
No topic	2	

Error condition	**Number of occurrences**
Face recognition failure	2
Other system errors	2

In the 4 cases where "Days since last visit (typical)" was selected, the person had come regularly to speak to the robot each day. In these cases, the robot said, "You come to shop here every day, don't you? It's always good to see you."

"Time of day (typical)" was used for four visits, twice for customers who regularly came around mid-day, and twice for customers who came in the late afternoon. In these cases, the robot said "You always come in the (morning/afternoon), don't you?"

"Same day visit count (unusual)" was used when a customer came twice in the same day, when they had previously always come on different days. In this case, the robot said, "Hey, I just spoke to you, didn't I? I'm happy to see you again!"

For two of the visits, "No topic" was selected. In these cases, the robot simply told the person it remembered them and was happy to see them again.

Four of the 17 visits encountered technical failures. Face recognition failed 2 times. Facial recognition failures were often due to the customer wearing a mask (which is typical in Japan if someone has a cold) or to variations in lighting or clothing, *e.g.* wearing a hat. In the remaining two failure cases, problems with connectivity to the visit history database over the wireless network prevented the system from predicting an utterance.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

PARTICIPANTS' RESPONSES TO INTERVIEW QUESTIONS.

	Did you feel familiarity ("shitashimi") with the robot?	**Did the interaction feel natural?**	**If you had the opportunity, would you like to interact with the robot again?**
Interview 1	Yes, because the robot reacts to the fact that I have come back several times.	No, because the robot is slow, and I need to adjust my timing to it.	Yes, I felt familiarity with the robot and I am interested in the robot.
Interview 2	Yes, because my child always looks forward to meeting the robot.	Yes, I am happy that the robot remembered me.	Yes.
Interview 3	Yes, I feel familiarity with it because it remembers me. Also, its nodding behavior is cute.	Yes. It is amazing that every time I meet the robot it has something new to say.	Yes.
Interview 4	Yes, I have become more familiar with the robot over time and I have learned how the robot reacts. I feel familiar with the robot because I have met the robot many times.	Yes. After interacting with the robot several times, talking with the robot has come to feel natural.	Yes, I often came to the shopping mall, and if the robot is at the entrance, I would like to meet the robot and greet it. I'm not young, and it's difficult for me to remember people's faces. I'm impressed with how well the robot does it.

In these failure cases, the operator, who remembered the faces of the frequent customers, manually set the selection to "no topic" so that the interaction could proceed. Afterwards, she logged and manually corrected the recognition errors in the visit history database. Among the entire set of 368 interactions, she logged 12 detection failures due to people wearing masks and/or sunglasses, and 55 other instances where no face was detected in the video image. Among the cases when a face was detected, there were 10 instances where a person was not correctly re-recognized. The operator estimated that two of these were related to wearing a mask or glasses.

4.3.4 Interviews

We asked each customer who came at least 3 times to provide an interview about their experience. As customers were not paid participants, interviews were conducted on a volunteer basis only, and we were able to obtain 4 interviews. In the interviews we asked three questions, summarized in Table 8.

The first question was whether they felt "shitashimi" with the robot. "Shitashimi" is a Japanese term which describes a feeling of intimacy, affection, and familiarity. All participants said "yes."

Next, participants were asked if the robot's interaction was natural. Three responses were "yes," and one participant responded "no," because she said the timing of the robot's speech was too slow for natural conversation. We think this could have been due to network and system delays or a slow reaction by the teleoperator. Finally, we asked if they would like to talk with the robot again, given the chance. All respondents said "yes."

Although factors like the novelty of the robot and the timing of its speech were mentioned by participants, every participant positively commented about the robot's ability to recognize and remember them as individuals. Since the robot did not call anyone by name, the only way this individual recognition was conveyed was through use of our personal greeting technique. This provides qualitative evidence that the function provided by our system was appreciated by the customers.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

5 DISCUSSION AND CONCLUSIONS

5.1 Discussion

5.1.1 Number of participants

It was unfortunate that the system could only be evaluated through 11 interactions with 8 participants in the field trial. A likely factor contributing to this difficulty in gathering data is that the robot was not providing any particular service to the customers, and its conversational capability was quite limited and simple. In future studies it will be important to consider creating a more interesting or more useful service for the robot to provide, in order to enable the collection of more data.

