

# Multimodal Machine Learning

Ronald Caravaca Mora

Instituto Tecnológico de Costa Rica.  
Heredia, Costa Rica, Email: ronycaravaca0691@gmail.com

**Resumen**—En este documento se desarrolla el estado del arte sobre Machine Learning Multidominal, se aborda su concepto, su importancia y los desafíos técnicos que se involucran. Además, de los métodos actuales que se utilizan para solventar estos desafíos.

**Index Terms**—Multimodal, fuente unimodal, representación, traducción, alienamiento, fusión, co-aprendizaje, Machine Learning.

## I. INTRODUCCIÓN

LA realidad que nos rodea envuelve una gran cantidad de modalidades que en general las personas perciben a través de sus sentidos, por lo que la multi-modalidad se puede asociar a *modalidades sensoriales* que representan nuestros canales de comunicación como la visión o el habla. Los dispositivos portátiles han dado lugar a un creciente interés en aplicaciones de detección complejas, ya que traen una variedad de sensores que permiten obtener datos multimodales permitiendo su uso en aplicaciones como predicción de movimiento, detección de objetos, rastreo, etc. Cada modalidad de detección posee una perspectiva única, las combinaciones de múltiples modalidades pueden aumentar la calidad de detección y superar su potencial.

Cuando un problema contiene datos de diferentes modalidades se cataloga como *multimodal*. El uso de múltiples modalidades puede proporcionar información complementaria y aumentar la precisión del proceso general de toma de decisiones, mejorar la predicción de eventos u obtener información desconocida que solamente mediante un proceso de análisis multimodal puede ser conocida.

Para que la Inteligencia Artificial progrese en la comprensión del mundo que nos rodea, esta debe ser capaz de interpretar y entender mejor su naturaleza multimodal.

El campo de investigación del Machine Learning o aprendizaje automático multimodal trae algunos desafíos únicos para los investigadores debido lo heterogéneo de los datos. Sin embargo, aprender de fuentes multimodales ofrece la posibilidad de capturar las correspondencias entre las modalidades y obtener una comprensión mas profunda de los fenómenos naturales.

## II. DESAFÍOS TÉCNICOS

De acuerdo con Baltrusaitis et al. [4], el *Machine Learning* multimodal busca crear modelos que puedan procesar y relacionar datos provenientes de diferentes modalidades, dando la

posibilidad de obtener correspondencias entre estas y dar un mejor entendimiento del problema. Baltrusaitis et al. [4] identifica cinco retos técnicos principales para la implementación de *Machine Learning* multimodal, los cuales son:

1. *Representación*: El primero de los desafíos es el problema de como representar los datos multimodales de manera que se aproveche la redundancia y complementariedad de las múltiples modalidades. Por ejemplo, el lenguaje se representan a través de símbolos mientras que la audición y la visión se representan con señales.
2. *Traducción*: Otro de los retos es como traducir o “mapear” datos de una modalidad a otra. Los datos no solo son heterogéneos, sino que la relación entre las modalidades es a menudo abierta o subjetiva. Por ejemplo, existen varias formas correctas de describir una imagen y es posible que no exista una traducción perfecta.
3. *Alineamiento*: Un tercer desafío es identificar las relaciones directas entre elementos de dos o más modalidades. Por ejemplo, es posible que queramos alinear los pasos de una receta con un video que muestre el plato que se está preparando. Para enfrentar este desafío, necesitamos medir la similitud entre las diferentes modalidades y abordar las posibles dependencias y ambigüedades a largo plazo.
4. *Fusión*: Un cuarto desafío es unir información de dos o más modalidades para realizar una predicción. Por ejemplo, para el reconocimiento audiovisual del habla, la descripción visual del movimiento del labio se fusiona con la señal del habla para predecir las palabras habladas. La información proveniente de diferentes modalidades puede tener una potencia predictiva y una topología de ruido variables, posiblemente con datos faltantes en al menos una de las modalidades.
5. *Co-aprendizaje*: Un quinto desafío es transferir el conocimiento entre las modalidades, su representación y sus modelos predictivos. Esto se ejemplifica mediante algoritmos de co-entrenamiento. El aprendizaje explora cómo el aprendizaje de conocimientos de una modalidad puede ayudar a un modelo computacional entrenado en una modalidad diferente. Este desafío es particularmente relevante cuando una de las modalidades tiene recursos limitados.

