

Detección de expresiones faciales con base en métodos de reducción de dimensión

Jorge Luis Ramos Zavaleta

Centro de Investigación en Matemáticas. Unidad Monterrey.

Email: jorge.ramos@cimat.mx

Resumen—Un rostro humano puede ser representado conceptualmente como un conjunto de partes distribuidas de manera esparcida: ojos, nariz, boca, etc. Esta colección de elementos nos permiten encontrar patrones específicos lo que nos permite diferenciar entre el rostro de un hombre y el de una mujer, así como distintas expresiones. En este reporte se presentan los resultados de aplicar NMF(non-negative matrix factorization) y LNMF(Local NMF) a diversas imágenes expresando una expresión distinta con el fin de generar un clasificador para dichas expresiones comparando ambos métodos para conocer su efectividad en dicha tarea.

I. INTRODUCCIÓN

Debido a los problemas que se generan en altas dimensiones ya sea por redundancias o superposiciones de características, el análisis de objetos en el espacio de los datos originales puede ser computacionalmente exhaustivo. Por lo que problemas que requieren el análisis de imágenes en particular de rostros generalmente se transforman a uno con un subespacio de dimensión más baja para que dicho problema sea más fácilmente tratable.

Asimismo debe considerarse que una representación en un subespacio suficiente bueno puede exponer estructuras escondidas y características inherentes al tipo de objetos con el que se quiere trabajar. Inclusive puede considerarse alguna bondad de dicho subespacio en el que trabajar con diferentes variaciones como de iluminación o traslaciones sea mucho más sencillo que si se tratara con el espacio original.

Esto nos lleva a establecer que la elección de una apropiada representación de los datos en un subespacio de menor dimensión que el original puede volverse imperativo para el éxito y eficacia de un sistema de reconocimiento de objetos o rostros como es nuestro caso.

Métodos de base lineal como PCA y análisis de factores son usualmente usados para generar representaciones de datos multivariados. En el caso de PCA, se considera una representación global de los datos conseguida a través de mantener solo los primeros k componentes que nos generen una cierta varianza para trabajar y desechando el resto de los componentes.

Para realizar PCA se eligen las direcciones de las proyecciones que maximicen la varianza total sobre dichas proyecciones. Un problema asociado con PCA es que se

pierde la esparcidad que pudieron tener los datos originales, además de que en el caso de imágenes los eigenvectores tienen cargas negativas lo que puede llevar a que se retengan variaciones en la imagen que pueden afectar negativamente el resultado del clasificador¹.

Un enfoque relativamente nuevo para la reducción de dimensionalidad de datos multivariados es la conocida factorización no negativa de matrices (NMF). Este enfoque ha comenzado a atraer una significativa atención a su uso. Cabe mencionar que el esquema de NMF pertenece a un tema mucho más amplio de estudio conocido como aproximación de rango bajo en donde se pueden encontrar algunos otros métodos muy conocidos como escalamiento dimensional y correlación canónica.

Para el caso de NMF como el nombre implica, se establecen restricciones de no negatividad sobre la factorización que se genera de la matriz de datos, permitiendo que solo se generen sumas, por lo que no hay cancelaciones o interferencia de patrones debido a algún valor negativo en las características.

Debido a este último punto se considera que dicha factorización lleva de manera natural a la noción de representación de imágenes basadas en sus partes. Esto debido a que con dichas restricciones NMF es capaz de generar representaciones basadas en partes localizadas. Particularmente en el caso de reconocimiento de rostros dichas representaciones pueden verse como una colección pequeña de características: ojos, boca, nariz que están distribuidos a lo largo del rostro.

Cabe aclarar que NMF al igual que PCA son enfoques basados en características globales, en nuestro caso de un rostro, por lo que las bases contendrán más de una sola característica en cada vector, por lo que en algunos problemas más específicos como el que tratamos pueden no ser la mejor opción.

