

Reconocimiento de emociones en video mediante servicios cognitivos

Francesc Boixader Estévez, Julio Iglesias Lavirgen, Albert Balaguer Antolí,
Daniel Palma Robleda, Noelia Camargo Fernández
Escola Universitària d'Informàtica Tomàs Cerdà (Adscrita a la Universitat Autònoma de Barcelona)
Sant Cugat del Vallès
{francesc.boixader,julio.iglesias, albert.balaguer, daniel.palma}@eug.es,
ncamargo@campus.eug.es

Resumen

El uso de video tutoriales se ha convertido en una práctica habitual en la docencia, incluso a nivel presencial. Diseñar video tutoriales de calidad es un reto para aquel docente que pretende mantener la atención del estudiante a lo largo de todo el visionado. No obstante, la utilización de este recurso de manera inapropiada puede generar el efecto contrario al que se persigue. El visionado de vídeos con contenidos obsoletos, anacrónicos, y/o con una proporción de texto excesiva, puede provocar desinterés y desmotivación en los alumnos.

Esta problemática fue la que motivó a una estudiante del *Grado en Informática y servicios* que se imparte en nuestro centro, a presentar una propuesta de Trabajo de fin de grado (TFG) con un claro objetivo: detectar los distintos estados de ánimo de un estudiante cuando está visionando un video tutorial, con la intención de aprovechar esta información para un mejor diseño de los video tutoriales utilizados en clase.

En este trabajo se presenta el diseño y la implementación de una aplicación de escritorio para Windows, escrita en lenguaje de programación C#, que ayuda al docente a evaluar, de manera asíncrona, la idoneidad del material audiovisual utilizado como soporte, observando qué impacto tienen los video tutoriales sobre los alumnos y analizando cómo éstos responden a su contenido. Esta aplicación permite la extracción de datos cuantitativos, acerca de las emociones detectadas en videos almacenados localmente.

La metodología utilizada consistió en efectuar grabaciones de varios estudiantes mientras visualizaban un video tutorial para, con posterioridad, analizar dichas grabaciones mediante el uso de las herramientas disponibles en servicios cognitivos de Microsoft. Estos servicios permiten el reconocimiento

de emociones tales como sorpresa, disgusto, miedo, felicidad, entre otras. La utilización de dichas funcionalidades generó algunos problemas técnicos que, una vez superados, permitieron asociar las emociones mostradas por el estudiante durante el visionado, con los instantes del video tutorial causantes de dichas emociones.

Los resultados obtenidos han sido satisfactorios hasta el momento, aunque no se ha podido efectuar un número de análisis suficientemente grande como para extraer conclusiones de peso. Lamentablemente, la temporalidad propia del TFG ha condicionado el tiempo que se ha dedicado a probar la herramienta implementada. No obstante, se ha abierto una línea de trabajo que será seguida por otros estudiantes en el futuro.

Abstract

The use of video tutorials has become a common practice in teaching, even at the classroom level. Design quality video tutorials is a challenge for the teacher who intends to maintain the student's attention throughout the viewing. However, the use of this resource inappropriately can generate the opposite effect to the one pursued. The viewing of videos with obsolete, anachronistic content and/or with an excessive proportion of text can cause disinterest and demotivation in the students.

This problem was the one that motivated a student of the Degree in *Computer Science and services* that is taught in our center, to present a proposal for *Trabajo de fin de grado* (TFG) with a clear objective: detect the different emotions of a student when is watching a video tutorial, with the intention of taking advantage of this information for a better design of the video tutorials used in class.

This paper presents the design and implementation of a desktop application for Windows, written in C #

programming language, which helps the teacher to evaluate, asynchronously, the suitability of the audiovisual material used as support, observing what impact the video tutorials have on the students and analyzing how they respond to their content. This application allows the extraction of quantitative data, about the emotions detected in videos stored locally.

The methodology used consisted of making recordings of several students while viewing a video tutorial to later analyze those recordings by using the tools available in Microsoft cognitive services. These services allow the recognition of emotions such as surprise, disgust, fear, happiness, among others. The use of these functionalities generated some technical problems that, once overcome, allowed to associate the emotions shown by the student during the viewing, with the instants of the video tutorial that caused those emotions.

The results obtained have been satisfactory up to now, although it has not been possible to carry out an analysis number large enough to draw conclusions of weight. Unfortunately, the temporality of the TFG has conditioned the time that has been dedicated to test the implemented tool. However, a line of work has been opened that will be followed by other students in the future.