En la tabla I se muestran como se distribuyen estos retos en algunas aplicaciones de *Machine Learning*, como se muestra en la tabla, en el análisis de emociones y afecto están presentes

cada uno de estos retos.

	Representación	Traducción	Alineamiento	Fusión	Co-aprendizaje
Detección de voz	x		x	x	x
Detección de eventos	x		x	x	x
Emociones y afecto	x	x	x	x	x

Tabla I: Retos para cada aplicación de *Machine Learning*. [4]

### III. REPRESENTACIÓN

La representación de los datos en un modelo computacional representa un reto importante en campo del Machine Learning. Una representación multimodal es una representación de datos provenientes de diferentes fuentes o modalidades, como imágenes, audio o texto, donde una representación hace referencia a un vector o tensor de cada una de estas modalidades. Esto trae algunas complejidades como el manejo de datos heterogéneos, manejo de diferentes niveles y tipos de ruido, además de datos faltantes en alguna modalidad.

Para tener una buena representación se definen las siguientes propiedades: suavidad, coherencia temporal y espacial, esparcidad, correlación y agrupamiento natural de los datos.

El avance en las técnicas de Machine Learning ha ayudado a mejorar la representación y manejo de datos multimodales. Anteriormente, la representación de modalidades se hacía mediante un “diseño manual” dependiendo de la naturaleza de la misma. Por ejemplo, para la representación de imágenes se utilizaba métodos como “Scale Invariant Feature Transform (SIFT)”, una señal de voz se representaba mediante “Mel-frequency cepstral coefficients (MFCC)”, pero actualmente utilizando técnicas de Machine Learning se pueden representar imágenes y señales de audio mediante descripciones aprendidas por una red neuronal.

De acuerdo con Baltrusaitis et al. [4] existen dos maneras de hacer la representación multimodal: Representación unida (*joint*) y Representación coordinada (*coordinated*).

#### III-A. Representación unida

Esta representación combina las señales unimodales en un mismo espacio multimodal. Matemáticamente se expresa como:

$$\mathbf{x}_m = f(\mathbf{x}_0, \dots, \mathbf{x}_n) \quad (1)$$

Donde la representación  $\mathbf{x}_m$  es realizada mediante una función  $f$  que puede ser una red neuronal, modelo gráfico probabilístico o modelo secuencial. En la figura 1 se muestra el concepto de la representación unida.

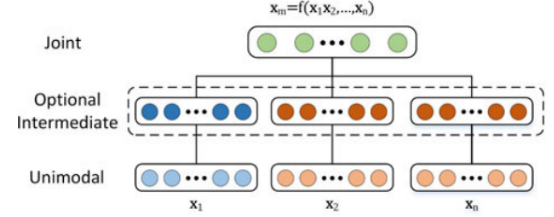


Figura 1: Representación unida. [4]

Las *redes neuronales* son comúnmente utilizadas debido a su capacidad de ser pre-entrenadas de manera no supervisada cuando se tienen pocos datos etiquetados. Los *modelos probabilísticos* utilizan modelos de variables aleatorias para la presentación, son útiles debido a que no necesitan entrenamiento no supervisado y a partir de ellos se pueden obtener redes neuronales. Los *modelos secuenciales* utilizan redes neuronales recurrentes para la representación, son útiles cuando se deben representar secuencias no finitas de alguna modalidad.

