Propuesta de Proyecto de Grado

Reconocimiento multimodal de estados afectivos para el robot social LRS2

Por

Br. Nerio Moran

Tutor: Prof. Jesús Pérez

Abril 2018

Reconocimiento multimodal de estados afectivos para el robot social LRS2

Br. Nerio Moran

Propuesta de Proyecto de Grado — Sistemas Computacionales, 8 páginas
Escuela de Ingeniería de Sistemas, Universidad de Los Andes, 2018

Resumen: En este trabajo se presenta una propuesta para el reconocimiento de emociones multimodal a partir de señales, específicamente expresiones faciales y el habla para los canales no-verbal y verbal respectivamente, en el contexto de interacción humano-robot. El problema planteado hace parte de los estudios actuales que buscan mejorar la comunicación humano-robot mediante la inclusión de diferentes señales que permitan determinar el estado afectivo de los humanos de una manera continua y precisa. El sistema propuesto aplica técnicas de aprendizaje de máquina para lograr la detección de emociones en una persona mientras interactúa con el robot LRS2.

Palabras clave: Interaccion Humano Robot, Aprendizaje de Máquina, emociones, Redes neuronales, Multimodal

Índice general

Introducción

Dentro del área de la interacción humano-robot, se ha generado una gran iniciativa de investigación en la detección de emociones, esto bajo la justificación de acortar la brecha de comunicación entre humanos y robots. Mejorar la capacidad de interacción mediante el reconocimiento del estado emocional del humano es de gran importancia para permitir que los robots puedan desempeñar tareas cooperativas y de una manera natural con los humanos.

Actualmente las áreas de aplicación para la interacción humano-robot son diversas y se mantienen en expansión. Alguna de ellas: robots sociales, entornos inteligentes, enseñanza, terapias y tratamientos, asistencia, entre otros.

La importancia que ha adquirido durante los últimos años la interacción entre humanos y robots es debido a la necesidad de desarrollar tareas de cooperación entre ellos, de tal forma que cualquier persona pueda interactuar de forma natural sin la necesidad de entender el robot.

La creación de sistemas lo suficiente robustos para permitir este tipo de interacción requieren de distintas señales de entrada independientes que se transformaran en información multimodal. El procesamiento de información proveniente de múltiples fuentes es un problema que se encuentra en diferentes áreas de investigación como la inteligencia artificial y la robótica.

Capítulo 1

Aspectos preliminares de la investigación

1.1. Antecedentes

Los antecedentes de esta investigación se dividen principalmente en dos categorías: primero, trabajos relacionados con el desarrollo de modelos de aprendizaje para distintos tipos de señales de entrada; segundo, investigaciones que utilizan más de 1 señal de entrada en sistemas de reconocimiento multimodal.

Faria et al., [2] realizó un articulo titulado "Affective Facial Expressions Recognition for Human-Robot Interaction" en el cual desarrollo un marco de trabajo para el reconocimiento de emociones en la interacción humano-robot. Para el marco de trabajo utilizo una base de datos "Karolinska Directed Emotional Faces (KDEF)", la cual utilizó para entrenar mediante aprendizaje supervisado un clasificador. El modelo se enfocó en detectar las 7 emociones universales, para esto utilizaron un modelo dinámico bayesiano mezclado, el cual es una especialización de las redes dinámicas bayesianas.

Razuri et al., [5] estudió las características cruciales del habla para el reconocimiento de emociones es su investigación titulada "Speech emotion recognition in emotional feedback for Human-Robot Interaction". En su investigación utilizó diferentes tipos clasificadores para probar cuales de ellos se acercaba a un sistema de tiempo real, la base de datos de entrenamiento fue eNTERFACE05. La investigación tuvo como

conclusión que los clasificadores: Maquina de vectores de soporte (SVM) y las redes bayesianas son buenos candidatos para los sistemas de tiempo real.