Palabras clave

Reconocimiento de emociones en video, detección de expresiones faciales, servicios cognitivos, inteligencia artificial, aprendizaje máquina, comunicación no verbal.

1. Introducción

Las expresiones faciales son utilizadas por los seres humanos y los animales para su comunicación, interacción social y, especialmente, para transmitir las emociones a sus pares [1]. A través de la expresión facial es posible mostrar estados emocionales específicos, aportando de esta manera información motivacional y de comunicación [2].

El rostro humano es considerado el principal sistema de señales para mostrar las emociones, además de ser el área más importante y compleja de la comunicación no verbal. El rostro es la parte del cuerpo que más de cerca y detalladamente se observa durante cualquier tipo de interacción social [3]. A través de la cara se pueden extraer infinidad de claves que proporcionan información útil para esa interacción y la adecuada comunicación; de igual manera permite identificar emociones y predecir actos respecto a lo que observamos.

Hace casi 150 años que existe discusión sobre la universalidad de las expresiones faciales de las emociones a través de las diferentes culturas. Darwin fue el primero en documentar la presencia de las emociones como características del ser humano [4], y algunos autores han identificado más de 100 emociones diferentes, aunque, en realidad, todas son variaciones o mezclas de unas pocas consideradas como básicas (ver Figura 1).

Ekman y Friesen afirmaron en sus estudios que hay seis emociones básicas involuntarias, inconscientes y universales [5]. A partir de sus trabajos, se concluye que todas las personas, sin importar su edad, género, etnia o cultura, poseen la habilidad de manifestar algunas emociones exactamente de la misma manera, expresadas principalmente a través del rostro.

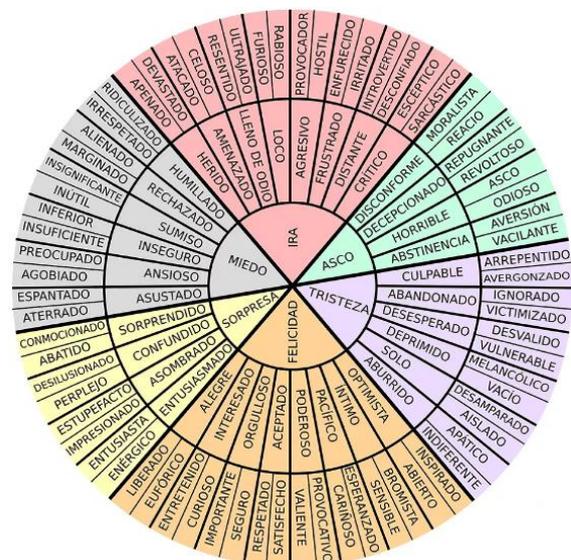


Figura 1: La Rueda de las Emociones.

Se puede definir al reconocimiento facial de la expresión emocional como la capacidad de todos los individuos de reconocer formas básicas de manifestación afectiva mediante expresiones faciales de la emoción, universales e innatas [6, 8].

El reconocimiento facial sigue siendo un área de investigación en curso. Los avances en nuevas tecnologías, concretamente la aparición y actual perfeccionamiento, de un amplio abanico de servicios cognitivos, ha abierto la posibilidad de progresar enormemente en el área. Actualmente se aplica en una gran variedad de sectores como la seguridad, salud, educación, etc.

Los servicios cognitivos son aquellos que permiten dotar a los equipos informáticos de cualquier tipología, de las capacidades que tenemos los humanos de percibir y procesar aquello que ha sido percibido,

apoyándose en las capacidades que proporciona la Inteligencia artificial (IA) y el aprendizaje máquina o *Machine learning* (ML).

Las emociones tienen una función adaptativa. Es un estado que sobreviene súbitamente, de forma más o menos violenta, y más o menos pasajera. Además, las emociones juegan un papel muy importante en nuestra educación: no hay aprendizaje sin emoción. Todo aprendizaje implica una experiencia afectiva porque los centros nerviosos y los neurotransmisores que regulan la motivación y la emoción están involucrados en sus procesos¹.

En este artículo se presenta *EmotionInVideo* (EIV), una aplicación de escritorio que faculta al docente con datos válidos para valorar la idoneidad del material audiovisual utilizado como soporte en su materia. El uso de video tutoriales se ha convertido en una práctica habitual en la docencia, incluso a nivel presencial, ya sea como complemento de la sesión, ya sea como un elemento de consulta para aquel alumno que no puede asistir a una determinada clase. Diseñar video tutoriales de calidad se ha convertido en una necesidad para aquel docente que pretende mejorar la calidad de la docencia que imparte.