Para ello se propone hacer uso de Local NMF, que surge como una propuesta de extensión a NMF en el que se permite encontrar características de manera local en el espacio de las

¹P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," 19(7), pp. 711–720, 1997.

imágenes. En nuestro caso, podría ser capaz de encontrar solo una característica particular para permitarnos discernir todo un rostro, como puede ser la forma de los labios, el tamaño de los ojos, la separación entre los ojos, tipo de cejas.

Debido a que nuestro propósito es descubrir la expresión facial de una persona en una imagen este enfoque puede poseer una ventaja ya que solemos discernir expresiones faciales a partir de características muy específicas como la forma de la cara, forma de la ceja, forma de los labios. Es decir, podríamos destacar algunos aspectos más específicos para ubicar expresiones más complejas.

II. PCA, NMF Y LOCAL NMF

PCA es uno de los métodos más usados en estadística y es por ello que es muy conocido y aplicado para hacer reducción de dimensión así como para intentar encontrar características en datos multivariados. El método consiste en encontrar direcciones de sobre las cuales las proyecciones de los datos generen la máxima varianza.

El enfoque de detección de rostros usando PCA es muy conocido y recibe el nombre de eigenfaces. Es un esquema muy simple para aplicar PCA a diversos rostros. Para ello se genera una matriz con cada rostro vectorizado y colocado como una columna de dicha matriz después se calcula la imagen media y se le resta a cada imagen de la matriz para centrar las imágenes y evitar problemas dados por traslaciones.

A esta matriz centrada se le calcula su factorización SVD y se considera el número de componentes a usar dependiendo el criterio que se elija para esto. Ya elegidos los componentes se proyectan las imágenes de la matriz centrada en dicho subespacio y cada nueva imagen se proyecta en dicho subespacio para clasificarla. Una manera simple de clasificar esta nueva observación es haciendo uso de alguna métrica para indicar su cercanía con alguna de las imágenes de la matriz centrada.

Por otro lado tanto NMF como su versión local son esquemas de aproximación de una matriz considerando un rango más bajo que el de la matriz original, por lo que podría parecer que no contienen propiedades que permitan ayudar a realizar una identificación de características que permitan realizar una clasificación eficaz. Sin embargo su naturaleza inherente de no negatividad permite entrever que podría acumular información muy importante de imágenes, texto o señales.

Para el caso de NMF se tiene que puede verse como una generalización de k-means² con lo que tenemos por seguro que podremos encontrar objetos relevantes dentro de una

imagen. En el caso específico de rostros esperamos que pueda identificar cosas como ojos, boca y la nariz en conjunto para permitarnos clasificar eficazmente.

II-A. NMF

A finales de los 1970's ya se había planteado la idea de usar una factorización de matrices con restricción de no negatividad, sin embargo el problema de optimización que esto conlleva es de tipo Np-hard por lo que se abandono la idea, y fue hasta que en 1999 Lee et al³ publican un resultado donde encuentran una forma de resolver dicho problema de optimización haciendo uso de una regla de actualización multiplicativa. A partir de dicho resultado se ha comenzado una extensa investigación este tipo de aproximación de rango bajo y de como generar extensiones del método para considerar problemas más complejos.

Cabe mencionar que a partir del trabajo de Lee se encontrado otras formas de resolver el problema de optimización, sin embargo uno de los problemas asociados a estos nuevos descubrimientos es que la factorización no es única. Aunque en la práctica se consideran algunas restricciones de esparcidad para generar dicha unicidad.