#### III-B. Representación coordinada

Matemáticamente se expresa como:

$$f(\mathbf{x}_1) \sim g(\mathbf{x}_2) \quad (2)$$

Donde cada modalidad tiene una representación propia, ya sea  $f$  o  $g$  que luego se mapea al un espacio multimodal. La proyección al espacio multimodal es independiente de cada modalidad resultando en un espacio coordinado entre ellas ( $\sim$ ). Lo que se busca resaltar en el espacio coordinado es la *semejanza* entre las representaciones unimodales, minimizando la distancia entre ellas. Las redes neuronales también son utilizadas de esta manera, debido a su capacidad de aprendizaje son entrenadas para que aprendas a distinguir representaciones coordinadas. En la figura 2 se muestra el concepto de la representación coordinada.

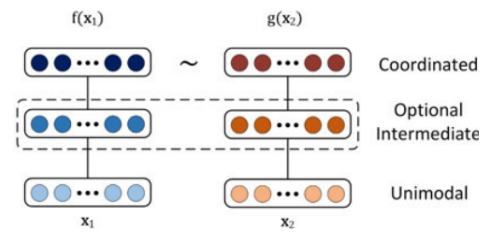


Figura 2: Representación coordinada. [4]

### IV. TRADUCCIÓN

Uno de los problemas mas comunes del procesamiento multimodal tiene que ver con la traducción o mapeo de una modalidad a otra, esto es, dada una información en una modalidad, la tarea consiste en generar la misma información en otra modalidad. Por ejemplo, dada una imagen generar una descripción textual de esa imagen, o dada la descripción generar la imagen. Los métodos utilizados se puede categorizar en dos tipos; *basado en ejemplos* y *generativos*.

#### IV-A. Basado en ejemplos

Estos utilizan diccionarios para la traducción entre modalidades, por lo que su entrenamiento esta restringido a estos. Entre sus variantes hay algoritmos que utilizan la traducción directamente mientras que otros utilizan reglas mas complejas para definir la traducción.

#### IV-B. Generativos

Estos pueden generar la traducción basados únicamente en las fuentes unimodales. Algunos algoritmos utilizan la gramática para generar oraciones especificas basándose en objetos o verbos, inician detectando conceptos a alto nivel como objetos en imágenes o eventos en videos, luego se generan procedimientos basados resultados gramáticos predefinidos. Otros métodos codifican la información de la fuente unimodal a una representación que pueda ser decodificada para generar la modalidad objetivo, esto a través de una red neuronal entrenada para la codificación y decodificación. Por otro lado, hay métodos que continuamente generan la modalidad objetivo basándose en una entrada constante (*stream*) de datos de la fuente unimodal.

### V. ALINEAMIENTO

Se define como *alineamiento* el proceso de encontrar relaciones o correspondencias entre subcomponentes de las fuentes unimodales. Por ejemplo, dada una imagen y su descripción, encontrar los elementos descritos en la imagen. Esto es particularmente importante para recuperación de datos multimedia. Como buscar el contenido de videos basado en texto o encontrar escenas en una película donde aparezca un objeto o suceda un evento en especifico. Se puede categorizar el alineamiento multimodal en dos tipos; *explícito* e *implícito*.

#### V-A. Explícito

En este tipo de alineamiento se busca explícitamente alinear subcomponentes entre modalidades donde la métrica mas importante es la semejanza. Esta semejanza puede ser definida manualmente o aprendida mediante un red neuronal.

#### V-B. Implícito

Este alineamiento es usado como un paso intermedio en busca de obtener un objetivo diferente al alineamiento. Se han utilizado en aplicaciones de “speech recognition”. Debido a que el objetivo no es alineamiento, el modelo no se entrena con ese objetivo, pero puede implícitamente hacer alineamientos a través de aprendizaje no supervisado.