El objetivo es observar qué impacto tienen los video tutoriales sobre los alumnos, y analizar cómo éstos responden al contenido percibido a lo largo del tiempo. El análisis de los resultados obtenidos proporcionará al docente elementos objetivos en base a los cuáles podrá mejorar el material utilizando durante su práctica habitual.

En contraste con otras alternativas disponibles, este *software* no analiza capturas en tiempo real, sino que trabaja de manera asíncrona, reconociendo emociones en archivos previamente almacenados en disco. Primero extrae cada uno de los fotogramas, o *frames*, que forman un video, y los envía al módulo de reconocimiento facial *Face API*² de los servicios cognitivos de Azure, los servicios de computación en la nube de Microsoft.

Con la implantación de esta herramienta se pretende aportar un instrumento para garantizar y, en la medida de lo posible, mejorar, la calidad del material empleado en la docencia a nivel universitario. Adicionalmente, podría aprovecharse la aplicación para hacer lo mismo con sesiones prácticas autónomas, asistencia a seminarios, e incluso evaluar el desarrollo de las sesiones teóricas.

¹ Francisco Mora en <http://www.biotmr.com/2014/04/15/la-neurociencia-demuestra-que-el-elemento-esencial-en-el-aprendizaje-es-la-emocion>.

² *API* es la abreviatura para el término anglosajón *Application Programming Interface* y hace referencia a una interfaz de programación de aplicaciones.

2. Trabajo relacionado

Existen múltiples soluciones que sacan buen provecho de los diferentes servicios cognitivos para reconocimiento facial disponibles. Muchas trabajan a partir de imágenes estáticas individuales, o mediante la *webcam* como fuente de video en tiempo real, pero a diferencia de EIV, ninguna permite reconocer emociones en videos almacenados de forma previa al análisis.

Dentro de la documentación oficial de los servicios cognitivos de Microsoft, integrados en la plataforma *cloud* de Azure, podemos encontrar varios ejemplos de uso en tiempo real³.

También pueden hallarse proyectos basados en otros proveedores de servicios cognitivos, como aquéllos que sacan provecho de la librería gratuita OpenCV, enfocada principalmente en la visión por computador en tiempo real. Encontramos muestras de aplicaciones que permiten detectar y reconocer caras en el video en tiempo real que captura la *webcam* en, por ejemplo, contribuciones a la documentación de la *API* de OpenCV⁴.

El código fuente es público y, con alguna modificación, permitiría reconocer emociones, pero esta opción no ha sido contemplada por los autores, por lo que son necesarios conocimientos en lenguajes de programación de alto nivel, como C#, en este caso, y Python o R, en muchos otros ejemplos.

3. EmotionInVideo

En este apartado se detalla cómo se ha abordado el recurso docente como instrumento para valorar la calidad e idoneidad de algunas prácticas pedagógicas, empleando como ejemplo los video tutoriales, utilizados habitualmente como soporte didáctico por el profesorado.

EIV se propone como un sistema automático de obtención de datos numéricos, fáciles de exportar a gráficos, para analizar y comparar los fotogramas que componen un video en el que se muestre a un alumno consumiendo un video tutorial. Esto es posible gracias al reconocimiento de emociones a partir de las expresiones faciales detectadas en las videgrabaciones. Para ello se emplean las funcionalidades de inteligencia artificial (IA) disponibles a través de la interfaz de *Face API*: tomando una expresión facial en una imagen como entrada, devuelve la confianza a través de un conjunto de emociones para cada cara de la imagen utilizando la *API*. Las emociones detectadas son ira, disgusto, miedo, felicidad, tristeza y sorpresa, las seis emociones básicas; además de neutralidad y desprecio. Se entiende que estas emociones se comunican

intercultural y universalmente con expresiones faciales particulares.

Además, se introduce la interfaz de usuario de la aplicación, escrita en lenguaje de programación C#, que permite al profesorado seleccionar un video y obtener un archivo de texto con una relación de la confianza para las emociones detectadas en cada una de las imágenes, así como su correspondiente número de fotograma respecto al video de origen.

3.1 Sistema de reconocimiento

Tal y como se muestra en la Figura 2, la idea principal de nuestra herramienta es proporcionar un entorno robusto que admita manejar videos capturados previamente. Estos videos pueden contener, por ejemplo, grabaciones de alumnos mientras miran video tutoriales en clase y, mediante llamadas a los servicios cognitivos de Azure, se procesarán las expresiones faciales detectadas en cada una de las imágenes que conforman el video.