Conversely, the fact that so many people returned to talk with the robot again, even though it provided no useful information or service, could be seen as a positive argument for the effectiveness of using techniques for designing human-robot interactions based on memory or history. Visitors found it interesting that the robot could remember them, and many came back to talk with it again, even though the interaction served no practical purpose.

5.1.2 Possible features

Despite the apparent simplicity of the features used in the field study (all were based on time of arrival), customers were still surprised and impressed that the robot remembered them and noticed their behavior, and our data collection demonstrated that features reflecting sensor data such as walking speed could also be used. However, the addition of more features would give the interactions more variety and probably increase the feeling that the robot's utterances are truly personalized. As recognition technologies improve, it would be interesting to use detections of more sophisticated features such as clothing or hairstyle with the technique presented here.

5.1.3 Supervised vs. Unsupervised Learning

It might seem that the pattern consistency classification task in this study could have been achieved using an unsupervised learning technique, such as those used in some anomaly detection systems. We would argue that since the task is to reproduce human judgments, it is fundamentally necessary to use some kind of human judgment, obtained by coding or by observation of human behavior, as a basis for training the system. In particular, it is important to consider that human judgments of social appropriateness should not necessarily be expected to agree with statistical measurements of consistency. For example, some observations of time could be statistically classified as outliers but below the threshold of human detection (if a customer came at exactly 2:30pm each day, then came at 2:33pm one day, this probably would not be worthy of commenting on socially, and might not even be noticed, although it might statistically be an outlier). Furthermore, perception of features in different time ranges might be nonlinear due to social conventions – for example, a person might consider "before lunch" vs "after lunch" to be qualitatively very different, but perceive very little difference between "early afternoon" and "mid-afternoon", even though mathematically the time differences are identical. For more complex features, such as observations of hairstyle and clothing, it would probably be even more difficult to statistically judge pattern consistency, and training based on human inputs would be even more important.

5.1.4 Use of SVM

Given the relative simplicity of the features used in this study, the use of SVM classifiers may seem like an excessively complex approach to a simple problem. While the classification tasks in this study could arguably have been achieved in simpler ways, our stance was to develop a generalizable system in anticipation of future sensing techniques, which may provide data with higher dimensionality and greater complexity. The SVM classifiers we used were effective for the simple inputs we provided, while leaving open the future possibility of using more complex kinds of input data.

One could also argue that some form of supervised anomaly detection system would be more appropriate as the underlying classifier, rather than SVM. In this case, the question is which approach makes the most sense for the classification problem at hand.

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

In our scenario, we obtain a set of "SAME", "DIFFERENT", or "NEITHER" labels from a human coder, with the goal of reproducing the judgment leading to those labels. This fits the format of a classic multi-label classification problem, in which case a technique like SVM would be a typical solution.

Furthermore, although the "DIFFERENT" label in our study does seem to correspond conceptually to some level of anomaly detection, the "SAME" and "NEITHER" labels do not have such analogies. A rating of "NEITHER" could mean that no pattern was discernable at all, or it could mean that although a clear pattern is discernable, the most recent observation may not clearly conform to or diverge from that pattern. It could even mean that, although the observed behavior clearly conforms to or differs from a pattern, it is not the sort of pattern that is worthy of comment.

For these reasons we believe that multi-label classification seems more appropriate for this type of problem than anomaly detection.

5.1.5 Scalability of application

The contribution of this technique goes beyond simply automating a manual operation task. As this technique is data-driven, it is equally effective regardless of the number of customers. Remembering details about hundreds or thousands of customers is a task that is truly beyond the normal capability of a human teleoperator. Thus, this technique is not merely a matter of offloading a manual task to an automated system; it is a way in which automation can transcend the limitations of a human teleoperator to enable a task which would previously have been impossible.