Por otro lado, Ménard et al., [3] utilizó como señales de entrada el ritmo cardiaco y la conductancia de la piel para determinar las emociones en su investigación titulada "Emotion Recognition Based on Heart Rate and Skin Conductance ". Ambas entradas fueron usadas en clasificadores distintos, crearon su propia base de datos utilizando sensores de respuesta biológica. Utilizaron como clasificador para ambas entradas una maquina de vectores de soporte (SVM) y el cual fue entrenado con los coeficientes de fourier de ambas entradas. Se determinó que ambas entradas fisiológicas sirven como fuente de información para determinar una emoción y a diferencia de otras entradas como la voz o la imagen digital estas señales tienen disponibilidad continua.

En los sistemas de reconocimiento multimodal Perez et al., [4] en su tesis de maestría titulada "Identificación de señales multimodales para reconocimiento de emociones en el contexto de interacción humano-robot ", realiza un sistema de reconocimiento bimodal cuyas entradas son imágenes y el habla. Para las imágenes utiliza una red neuronal convolucional y para el habla utiliza una maquina de vectores de soporte (SVM). Para la de fusión utiliza un sistema basado en decisiones utilizando el resultado de ambas entradas.

Castillo [1] en su investigación titulada "Emotion Detection and Regulation from Personal Assistant Robot in Smart Environmentrealiza un diseño para un robot social asistente el cual cuenta con sistema de reconocimiento bimodal basado en imágenes y el habla. Para detectar la emoción del usuario con el que interactuá utiliza la representación continua de las emociones basada en valencia y excitación, solo detecta 4 estados emocionales: Sorpresa, Felicidad, tristeza y neutralidad. Las características del audio fueron: Pitch, Flux, Rolloff-95, Centroid, Zero-crossing rat, SNR y el ritmo comunicativo. Para el audio se utilizaron dos modelos de clasificación uno basado en un árbol de decisión y el otro basado en reglas de decisión. Para el reconocimiento de las imágenes utilizaron dos softwares CERT y SHORE. En el componete de fusión utilizaron un método basado en reglas de decisión.

1.2. Planteamiento del problema

La interacción humano-robot (HRI) es un tema ampliamente estudiado en la actualidad. En el área especifica de HRI, se espera que para el año actual los robots cuenten con algoritmos que les permitan detectar las emociones de una persona mientras se comunica de manera natural; de manera que para el año 2023 los robots puedan emplear la información suministrada acerca de la detección del estado emocional de la persona, para tomar decisiones de forma autónoma de acuerdo a la emoción del humano.

En este sentido y siguiendo con la tendencia actual, para lograr la detección de emociones a partir de una comunicación natural es indispensable tratar información multimodal. En este caso las emociones se detectan a partir de la información de los gestos faciales y el habla correspondientes a los canales no-verbal y verbal; por ello en este proyecto se abordará el problema especifico de reconocimiento multimodal de estados afectivos en el área de HRI.

En la Figura 1.1 se muestra un escenario de ejemplo de interacción entre una persona y un robot, en la cual el reto fundamental se centra en la identificación de las señales de la expresión facial y la voz (emitidas por un humano de manera que con esta información sea posible clasificar que emoción experimenta la persona en ese momento.

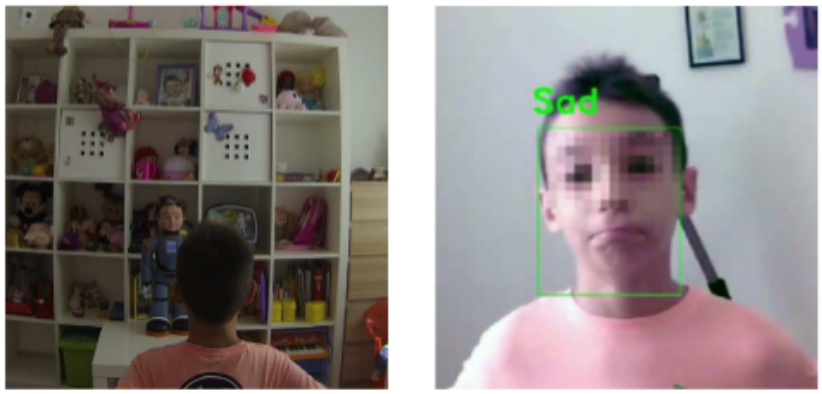

Figura 1.1: HRI: Detección de emociones a través del canal no-verbal y verbal

Como se ha mencionado anteriormente, en este caso es interesante estudiar HRI desde la perspectiva del análisis de las emociones con la intención de que un robot humanoide pueda interactuar de forma correspondiente a la emoción detectada en la