Tanto para NMF como para Local NMF en este trabajo se considera una como función de costo una versión adaptada de la divergencia de Kullback-Leibler. Para el caso de NMF se tiene que la forma de la divergencia de Kullback-Leibler es la siguiente

$$D(X|Y) = \sum_{i,j} \left(X_{ij} \log \frac{X_{ij}}{y_{ij}} - X_{ij} + y_{ij} \right)$$

Por lo que nuestro problema es

$$\min D(X|BH)$$

$$s. a. B \geq 0 \text{ y } H \geq 0$$

Y para resolver este problema de la minimización de dicha función se tiene la siguiente actualización multiplicativa

$$h_{pj} = h_{pj} \sum_{i=1}^n \left(\frac{b_{ip} a_{ij}}{\sum_{k=1}^r b_{ik} h_{kj}} \right)$$

$$b_{ip} = b_{ip} \sum_{j=1}^m \left(\frac{a_{ij} h_{pj}}{\sum_{k=1}^r b_{ik} h_{kj}} \right)$$

$$b_{ip} = \frac{b_{ip}}{\sum_{k=1}^n b_{kp}}$$

Estas funciones de actualización multiplicativa tienen garantizado que la divergencia de Kullback-Leibler sea no creciente con cada actualización.

²Ding, C. H., Li, T., & Jordan, M. I. (2010). Convex and semi-nonnegative matrix factorizations. IEEE transactions on pattern analysis and machine intelligence, 32(1), 45-55.

³Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755), 788.

II-A1. Local NMF: Como ya se menciono anteriormente, local NMF es una extensión del algoritmo original de NMF en el que se busca que se haga una búsqueda en el espacio local para encontrar características locales en lugar de las características globales que encuentran PCA y NMF. Por lo que en nuestro problema a resolver, permitiría encontrar características específicas que pueden determinar lo que define un rostro masculino o un rostro femenino.

Como es una extensión de NMF entonces debe considerar condiciones de optimalidad mayores que los de NMF. Nuestro problema inicial en este caso es dada la factorización

$$A \approx BH$$

Si definimos:

$$U = B^T B \text{ y } V = H H^t$$

Entonces para lograr que se establezca una búsqueda local se requiere que se cumplan las siguientes restricciones

- Máxima esparcidad en H.- Esta se logra minimizando la diagonal de U
- Máxima expresividad en B.-Esta se logra maximizando la traza de V
- Ortogonalidad de las bases.- Esta se logra minimizando la suma de las entradas de la matriz U

por lo que nuestra divergencia de Kullback-Leibler toma la siguiente forma

Divergencia de Kullback-Leibler

$$D(X|BH) = \sum_{i,j} \left(X_{ij} \log \frac{X_{ij}}{BH_{ij}} - X_{ij} + BH_{ij} \right) + \alpha \sum_{i,j} u_{ij} - \beta \sum_i v_{ii}$$

Por lo que el problema de minimización a resolver es idéntico al que se resuelve con NMF.

Y la nueva regla de actualización multiplicativa es

$$h_{pj} = \sqrt{h_{pj} \sum_{i=1}^n \left(\frac{b_{ip} a_{ij}}{\sum_k b_{ik} h_{kj}} \right)}$$

$$b_{ip} = b_{ip} \sum_j \left(\frac{a_{ij} h_{pj}}{\sum_i b_{ip} h_{pj}} \right) / \sum_j h_{pj}$$

$$b_{ip} = \frac{b_{ip}}{\sum_{k=1}^n b_{kp}}$$

y con ello se consigue alcanzar un mínimo local.

II-A2. Clasificando con NMF y Local NMF: Tanto para NMF como para LNMF se considera que la matriz B contiene la base de imágenes sobre el espacio reducido. Si x es una nueva imagen entonces para verla en el espacio de las imágenes reducidas se calcula la pseudoinversa de B y se multiplica por x menos la imagen media del conjunto de entrenamiento, y se procede a usar cualquier función de distancia con respecto a los vectores en el nuevo espacio

$$B^\dagger (train - imagen\ media)$$

III. RESULTADOS

Para el estudio se utilizo el conjunto de imágenes JAFFE16 consistente de fotografías de 10 mujeres asiáticas con diferentes expresiones cada una. El número de imágenes por mujeres es variable entre 17 y 19, y también se presenta variabilidad en el número de fotos por expresión facial en cada mujer. Siendo en total un conjunto de 180 imágenes.