### VI. FUSIÓN

Según Atrey et al. [3], el proceso de fusión multimodal trae algunos desafíos importantes, que se describen a continuación:

#### VI-A. Niveles de fusión

Una de las primeras consideración a tomar cuando se fusionan múltiples modalidades es decidir cual estrategia fusión utilizar. La estrategia mas utilizada es *fusión a nivel de features*, que es conocida como *fusión temprana*. La otra estrategia es *fusión a nivel de decisión* o *fusión tardía*. Una combinación de ambas estrategia se conoce como *fusión híbrida*.

**VI-A1. Fusión a nivel de features:** En esta estrategia los features son extraídos de los datos, se combinados y se utilizan como entrada a una única unidad de procesamiento que hace el análisis. Por ejemplo, en un modelo de detección de rostros se combinan los features multimodales como el color de la piel y señales de movimiento en un vector de features más grande que se toma como la entrada para el modelo de detección. La figura 3 describe este proceso, donde  $F_n$  representa los features,  $FF$  la unidad de fusión de features y  $AU$  la unidad de procesamiento.

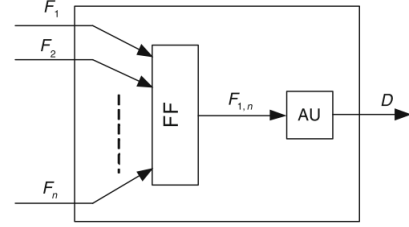


Figura 3: Fusión a nivel de features.

La fusión a nivel de features tiene la ventaja de utilizar la correlación entre múltiples features de diferentes modalidades en una etapa temprana que ayuda a mejorar el procesamiento. Además, solo requiere una fase de aprendizaje. Sin embargo esta proceso puede enfrentar problemas de sincronía entre features.

**VI-A2. Fusión a nivel de decisión:** En esta estrategia las unidades de análisis proveen las decisiones  $D_n$  locales basadas los features  $F_n$  de cada modalidad individual. Las decisiones son fusionadas en un vector de decisión que es analizado mas adelante para obtener la decisión final, como lo muestra la figura 4.

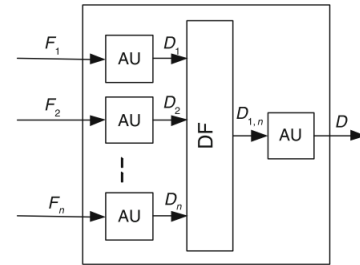


Figura 4: Fusión a nivel de decisión.

La decisión final la toma una unidad de análisis a nivel semántico. Lo cual es una ventaja en comparación con la estrategia de features, ya que en este caso a nivel semántico se tiene una misma representación para las decisiones, lo cual hace que el análisis sea menos complejo. También permite adecuar el análisis independiente para cada modalidad de acuerdo a su naturaleza y utilizar los métodos (HMM, RCC, SVM, ...) adecuadamente. Por otro lado, este proceso tiene la desventaja que desaprovecha la correlación entre features. Además, el proceso de entrenamiento puede ser tedioso consumir mas tiempo.

**VI-A3. Fusión híbrida:** Para explotar las ventajas de las estrategias de fusión tanto a nivel de features como a nivel de decisión, se puede por utilizar una estrategia de fusión híbrida, que es una combinación de ambas estrategias, como se muestra en la figura 5. Donde los features son fusionados primero por una unidad  $FF$  y luego el vector de features es analizado por una  $AU$ . Al mismo tiempo, otros features individuales son analizadas por diferentes  $AU$  y sus decisiones se fusionan usando una unidad  $DF$ . Finalmente, todas las decisiones obtenidas de las etapas anteriores son fusionadas por una unidad  $DF$  para obtener la decisión final.

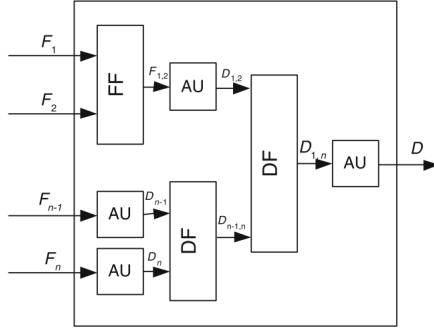


Figura 5: Fusión híbrida.