persona que le habla y emite gestos faciales; pero para lograrlo es necesario construir un sistema robusto que incluya algoritmos de inteligencia artificial y procesamiento de señales de video y audio.

Con este objetivo, para el proyecto se propone implementar un sistema que pueda capturar las señales expresión facial y voz de los canales de comunicación no-verba y verbal, respectivamente y a partir de ellas procesarlas y clasificarlas para lograr la identificación de la emoción o estado afectivo que la persona transmite a través de los canales.

Finalmente se considera un conjunto de posibles emociones a detectar las cuales son: alegría, enfado, neutralidad, sorpresa, disgusto, tristeza, miedo y desprecio.

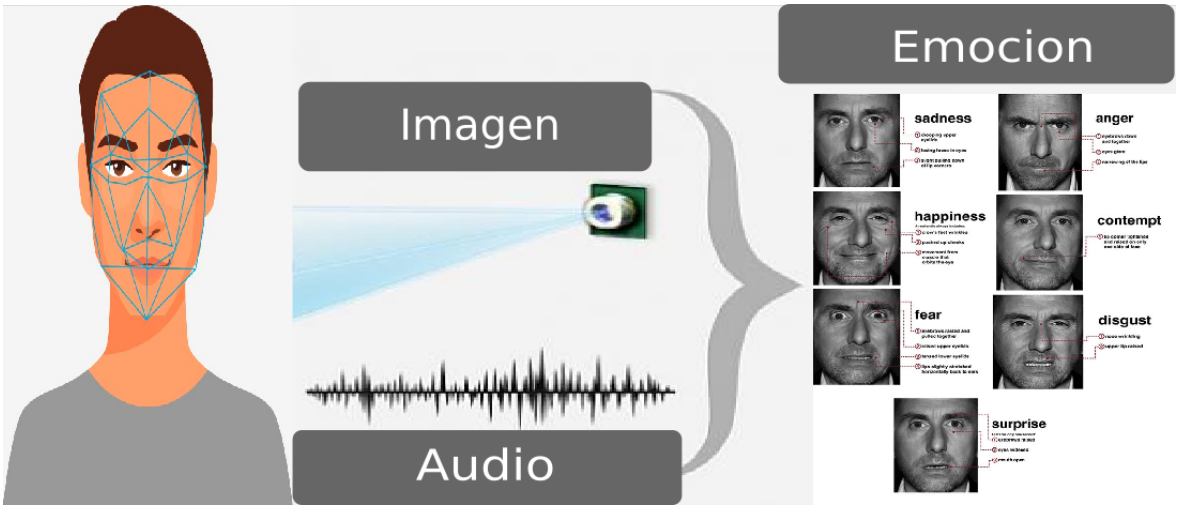

Figura 1.2: Descripción grafica del planteamiento del problema

1.3. Justificación

Hoy en día, existe una necesidad muy grande de encontrar sistemas que permitan reconocer emociones para mejorar la interacción con humanos. Muchos de estos sistemas requieren no solo reconocer emociones, sino tomar decisiones de manera inteligente en base a la emoción detectada.

La mayoría de los sistemas de reconocimiento utilizan como entrada una sola señal. Ejemplo de ello se encuentra en las investigaciones desarrolladas por Faria et all.,[2] y Rzuri et al., [5] para el reconocimiento de emociones a través de los gestos y el habla respectivamente. Sin embargo existen limitantes en la forma en que se recibe la señal, esto es debido a que no siempre podemos enfocar el rostro de un sujeto en la posición correcta y no siempre es posible escuchar correctamente al sujeto, por lo que, recibir y procesar múltiples señales permite el desarrollo de sistemas de reconocimiento mucho mas precisos.