5.1.6 Scalability of training

The utility of the proposed technique should be considered in conjunction with the cost of training the classifiers. Human coders will need to label each feature in the training data set, and this scales linearly with the number of features and the number of examples in the data set, potentially requiring a significant amount of coding effort.

However, the amount of manual coding needed for our proposed technique is dictated only by the complexity of the model of human judgment. It is not necessary to code the entire data set, but just enough variety of samples to provide an accurate model of the Same/Neither/Different judgment patterns for each feature. In our study, we coded only a subset of three days of data, which could probably be reduced even further if necessary. With careful selection of data and optimization of the process, it should be possible to reduce the amount of annotation substantially while still providing the data necessary to train a model consistent with the judgments of multiple independent coders.

The first-stage classifiers trained to detect novelty or sameness can only detect novel or typical situations that are represented in the training data set, which is a limitation on the system's ability to detect particularly rare situations. For most of the features used in this study, a good variety of "typical", "unusual", and "neither" examples were present in the data set that was coded, but ensuring sufficient coverage of training examples is still an important consideration in the process of choosing the data used for coding.

5.1.7 Other limitations

The height assumption of 160 cm used for fusion between the face detection system and the pedestrian tracking system can be a source of error in sensor fusion if people are much taller or shorter than the assumed height, although this can be moderated to some extent by adjusting error thresholds for the nearest-neighbor matching. The lack of height information is one limitation of a 2D LRF-based tracking system. Some pedestrian tracking systems have used other techniques, such as 3D depth sensors, eliminating the problem of this height assumption [26].

5.2 Conclusion

In this study we have demonstrated a working example of how data-driven machine learning techniques can be used to enhance social human-robot interactions by helping to personalize interaction contents.

The originality of this work lies in its modeling of human perception of when a feature is "novel enough" or "typical enough" to merit comment, and in its application as a tool for enhancing and personalizing social human-

This is the authors' preprint version of the accepted manuscript.
The final publication is available at Springer via http://dx.doi.org/10.1007/s12369-016-0385-4

robot interactions which would otherwise seem highly impersonal. This technique represents one way to incorporate history into interaction to give a person the impression that the robot remembers him/her and create a personalized feel to an interaction.

This technique for personalization is not based simply on static or attribution data such as a person's name, age, or gender, but rather it is a reflection of actual shared experience between the robot and the human. By using a pattern consistency classifier and utterance generator trained to reproduce the judgment and utterances of humans, the robot is able to produce humanlike greetings in response to these real shared experiences, creating a feeling of personal familiarity with the robot's conversation partner.

Compliance with Ethical Standards

This research was conducted in compliance with the standards and regulations of our company's ethical review board, which requires every experiment we conduct to be subject to a review and approval procedure according to strict ethical guidelines.

Conflict of interest: The authors declare that they have no conflicts of interest.

Acknowledgment

Funding: This research was supported by the Ministry of Internal Affairs and Communications of Japan.

We would like to thank Dr. Satoshi Koizumi, Tony Han, Benoit Toulmé, Peace Cho, and the management of the APiTA Town Keihanna shopping mall for their help in the organization and execution of the data collection.

References

[1] 1. Friedman B, Kahn PH, Hagman J (2003) Hardware Companions?-What Online AIBO Discussion Forums Reveal about the Human-Robotic Relationship. Paper presented at the ACM Conference on Human Factors in Computing Systems (CHI2003),

[2] 2. Kahn PH, Kanda T, Ishiguro H, Freier NG, Severson RL, Gill BT, Ruckert JH, Shen S (2012) "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. Developmental psychology 48 (2):303

[3] 3. Wada K, Shibata T (2007) Living With Seal Robots-Its Sociopsychological and Physiological Influences on the Elderly at a Care House. IEEE Transactions on Robotics 23 (5):972-980

[4] 4. Sabelli AM, Kanda T, Hagita N (2011) A Conversational Robot in an Elderly Care Center: an Ethnographic Study. Paper presented at the ACM/IEEE int. Conf. on Human-Robot Interaction (HRI2011),