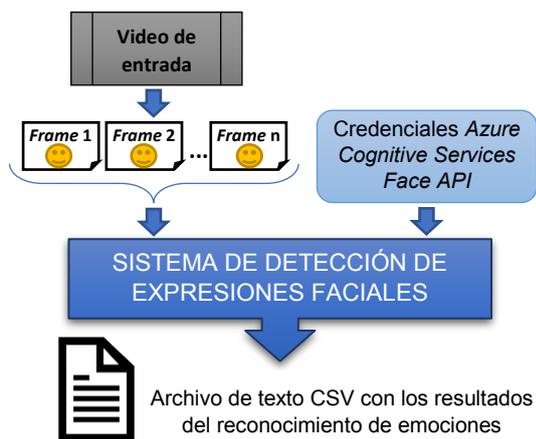


Figura 2: Diagrama conceptual del sistema de reconocimiento de emociones.

El sistema genera un archivo de salida que permite obtener datos numéricos acerca de las emociones reconocidas en dichas imágenes.

El objetivo es disponer de un instrumento que permita al docente comprobar con seguridad, contrastando con datos cuantificables, fiables y relevantes, si el material empleado durante el proceso de enseñanza-aprendizaje, tiene el impacto esperado

³ https://docs.microsoft.com/en-us/azure/cognitive-services/face/face-api-how-to-topics/howtoanalyzevideo_face.

⁴ https://docs.opencv.org/2.4/modules/contrib/doc/face_rec/tutorial/face_rec_video_recognition.html.

en los alumnos. De esta forma, los profesores pueden evaluar si su uso didáctico es eficaz o no, y mejorar la eficiencia de sus clases, optimizando el tiempo empleado en estas actividades (visionado de video tutoriales, o sesiones prácticas individuales).

3.2 Interfaz

La aplicación dispone de una simple ventana, donde se muestran los distintos elementos con los que el usuario puede interactuar: la zona de configuración de las credenciales de acceso a Face API, dos botones: uno para cambiar entre los modos de uso *Auto* o sencillo, y *Advanced*, con más opciones disponibles y, finalmente, otro botón, de ayuda, que muestra una pequeña lista de instrucciones paso a paso, diferente para cada uno de los modos de uso de la aplicación.

Por defecto, el software se inicia en el modo *Auto*. Operando en este modo, la única opción habilitada es la de seleccionar el archivo de video a procesar. El sistema admite archivos en formato MP4 y AVI.

En el modo *Advanced* se muestran dos botones adicionales. Si se dispone de los fotogramas que forman un video previamente convertidos en ficheros locales de imagen estática BMP o JPEG, haciendo clic en el botón *Select frames folder* se puede seleccionar la carpeta donde se almacenan estos archivos de imagen.



Figura 3: Captura de la interfaz de EIV durante el reconocimiento de emociones.

Una vez iniciado el proceso de reconocimiento de emociones existentes en los rostros de las imágenes, que se van mostrando por pantalla, aparece un botón *Stop detection* para detener la ejecución en cualquier instante.

Además, en el espacio situado a su derecha, van apareciendo los valores detectados para cada emoción, en cada una de las imágenes. Todos los mensajes relevantes para el usuario se muestran en esta misma ubicación, en el centro del margen inferior.

Cuando se ha terminado la detección en alguna de las imágenes, se muestra un texto con los valores

detectados para cada emoción y, además, aparece en la imagen un recuadro que rodea el rostro procesado.

En el caso de haber más de una cara por *frame*, el *software* solamente tendrá en cuenta una de ellas. Por otro lado, si EIV no detecta ninguna cara en alguna de las imágenes, establecerá a cero, para ese fotograma, el valor para todo el conjunto de emociones.

Finalmente, el usuario encontrará, en la esquina inferior derecha, un botón *Exit* para salir de la aplicación. Una vez clicado, se cierra EIV y se abre automáticamente el último archivo de salida CSV generado, de existir. A la izquierda de este botón, y únicamente si se ha procesado algún video de forma satisfactoria, aparece un botón *Open* que abre el último archivo CSV con resultados, generado por la aplicación, en un editor de texto.