Un enfoque de fusión híbrida puede utilizar las ventajas de las estrategias de fusión temprana y tardía.

#### VI-B. Como fusionar?

Existen diferentes métodos que se utilizan en la fusión multimodal de datos, los cuales pueden dividirse en tres categorías como se muestra en la tabla II.

Categoría	Métodos
<i>Rule-based methods</i>	Linear Weighted Fusion Majority Voting Rule Custom-defined Rule
<i>Classification-based methods</i>	Support Vector Machine Bayesian Inference Dampster-Shafer Theory Dynamic Bayesian Networks Neural Networks Maximum Entropy Model
<i>Estimation-based methods</i>	Kalman Filter Extended Kalman Filter Particle Filter

Tabla II: Métodos de fusión multimodal [3]

**VI-B1. Rule-based fusion methods:** Se basan en una variedad de reglas básicas para combinar información. Estos incluyen métodos estadísticos basados en reglas como la fusión ponderada lineal (suma y producto), MAX, MIN, AND, OR, votación mayoritaria. Aprovechan la sincronía y la correlación entre features, por lo que son bastante

utilizados en la estrategia de fusión a nivel de features. Sin embargo, se han utilizado como métodos de fusión a nivel de decisión. La tabla III se muestran algunos de estos métodos, las modalidades utilizadas y sus aplicaciones.

**VI-B2. Clasification-based fusion methods:** Son métodos regularmente utilizados en problemas de clasificación. Estos métodos se pueden clasificar aún más como modelos generativos y discriminativos desde la perspectiva del aprendizaje automático. Por ejemplo, Bayesian inference y Dynamic Bayesian networks son modelos generativos, mientras que la máquina de vectores de soporte y las redes neuronales son modelos discriminativos. Se utilizan tanto a nivel de features como a nivel de decisión. La tabla IV se muestran algunos de estos métodos, las modalidades utilizadas y sus aplicaciones.

**VI-B3. Estimation-based fusion methods:** En esta categoría se incluyen los métodos de filtro de Kalman, filtro de Kalman extendido y filtro de partículas. Estos métodos se han utilizado principalmente para estimar mejor el estado de un objeto en movimiento basado en datos multimodales. Por ejemplo, para la tarea de seguimiento de objetos, se fusionan múltiples modalidades como audio y video para estimar la posición del objeto. Los detalles de estos métodos son los siguientes. La tabla V se muestran algunos de estos métodos, las modalidades utilizadas y sus aplicaciones.

## VII. CO-APRENDIZAJE

El último de los desafíos es el co-aprendizaje, el cual ayuda a procesar una modalidad con datos pobres utilizando el conocimiento obtenido de otra modalidad. Es particularmente utilizado en aplicaciones donde se tiene recursos limitados de datos, ruido en alguna de las modalidades o datos no confiables. Este método es utilizado únicamente en el proceso de entrenamiento. Existen tres enfoques para el co-aprendizaje; *paralelo*, *no paralelo* e *híbrido*.

#### VII-A. Paralelo

Este enfoque es utilizado en aplicaciones donde las observaciones de diferentes modalidades están claramente enlazadas entre sí. En otras palabras, los datos multimodales provienen de las mismas instancias, como datos de audio y video de una misma fuente. El objetivo general es crear etiquetas para los datos no etiquetados a partir de las otras modalidades, transfiriendo así información entre modalidades.

#### VII-B. No paralelo

Estos métodos no requieren que las modalidades compartan las instancias pero sí categorías o conceptos. Este enfoque puede ayudar a aprender representaciones, permitir una mejor comprensión del concepto semántico e incluso realizar un reconocimiento de objetos invisibles. Este tipo de aprendizaje a menudo se logra mediante el uso de representaciones multimodales coordinadas.