[5] 5. Eyssel F, Reich N Loneliness makes the heart grow fonder (of robots)—On the effects of loneliness on psychological anthropomorphism. In: Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on, 2013. IEEE, pp 121-122

[6] 6. Bickmore TW, Picard RW (2005) Establishing and Maintaining Long-Term Human-Computer Relationships. ACM Transactions on Computer-Human Interaction (TOCHI) 12 (2):293-327

[7] 7. Gockley R, Bruce A, Forlizzi J, Michalowski M, Mundell A, Rosenthal S, Sellner B, Simmons R, Snipes K, Schultz AC, Jue W Designing robots for long-term social interaction. In: Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on, 2-6 Aug. 2005 2005. gockley05b, pp 1338-1343. doi:10.1109/iros.2005.1545303

[8] 8. Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2012) Long-Term Interactions with Empathic Robots: Evaluating Perceived Support in Children. Social Robotics:298-307

[9] 9. Kidd CD (2008) Designing for long-term human-robot interaction and application to weight loss. Massachusetts Institute of Technology,

[10] 10. Pan Y, Okada H, Uchiyama T, Suzuki K (2015) On the Reaction to Robot's Speech in a Hotel Public Space. Int J of Soc Robotics 7 (5):911-920. doi:10.1007/s12369-015-0320-0

[11] 11. Riek LD, Paul PC, Robinson P (2009) When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. Journal on Multimodal User Interfaces 3 (1-2):99-108

[12] 12. Sakamoto D, Kanda T, Ono T, Kamashima M, Imai M, Ishiguro H (2004) Cooperative embodied communication emerged by interactive humanoid robot. International Journal of Human-Computer Studies:247-265

[13] 13. Sabelli AM, Kanda T, Hagita N A conversational robot in an elderly care center: An ethnographic study. In: Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on, 8-11 March 2011 2011. sabelli11, pp 37-44

[14] 14. Kanda T, Sato R, Saiwaki N, Ishiguro H (2007) A Two-Month Field Trial in an Elementary School for Long-Term Human-Robot Interaction. Robotics, IEEE Transactions on 23 (5):962-971. doi:10.1109/tro.2007.904904

[15] 15. Kanda T, Shiomi M, Miyashita Z, Ishiguro H, Hagita N (2010) A communication robot in a shopping mall. Trans Rob 26 (5):897-913. doi:10.1109/tro.2010.2062550

[16] 16. Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. ACM Comput Surv 41 (3):1-58. doi:10.1145/1541880.1541882

[17] 17. Huber PJ (1981) Robust statistics, vol 1. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York

[18] 18. Marsland S, Nehmzow U, Shapiro J (2005) On-line novelty detection for autonomous mobile robots. Robotics and Autonomous Systems 51 (2):191-206

[19] 19. Andry P, Gaussier P, Moga S, Banquet JP, Nadel J (2001) Learning and communication via imitation: an autonomous robot perspective. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 31 (5):431-442. doi:10.1109/3468.952717

[20] 20. Bonaccorsi M, Fiorini L, Cavallo F, Saffiotti A, Dario P (2016) A Cloud Robotics Solution to Improve Social Assistive Robots for Active and Healthy Aging. Int J of Soc Robotics 8 (3):393-408. doi:10.1007/s12369-016-0351-1

[21] 21. Glas DF, Miyashita T, Ishiguro H, Hagita N (2009) Laser-Based Tracking of Human Position and Orientation Using Parametric Shape Modeling. Advanced Robotics 23 (4):405-428. doi:10.1163/156855309x408754

[22] 22. Lao S, Kawade M (2005) Vision-based face understanding technologies and their applications. In: Advances in Biometric Person Authentication. Springer, pp 339-348

[23] 23. Yücel Z, Zanlungo F, Ikeda T, Miyashita T, Hagita N (2013) Deciphering the Crowd: Modeling and Identification of Pedestrian Group Motion. Sensors 13 (1):875-897

[24] 24. Chang CC, Lin CJ (2001) LIBSVM: A library for support vector machines.