3.3 Paso a paso

Para hacer un uso genérico de la herramienta, los pasos a seguir son los que se enumeran a continuación:

1. Introducir unas credenciales válidas de acceso a Face API. Es necesario tanto seleccionar la región de conexión, como insertar la clave de suscripción vinculada.
2. Hacer clic en el botón *Select video*. Seleccionar el archivo de video MP4 o AVI que se desea procesar.
3. Esperar a que se genere una carpeta de archivos de imagen JPEG, con cada uno de sus *frames*, a partir del video seleccionado. Aparecerá un mensaje de confirmación al terminar la conversión. Una vez finalizada, la aplicación procederá a llamar a los servicios cognitivos para cada una de las imágenes creadas, generando y mostrando un archivo de texto CSV con los resultados de la confianza de las emociones detectadas en cada una de ellas. Se puede detener el proceso en cualquier momento, pulsando el botón *Stop detection*.

Si se activa el modo avanzado pueden ejecutarse los pasos 2 y 3 del proceso anterior en dos etapas que el docente puede realizar por separado:

2. Pulsar el botón *Convert video*. Tras seleccionar un archivo de video en la ventana emergente, esperar a que se genere un directorio con los fotogramas correspondientes.
3. Hacer clic en *Select frames folder*. Tras seleccionar la carpeta recién creada, esperar a que se genere el archivo CSV con la relación de emociones detectadas.

En el instante en el que la primera imagen es procesada, aparece el botón *Open* a la izquierda del botón de salida.

3.4 Instalación en Windows 10

Para empezar a hacer uso de EmotionInVideo desde el sistema operativo es necesario descargar e instalar el paquete de la aplicación, alojado en DropBox⁴. Haciendo doble clic en el archivo ejecutable *setup.exe*, se inicia el asistente de instalación y, tras breves instantes, la herramienta está lista para su utilización.

3.5 Uso desde Microsoft Visual Studio

También puede emplearse la aplicación, de una forma más técnica, desde el entorno de programación. Es necesario disponer de Microsoft Visual Studio, preferiblemente en su versión 2017, así como de conocimientos de lenguajes de programación de alto nivel, en concreto del conocido como CSharp o C#. Para poder hacer uso de esta característica será necesario importar el paquete del proyecto *software* de EmotionInVideo, escrito en C#, repositorio que también se encuentra disponible en DropBox⁴. Además, deben añadirse los siguientes módulos extra desde el gestor de paquetes NuGet, incluido en el mismo Microsoft Visual Studio 2017:

- Accord.Video.FFMPEG (en su versión para PCs con arquitecturas x86)
- Microsoft.Azure.CognitiveServices.Vision.Face

Todo el código fuente se ha comentado para facilitar la comprensión de su lógica a desarrolladores de posibles proyectos futuros.

Por otro lado, hay que tener presente que el número máximo de fotogramas a procesar por segundo está limitado por la capacidad de cómputo del equipo informático utilizado para el procesado, y por la latencia máxima de llamadas permitidas a la API, según la suscripción de la que se disponga.

Con el acceso gratuito se permiten actualmente 30 mil llamadas mensuales, a una tasa máxima de 20 transacciones por segundo. Por este motivo, tanto la versión ejecutable de EIV como el código fuente presentado contienen una limitación temporal para adaptarse a la restricción que propone el proveedor de los servicios cognitivos.

Asimismo, existe la posibilidad de crear cuentas Azure gratuitas para estudiantes que permiten procesar 10 imágenes por segundo. Esta ha sido la opción escogida por los autores, dado que una de ellas cursa, actualmente, estudios universitarios de Grado en el ámbito de la Ingeniería Informática. Su colaboración en el presente proyecto constituye, esencialmente, su Trabajo fin de Grado (TFG).

⁴ https://www.dropbox.com/sh/qwl12j6n3d9tt7y/AABcEVC7L604c_I14BjtLJDa?dl=0

4. Resultados

Como ya se ha mencionado con anterioridad, cuando finaliza el proceso de reconocimiento de emociones se genera un fichero CSV (*Comma Separated Values*).

En este caso de estudio, se ha preferido el uso del punto y coma como separador con la intención de evitar posibles problemas provocados por la confusión entre la coma decimal y el símbolo utilizado para separar campos.

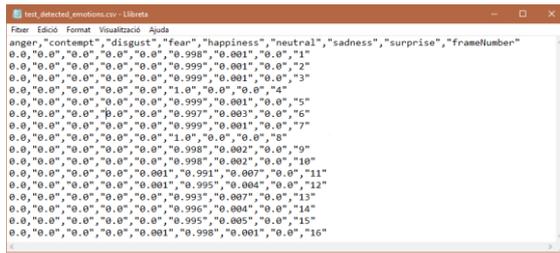


Figura 4: Ejemplo de archivo de texto CSV de salida

La Figura 4 muestra, a título de ejemplo, un contenido generado por el *software* tratado en el artículo, EmotionInVideo.