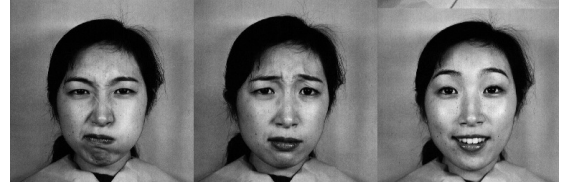


Figura 1. Imágenes representativas del conjunto JAFFE16

Cabe mencionar que este conjunto de imágenes contiene 2 tipos de etiquetado, en el primero se le pregunto a personas que indicaran la expresión facial que veían por lo que no se considero la expresión neutral, mientras que el segundo etiquetado corresponde a la expresión que se le pidió a las mujeres expresar en cada fotografía. Las etiquetas correspondientes a esta clasificación son las siguientes

HAP (happiness), SAD (sadness), SUR (surprise) ANG (anger), DIS (disgust) y NEU (neutral)

Se separo el conjunto de datos en imágenes de entrenamiento y prueba. Para el conjunto de entrenamiento se eligió una imagen de cada expresión facial por cada mujer por lo que se consideraron 60 imágenes para el entrenamiento y 120 para el conjunto de prueba. Aunque claramente el nivel de precisión aumentaría si se usaran más fotografías, el fin del trabajo es probar el poder de los métodos para reconocer gestos faciales. Aún cuando el nivel de información que se tiene para clasificar es menor de lo usual, de lograrse una buena clasificación indicaría que los métodos capturan elementos esenciales de las expresiones faciales.

A diferencia de PCA tanto NMF como Local NMF no tienen una regla específica de selección de rango sino que muchas veces es considerada una regla de pulgar considerando que la multiplicación del rango elegido por la suma de las dimensiones de la matriz original debe ser menor que la

multiplicación de las dimensiones de dicha matriz.

Debido a que el problema de resolver el algoritmo NMF y Local NMF es considerado NP-Hard el calculo de ambas factorizaciones se realizó en una interfaz de C++ con R, conocida como RCPP lo que permitió mejorar los tiempos de ejecución. Para establecer la clasificación se utilizó como métrica la correlación entre los vectores nuevos y los de la base, considerando la máxima correlación. Asimismo se consideraron diversos rangos para observar la variabilidad que se podría generar con el cambio entre rangos.

En las figuras 2 y 3 se puede observar que el nivel de precisión obtenido por ambos métodos coincide en general por lo que podemos sospechar que en general las expresiones faciales son una característica global mas que local.

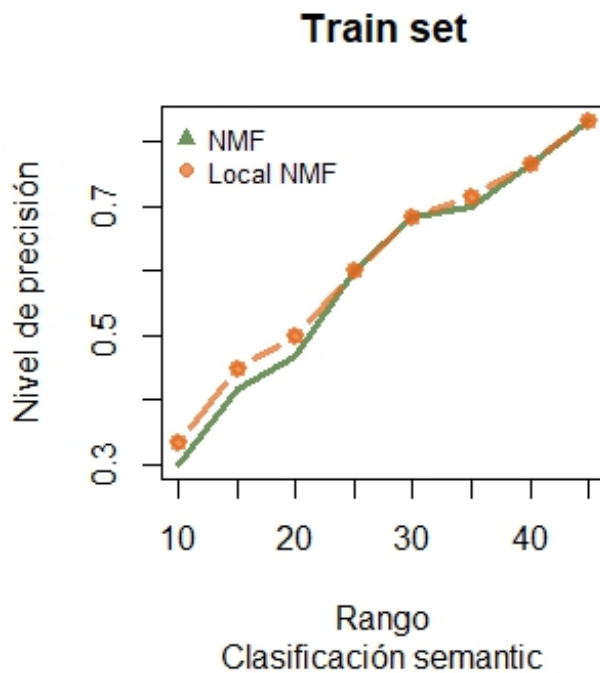


Figura 2. Nivel de precisión para varios rangos con NMF y LNMF para el conjunto de entrenamiento con el etiquetado semantic (5 categorías)

Por otro lado en las figuras 4 y 5 se puede observar la misma tendencia que en las anteriores pero el nivel de precisión aumento por lo que la entrada de una nueva categoría parece impactar de manera positiva la clasificación, esto es indicativo de que la categoría neutral suele ser bien reconocida por ambos algoritmos.