Método	Nivel de fusión	Modalidades	Aplicaciones
Linear weighted fusion	Feature	Video (trayectorias) Video (colores, texturas) Video (Movimiento)	Human tracking Face detection, monologue detection and traffic monitoring Monologue detection
	Decisión	Audio y video Imágenes Texto, audio y video	Semantic concept detection Speaker recognition Video retrieval Image retrieval
Majority voting rule	Decisión	Audio (Habla)	Speaker identification from audio
Custom-define rules	Decisión	Visual, texto Habla, gestos	Multimodal interaction with robot Human computer interaction

Tabla III: Rule-based fusion methods [3]

Método	Nivel de fusión	Modalidades	Aplicaciones
SVN	Decision	Video, Audio	Semantic concept detection
	Hibrida	Vlideo, Audio Imágenes, Texto	Multimedia data analysis Image clasification
Bayesian inference	Feature	Audio, video	Speech recognition
	Decision	Audio, video	Spoken digit recognition
Dempster-Shafer theory	Hibrida	Audio, video, texto, web log	Sports video analysis
	Feature	Video, audio	Segmentation of satellite images
Dynaminc Bayesian networks	Decision	Video, audio	Video clasification
	Hibrida	Video, audio	Human computar interaction
Neural networks	Feature	Audio, video, habla	Speech recognition
	Decision	Imágenes, video	Photo annotation
Maximun entroy model	Hibrida	Texto, audio	Topic clustaring in video
	Feature	Audio, video	Human tracking
	Decision	Carga de CPU, Cargas en redes, Imágenes	Human activity monitoring
	Hibrida	Images	Image recognition
	Feature	Taxto, imágenes	Semantic image indexing

Tabla IV: Clasification-based fusion methods [3]

Método	Nivel de fusión	Modalidades	Aplicaciones
Filtro Kalman y sus variantes	Feature	Audio, video, posicion	Multiple speakers tracking
	Decision	Audio,video	Single object localization
Filtro particula	Feature	Audio, video	Single speaker tracking
	Decision	Audio, video	Multiple speakers tracking

Tabla V: Estimation-based fusion methods [3]

### VII-C. Híbrido

En la configuración de datos híbridos, dos modalidades no paralelas están unidas por una modalidad compartida o un conjunto de datos.

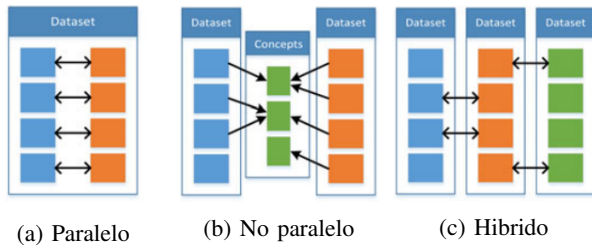


Figura 6: Tipos de co-aprendizaje

## VIII. TRABAJOS REALIZADOS MEDIANTE MACHINE LEARNING MULTIMODAL

A continuación una lista de trabajos realizados mediante técnicas de Machine Learning multimodal:

1. Automatic social interaction analysis with audio and visual nonverbal cues. Aran [1]

2. Fusing audio-visual non-verbal cues to detect dominant people in small group conversations. Aran and Gatica-Perez [2]
3. The MatchNMI dataset: a novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates. Cabrera-Quiros et al. [5]
4. Automatic nonverbal analysis of social interaction in small groups: A review. Image and Vision Computing. Gatica-Perez [6]
5. Estimating Communication Skills Using Dialogue Acts and Nonverbal Features in Multiple Discussion Datasets. Okada et al. [7]
6. Visual and multimodal analysis of human spontaneous behaviour: Introduction to the Special Issue. Pantic and Cohn [8]
7. It's not you, it's me: Detecting flirting and its misperception in speed-dates. Ranganath et al. [9]
8. Being bored? Recognising natural interest by extensive audiovisual integration for real-life application. Schuller et al. [10]