[25] 25. Shiwa T, Kanda T, Imai M, Ishiguro H, Hagita N How quickly should communication robots respond? In: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, Amsterdam, The Netherlands, 2008. ACM, shiwa08, pp 153-160. doi:10.1145/1349822.1349843

[26] 26. Brščić D, Kanda T, Ikeda T, Miyashita T (2013) Person Tracking in Large Public Spaces Using 3-D Range Sensors. Human-Machine Systems, IEEE Transactions on 43 (6):522-534. doi:10.1109/thms.2013.2283945

Dylan F. Glas received his Ph.D. in Robotics from Osaka University in 2013. He received his M.Eng in Aerospace Engineering from MIT in 2000 and S.B. degrees in Aerospace Engineering and in Earth, Atmospheric, and Planetary Sciences from MIT in 1997. From 1998-2000 he worked in the Tangible Media Group at the MIT Media Lab. He is currently at ATR in Kyoto, Japan (2005-), where he was group leader of the Department of Cloud Intelligence and is currently a Senior Researcher in Hiroshi Ishiguro Laboratories, under the JST ERATO Ishiguro Symbiotic Human-Robot Interaction Project. He is also a Guest Associate Professor at the Intelligent Robotics Laboratory at Osaka University. His research interests include social human-robot interaction design, cloud network robot systems, ubiquitous sensing, teleoperation for social robots, and machine learning techniques for reproducing human behavior.

Kanae Wada received her master's degree in engineering from Osaka University in 2012 and is currently working at 3D MEDiA Co., Ltd. From 2010-2012, she was an intern researcher at the Intelligent Robotics and Communication Laboratories (IRC) at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan. Her research interests include networked robots, human-robot interaction, and field experimentation.

Masahiro Shiomi received M. Eng. and Ph.D. degrees in engineering from Osaka University in 2004 and 2007. From 2004 to 2007, he was an intern researcher at the Intelligent Robotics and Communication Laboratories (IRC). He is currently a group leader in the Agent Interaction Design department at IRC, Advanced Telecommunications Research Institute International (ATR). His research interests include human-robot interaction, robotics for child-care, networked robots, and field trials.

Takayuki Kanda received the B. Eng., M. Eng., and Ph.D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively. From 2000 to 2003, he was an Intern Researcher with the Advanced Telecommunications Research Institute International (ATR) Media Information Science

Laboratories, Kyoto. He is currently a Senior Researcher at ATR Intelligent Robotics and Communication Laboratories. His research interests include intelligent robotics, human-robot interaction, and vision-based mobile robots. Dr. Kanda is a Member of the Association for Computing Machinery, the Robotics Society of Japan, the Information Processing Society of Japan, and the Japanese Society for Artificial Intelligence.

Hiroshi Ishiguro received a D.Eng. in systems engineering from the Osaka University, Japan in 1991. He is currently professor of the Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University (2009-) and distinguished professor of Osaka University (2013-). He is also group leader (2002-) of Hiroshi Ishiguro Laboratories at the Advanced Telecommunications Research Institute and an ATR fellow. He was previously a research associate (1992–1994) in the Graduate School of Engineering Science at Osaka University and associate professor (1998–2000) in the Department of Social Informatics at Kyoto University. He was also visiting scholar (1998–1999) at the University of California, San Diego, USA. He was associate professor (2000–2001) and professor (2001–2002) in the Department of Computer and Communication Sciences at Wakayama University. He then moved to Department of Adaptive Machine Systems in the Graduate School of Engineering at Osaka University as a professor (2002–2009). His research interests include distributed sensor systems, interactive robotics, and android science.

Norihiro Hagita received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University in 1976, 1978, and 1986. In 1978, he joined Nippon Telegraph and Telephone Public Corporation (Now NTT). He was a visiting researcher in the Department of Psychology, University of California, Berkeley in 1989-90. He is currently Board Director of ATR, ATR Fellow, and Director of the Intelligent Robotics and Communication Laboratories. He is also a visiting professor at Nara Institute of Science and Technology and Osaka University. His major interests are cloud networked robotics, human-robot interaction, ambient intelligence, pattern recognition and learning, and data-mining technology. He is currently a research supervisor of the Japan Science and Technology Agency (JST).