Por tal motivo, el trabajo aquí desarrollado, se orienta a el reconocimiento de emociones multimodal, con el objetivo de realizar un sistema de reconocimiento de estados efectivos utilizando señales verbales y no-verbales, en el contexto de la interacción humano robot.

1.4. Objetivos

1.4.1. Objetivo General

Diseñar e implementar un sistema de reconocimiento de estados afectivos multimodal, mediante los canales verbal y no-verbal correspondiente al habla y los gestos faciales respectivamente, basado en la aplicación de modelos de aprendizaje supervisado en la inteligencia artificial, en el contexto de la interacción humano - robot.

1.4.2. Objetivos Específicos

- Recolectar información para una base de datos de entrenamiento y prueba para ambos tipos de señales: gestos faciales y audio.
- Seleccionar los modelos de aprendizaje supervisado para ambos tipos de señales.
- Implementar un sistema de reconocimiento de señales para identificar las emociones de una persona que interactúa con el robot.
- Plantear pruebas de evaluación para la precisión del reconocedor de emociones de forma individual para cada modelo y de manera conjunta.
- Evaluar el sistema de reconocimiento implementado a través de métricas que permitan validar la eficacia del sistema.

1.5. Metodología

Para el desarrollo del proyecto propuesto se propone un diseño metodológico con enfoque cuantitativo, basado en características propias del sistema de inteligencia artificial construido para la identificación de las emociones a partir de señales multimodales. Como fuentes de información se tienen bases de datos como IEEE y Springer, eventos internacionales como IROS (Intelligent Robots and Systems), ICRA (International Conference on Robotics and Automation) y HRI (Human Robot Interaction), en la cual se describen avances en temáticas especificas de visión, inteligencia y aprendizaje de máquina en el contexto HRI. Partiendo del análisis del estado del arte y la revisión teórica de los temas relaciona- dos al proyecto planteado se seleccionan las herramientas matemáticas y computacionales que serán empleadas para la implementación del sistema de inteligencia artificial que permitirá la identificación de las emociones, obteniendo así el diseño de la solución. Posteriormente se implementa el sistema de inteligencia artificial diseñado para la detección y finalmente se valida el desempeño a través de pruebas para medir el error de clasificación del sistema propuesto. Finalmente se tomarán los resultados mediante un conjunto de mediciones definidas y se establecerán medidas de evaluación para dichos resultados, encontrando métricas de rendimiento del sistema propuesto.

1.6. Alcance y limitaciones

- Para la base de datos de entrenamiento se utilizaran la población estudiantil, específicamente estudiantes nuevo ingreso.
- Este proyecto se enfoca en el reconocimiento de emociones multimodal, la integración del sistema con el robot o los protocolos de comunicación no son parte de este.

Bibliografía

[1] J. Castillo, A. Castro Gonzalez, F. Alonso Martin, A. Fernandez Caballero, and M. Salichs, "Emotion detection and regulation from personal assistant robot in smart environment," pp. 179–195, 01 2018.

[2] F. C. F. Diego R. Faria, Mario Vieira and C. Premebida, "Affective facial expressions recognition for human-robot interaction," *IEEE*, pp. 805–810, 2017.

[3] M. Ménard, P. Richard, H. Hamdi, B. Daucé, and T. Yamaguchi, "Emotion recognition based on heart rate and skin conductance," pp. 26–322, 01 2015.

[4] A. K. Perez Hernandez, "Identificación de señales multimodales para reconocimiento de emociones en el contexto de interacción humano-robot," Universidad de los Andes Colombia, Facultad de Ingenieria, Departamento de Ingeniería de Sistemas y Computación, Proyecto de Maestria , 2016.

[5] J. G. Rázuri, D. Sundgren, R. Rahmani, A. Larsson, and A. M. Cardenas, "Affective facial expressions recognition for human-robot interaction," *(IJARAI) International Journal of Advanced Research in Artificial Intelligence,*, 2015.