Los ficheros de salida que crea EIV con los resultados del reconocimiento de emociones presentan el siguiente formato:

En la primera línea de cada archivo de texto CSV generado por la aplicación se crea una cabecera, para facilitar el análisis de los datos numéricos que aparecen en las líneas subsiguientes.

La primera cifra de cada una de ellas señala el número de *frame* que corresponde a la imagen de la que se han extraído los datos, dentro de la secuencia del archivo de video de origen. Los demás valores numéricos indican la confianza, determinada por la IA de los servicios cognitivos, para cada una de las emociones detectadas en el rostro del fotograma correspondiente.

Los archivos CSV resultantes del uso de EIV pueden ser importados desde cualquier aplicación o plataforma de representación gráfica de datos. En la Figura 5 se puede ver una gráfica de línea de tiempo cuyos datos de origen han sido generados mediante el uso del *software* sobre un video de prueba, previamente capturado para tal fin.

Asimismo, estos archivos de salida CSV también pueden ser importados desde aplicaciones de hojas de cálculo, tales como la contenida en la suite de OpenOffice, el servicio Spreadsheet de Google o Microsoft Excel.

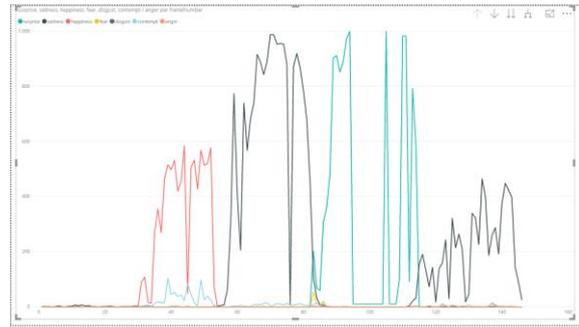


Figura 5: Gráfico basado en un archivo de texto CSV generado mediante el uso de EmotionInVideo.

El fichero de ejemplo de salida mostrado en la Figura 4 posee el aspecto que expone la Figura 6 tras su importación a hoja de cálculo. Los archivos de texto CSV son fácilmente interpretables como un cuadro por cualquiera de las tecnologías de hojas de cálculo mencionadas en el párrafo anterior, entre otras, además de por todos los editores de texto disponibles en el mercado.

| | A | B | C | D | E | F | G | H | I |
|----|-------|----------|---------|------|-----------|---------|---------|----------|----------|
| 1 | anger | contempt | disgust | fear | happiness | neutral | sadness | surprise | frameNum |
| 2 | 0 | 0 | 0 | 0 | 0 | 0,998 | 0,001 | 0 | 1 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0,999 | 0,001 | 0 | 2 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0,999 | 0,001 | 0 | 3 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0,999 | 0,001 | 0 | 5 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0,997 | 0,003 | 0 | 6 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0,999 | 0,001 | 0 | 7 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 8 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0,998 | 0,002 | 0 | 9 |

Figura 6: Muestra de la salida en formato cuadro.

Como ya se ha visto, los ficheros de salida que genera EIV incluyen una línea de datos por cada fotograma procesado.

El aplicativo supone que el primer archivo de imagen cargado es el primer fotograma del video a analizar y, desde este punto, va incrementando el valor en 1, hasta haber procesado el último de los *frames* de la carpeta seleccionada. Este es el valor que hallamos en la columna *frameNumber*.

Las demás columnas incluyen el valor de confianza estimado por los servicios cognitivos de reconocimiento facial para cada una de las ocho emociones evaluadas, enumeradas en el apartado 3 del presente documento. Esta cifra se refiere a cuán segura se encuentra la IA de que realmente esa sea la emoción mostrada en una imagen, a partir de la expresión facial detectada en el rostro que en ella aparezca. Siempre se encuentra dentro del intervalo entre 0 y 1, ambos

incluidos, siendo 0 muy improbable que el rostro de la imagen exprese esa emoción, y 1 una muy alta probabilidad de ser realmente la emoción manifestada.

La suma total de los valores de confianza para cada fotograma, o línea del archivo de resultados, es igual a uno.

Por otra parte, es necesario un trabajo previo por parte del docente para extraer conocimiento de los archivos generados como resultado del uso de EmotionInVideo. Requiere dedicarle cierto tiempo, según el nivel de ofimática del profesor, pero tras estos sencillos pasos, se pueden asociar las emociones detectadas con su correspondiente instante dentro del video que el alumno está viendo, y viceversa.