En este sentido cabe revisar las tablas de clasificación para observar si existe algún tipo de expresión que sea más fácilmente identificable de manera local. Para lo cual en las tablas de clasificación I y II se pueden observar una pequeña mejora en la clasificación de las categorías happiness y angry

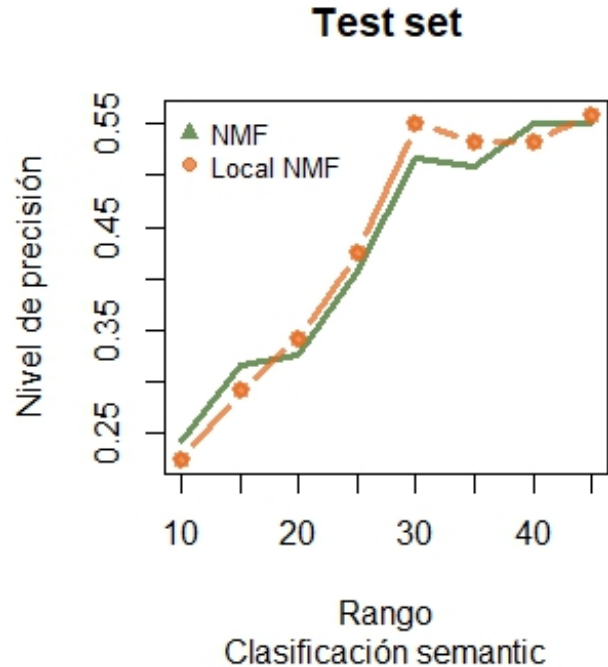


Figura 3. Nivel de precisión para varios rangos con NMF y LNMF para el conjunto de prueba con el etiquetado semantic (5 categorías)

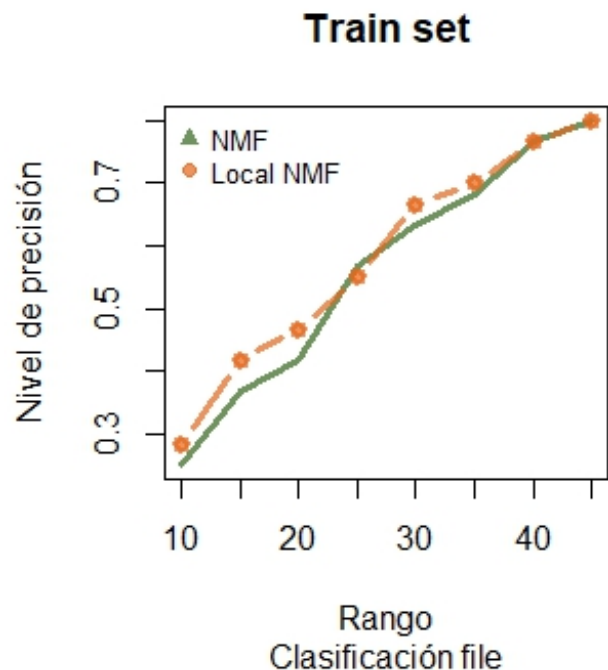


Figura 4. Nivel de precisión para varios rangos con NMF y LNMF para el conjunto de entrenamiento con el etiquetado file (6 categorías)

en el caso local, pero es muy pequeña la diferencia por lo que un enfoque de características globales parece ser suficiente