## IX. CONCLUSIÓN

El machine learning multimodal tiene gran aplicación hoy en día y resulta importante en cuando lo donde se busca es tener una mayor comprensión del entorno en que el vivimos. Ya que claramente nuestro mundo es multimodal por naturaleza, percibimos y procesamos grandes cantidades de información a través de nuestros sentidos, información que también podemos obtener mediante sensores y analizarlas en modelos computacionales. Esto representa una complejidad técnica que puede ser abordada mediante los desafíos expuestos en este documento y mediante el uso de los métodos de Machine Learning avanzados que se han expuesto.

## REFERENCIAS

- [1] O. Aran. Automatic social interaction analysis with audio and visual nonverbal cues — Görsel ve işitsel sözsüz öğeler ile otomatik sosyal etkileşim analizi. In *SIU 2010 - IEEE 18th Signal Processing and Communications Applications Conference*, pages 220–223, 2010. ISBN 9781424496716. doi: 10.1109/SIU.2010.5652754.
- [2] O. Aran and D. Gatica-Perez. Fusing audio-visual nonverbal cues to detect dominant people in small group conversations. In *Proceedings - International Conference on Pattern Recognition*, pages 3687–3690, 2010. ISBN 9780769541099. doi: 10.1109/ICPR.2010.898.
- [3] P.K. Atrey, M.A. Hossain, A. El Saddik, and M.S. Kankanhalli. Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems*, 16(6):345–379, 2010. doi: 10.1007/s00530-010-0182-0.
- [4] T. Baltrušaitis, C. Ahuja, and L.-P. Morency. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, 2019. doi: 10.1109/TPAMI.2018.2798607.
- [5] L. Cabrera-Quiros, A. Demetriou, E. Gedik, L.V.D. Meij, and H. Hung. The MatchNMingle dataset: a novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates. *IEEE Transactions on Affective Computing*, 2018. doi: 10.1109/TAFFC.2018.2848914.
- [6] Daniel Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, 27(12):1775 – 1787, 2009. ISSN 0262-8856. doi: <https://doi.org/10.1016/j.imavis.2009.01.004>. URL <http://www.sciencedirect.com/science/article/pii/S0262885609000109>.
- [7] Shogo Okada, Yoshihiko Ohtake, Yukiko I Nakano, Yuki Hayashi, Hung-Hsuan Huang, Yutaka Takase, and Katsumi Nitta. Estimating Communication Skills Using Dialogue Acts and Nonverbal Features in Multiple Discussion Datasets. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI 16, page 169176, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450345569. doi: 10.1145/2993148.2993154. URL <https://doi.org/10.1145/2993148.2993154>.
- [8] Maja Pantic and Jeffrey F Cohn. Visual and multimodal analysis of human spontaneous behaviour: Introduction to the Special Issue. *Image and Vision Computing*, 27(12):1741 – 1742, 2009. ISSN 0262-8856. doi: <https://doi.org/10.1016/j.imavis.2009.07.001>. URL <http://www.sciencedirect.com/science/article/pii/S0262885609001504>.
- [9] R. Ranganath, D. Jurafsky, and D. McFarland. It’s not you, it’s me: Detecting flirting and its misperception in speed-dates. In *EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009*, pages 334–342, 2009.
- [10] Björn Schuller, Ronald Mller, Florian Eyben, Jrgen Gast, Benedikt Hrnler, Martin Wllmer, Gerhard Rigoll, Anja Hthker, and Hitoshi Konosu. Being bored? Recognising natural interest by extensive audiovisual integration for real-life application. *Image and Vision Computing*, 27(12):1760 – 1774, 2009. ISSN 0262-8856. doi: <https://doi.org/10.1016/j.imavis.2009.02.013>. URL <http://www.sciencedirect.com/science/article/pii/S0262885609000316>.