En el momento de la captura de las grabaciones es importante marcar con claridad el inicio de los videos. Esto permitirá sincronizar los resultados con el video original. Cuando se ha convertido una grabación en imágenes, el docente debe eliminar todos aquellos fotogramas irrelevantes para que EIV no los procese.

Para conocer la equivalencia temporal de cada fotograma debe realizarse la siguiente operación:

$$\text{Marca de tiempo} = \text{N}^\circ \text{ fotograma} / \text{FPS}$$

Donde FPS (*Frames Per Second*) representa el número de fotogramas por segundo de un video dado. Este dato lo indica EIV en el momento de la conversión a *frames*. Lo ideal para graficar los datos directamente sobre un eje de marcas temporales (en segundos) es agregar una columna de datos nueva, tras importar un CSV generado por EIV a una hoja de cálculo.

En lo relativo a la representación gráfica, es recomendable obviar los registros del campo “neutral”, puesto que facilita la interpretación de las líneas temporales del resto de emociones.

Partiendo de los CSV obtenidos de las grabaciones de cada participante visualizando el mismo video se puede comparar el impacto ejercido sobre cada uno de ellos en el mismo instante de tiempo. Esto permite corroborar la idoneidad del material, incluso a nivel de los elementos individuales que componen el video en cuestión. Como muestra, tomemos un video tutorial de 10 minutos, parte del proceso docente de uno de los autores, en el que se pretende introducir a los alumnos en el ecosistema del BI, plagado de tecnicismos, conceptos abstractos y neologismos, la mayoría en inglés.

Si analizamos un instante del video en el que se incluye una referencia cómica de la cultura popular, en un intento de amenizar un contenido muy denso, esperaríamos obtener una reacción positiva en los alumnos: una amplia sonrisa o una mueca de sorpresa.

En este caso, se grafica únicamente el intervalo de tiempo a analizar, en el que aparece el estímulo ante

el que el docente espera obtener una reacción positiva por parte de los alumnos.

Sin embargo, como expone la Figura 8, esto sólo sucede con el “Participante 2”, que expresa algo de felicidad (línea azul cielo) en el instante que aparece la referencia. Cabe subrayar que dicho alumno nació en los 80. Los demás participantes mantienen un nivel constante de indiferencia e, incluso, de desprecio (línea naranja) y tristeza (azul marino) ante el contenido.

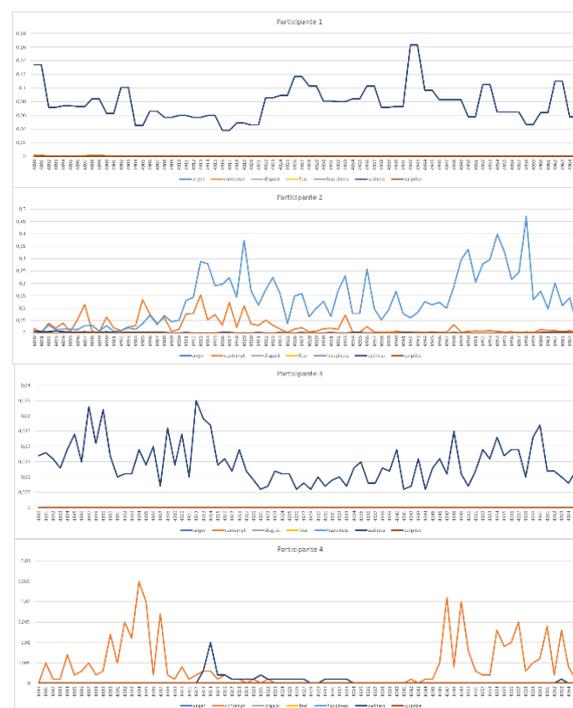


Figura 6: Reacciones de varios participantes a un elemento cómico incluido en un video tutorial.

Este sencillo análisis permite validar la hipótesis inicial que cuestionaba la idoneidad del material utilizado de soporte al proceso docente. De hecho, pone de manifiesto la obsolescencia de este material, demostrando la necesidad de ir adaptándolo para que logre impactar satisfactoriamente en las nuevas generaciones.

5. Conclusiones

Tras la experiencia de uso de EmotionInVideo, los autores consideran que este software provee al ecosistema docente de una eficaz herramienta para el desarrollo de proyectos relacionados con el análisis automatizado de expresiones faciales en archivos de video.