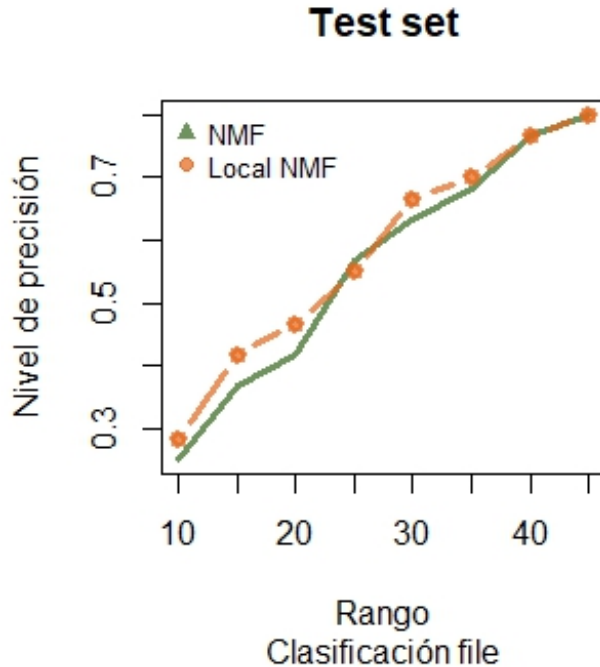


Figura 5. Nivel de precisión para varios rangos con NMF y LNMF para el conjunto de prueba con el etiquetado file (6 categorías)

Cuadro I
TABLA DE CLASIFICACIÓN PARA EL CONJUNTO DE PRUEBA CON 6 CATEGORIAS CON NMF CON RANGO 45

	ANG	DIS	HAP	NEU	SAD	SUR
ANG	14	2	4	2	5	1
DIS	3	15	0	0	1	1
HAP	1	0	9	3	4	4
NEU	0	0	1	14	1	1
SAD	2	2	4	1	10	2
SUR	0	0	2	0	0	11

para la clasificación.

IV. CONCLUSIONES

La clasificación de expresiones faciales parece seguir un esquema de características globales más que un esquema de características locales. Los niveles de clasificación encontrados para los conjuntos de prueba pueden parecer pequeños, pero dado que solo se quería comprobar el poder

directo de la reducción de dimensión los niveles de precisión son aceptables. Si se quisiera obtener niveles de precisión mayores se podría hacer uso de métodos de clasificación más sofisticados como SVM o AdaBoost multiclase para ello.

Cabe destacar el hecho de que el algoritmo Local NMF no es focalizado por lo que puede encontrar objetos en el rostro menos significativos para la clasificación y tendría que revisarse este mismo estudio con algunas versiones locales focalizadas como fisher NMF para asegurar que las expresiones faciales siguen en realidad un esquema global de características.

V. REFERENCIAS

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788.

Feng, T., Li, S. Z., Shum, H. Y., & Zhang, H. (2002). Local non-negative matrix factorization as a visual representation. In *Development and Learning, 2002. Proceedings. The 2nd International Conference on* (pp. 178-183). IEEE.

Li, S. Z., Hou, X. W., Zhang, H. J., & Cheng, Q. S. (2001). Learning spatially localized, parts-based representation. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.

Ding, C. H., Li, T., & Jordan, M. I. (2010). Convex and semi-nonnegative matrix factorizations. *IEEE transactions on pattern analysis and machine intelligence*, 32(1), 45-55.

Cuadro II
TABLA DE CLASIFICACIÓN PARA EL CONJUNTO DE PRUEBA CON 6 CATEGORIAS CON NMF CON RANGO 45

	ANG	DIS	HAP	NEU	SAD	SUR
ANG	15	2	3	2	3	1
DIS	3	13	0	0	2	1
HAP	0	0	11	3	4	4
NEU	0	0	1	14	2	2
SAD	2	4	4	1	10	2
SUR	0	0	1	0	0	10