Los resultados recogidos permiten pensar que esta aplicación puede constituir un aporte muy interesante para la comunidad docente, que puede hacer buen uso

de ella como herramienta para lograr mejoras pedagógicas en sus asignaturas, ya sea como útil para evaluar el impacto en los alumnos del material utilizado en el proceso de enseñanza-aprendizaje, tal y como se ha expuesto en el caso de ejemplo, o como fuente de inspiración o apoyo para otro tipo de soluciones.

Se ha contrastado la aplicabilidad y eficacia de EmotionInVideo en nuestro propio proceso docente, por lo cual creemos que puede ser de ayuda a otros docentes, incluso no relacionados con la docencia en el ámbito de la Informática. La intención a la hora de diseñar e implementar la herramienta ha sido la de proveer un entorno robusto que permita optimizar el tiempo y esfuerzos del profesorado, con el propósito de ayudar a mejorar, en la medida de lo posible, la calidad de su trabajo diario, permitiendo al docente observar si el impacto que el material audiovisual ejerce sobre su alumnado es el esperado y, tras una evaluación con datos empíricos, actuar en consecuencia, ya sea retirando o renovando el material, de ser necesario, o seleccionando para su materia solamente aquellos elementos que realmente sean de utilidad a los estudiantes.

Se ha procurado mantener la interfaz de la aplicación lo más sencilla posible para garantizar que, sea cual sea su nivel de conocimientos en informática, todo aquel que lo desee pueda utilizarla.

El análisis del apartado anterior revela que parte de nuestro material se estaba quedando obsoleto y necesitaba ser actualizado. Se trataba de recursos, por lo general bastante técnicos, relacionados con la inteligencia de negocios y la minería de datos [7], que incluían referencias a la cultura pop de principio de siglo en un intento por amenizar el visionado de los alumnos.

Sin embargo, los datos que obtuvimos demostraban que esas alusiones en clave de humor habían quedado antiguas y ya no ejercían el impacto positivo pretendido sobre los alumnos que cursan actualmente estudios de Grado en Informática pertenecientes, la mayoría de ellos, a la generación Y, comúnmente conocidos como *millennials*.

En base a esta información, se concluyó que se debía modificar dicho material, manteniendo todos los fundamentos teóricos, pero sustituyendo los elementos de *soporte emocional* por menciones a temas de actualidad que realmente exciten de la forma esperada a los estudiantes.

6. Agradecimientos

En primer lugar, agradecer al profesorado de las *Escoles Universitàries Gimbernat i Tomàs Cerdà* que ha participado en este proyecto, su colaboración. Sus aportaciones han resultado ser de un gran valor, tanto en lo que se refiere a conocimientos como, especialmente, en lo relativo a consejos y dedicación.

Por último, un especial reconocimiento y agradecimiento a los alumnos que participaron directamente en las videograbaciones, por el tiempo dedicado y por colaborar, de esta manera, a la generación de material fundamental para la redacción de este artículo.

Referencias

- [1] D. A. Álvarez y M. L. Guevara. Reconocimiento de expresiones faciales prototipo usando ICA. En *Scientia et Technica*, Año XV, número 41, Universidad Tecnológica de Pereira, Colombia, mayo 2009.
- [2] Ana María Anguas-Wong y David Matsumoto. Reconocimiento de la expresión facial de la emoción en mexicanos universitarios. En *Revista de Psicología*, Vol. XXV (2), páginas 277 – 293, Fondo Editorial, San Francisco, 2007.
- [3] Vicente E. Caballo. Manual de evaluación y entrenamiento de las habilidades sociales, Planeta, Madrid, 2005.
- [4] Charles Darwin. *The expression of the emotions in man and animals*, Penguin Classics, julio 2009 (publicado, por primera vez, por John Murray, en Londres, en 1872).
- [5] Paul Ekman y Wallace V. Friesen. *Constants across cultures in the face and emotion*. En *Journal of Personality and Social Psychology*, Vol. 17 (2), páginas 124 – 129, febrero 1971.
- [6] Paul Ekman, Richard E. Sorenson y Wallace V. Friesen. *Pan-cultural elements in facial displays of emotion*. En *Science*, 164(3875):86-8, Nueva York, mayo 1969.
- [7] María Pérez Marqués. *Minería de datos a través de ejemplos*, RC Libros, Madrid 2014.
- [8] James A. Russell. *Is there universal recognition of emotion from facial expression? A review of cross-culturally studies*. En *Psychological bulletin*, Vol. 115, páginas 102 – 41, febrero 1994.