# Resumen

Ver las fotos de un producto es muy util para su comercialización pero las herramientas de busqueda de imagenes actuales dan resultados erroneos o redundantes. En este articulo se presenta una nueva forma que autmoaticamente recoge imagenes relevantes explorando varios fuentes. Se utiliza la busqueda de amazon mas especifica junto a la de google mas general. Se propone una aproximacion de clustering condicional formulada ???? Utiliza ejemplos de Amazon y el ranking de imagenes de google. se pueden encontrar las misma imagenes en amazon y en google. el experimento muestra la efectividad de nuestra aproximacion.

# Intro

En las ultimas decadas se han hecho mucho esfuerzos para hacer buscadores de imagenes efectivos. La busqueda de imagenes de productos es un campo especifico y es muy util para visualizar los productos ya que ayuda a los usuarios y mejora las ventas.

La busqueda de imagenes de productos tiene sus propias propiedades. Suele haber imagenes de alta calidad asociadas a los productos creadas por los vendedores pero son muy limitadas y no nos dan la informacion que buscamos. En amazon solo hay 5 imagenes de cada preoducto.

Por otro lado en internet hay muchas mas fotos de esos productos que se pueden encontrar en google pero los resultados son inutiles o redundantes porque se indexan mediante informacion en texto.

REcoger imagenes correctas es util tanto para los usuarios como para los vendedores. El usuario puede hacerse una idea de como es el producto mirando diferentes imagenes. Un metodo que pueda recoger diferentes imagenes desde varios angulos y su descripcion para meterlo en una base de datos. La diversidad y la relevancia son importantes para la experiencia de usuario en la busqueda de imagenes.

En este trabajo se combinan los buscadores de google y amazon para recoger imagenes relevantes del producto. Se buscan las imagenes de un producto en amazon (creadas por los vendedores) y luego en google encontrando muchas imagenes incorrectas o redundantes. Se van a buscar las imagenes en amazon (de alta calidad) y algunas mas en google. Se propone una aproximacion de clustering condicional que se formula como un problema de afinidad de propagacion con las imagenes de amazon y teniendo en cuenta el ranking de imagenes de google. El conjunto de las imagenes seran las utilizadas para describir el producto.

Si no se encuntra un producto en amazon la busqueda se hace en google y tambien puede funcionar. Nuestro metodo esta diseñado para visualizar productos, es un algoritmo generalizado y se puede aplicar en aplicaciones similares.

*seccion 2 describe opiniones (textos) relativas al trabajo. la 3 introduce el metodo.4 se detalla como va. 5 introduce la estimacion usando el ranking de imagenes.6 experimentos. 7 trabajo*

# Related work

**Busqueda de imagenes de productos y diversificacion de busqueda de imagenes**

La busqueda de imagenes ha sido muy investigada pero la centrada en la imagen de producto no. Ya se ha probado aplicando un algoritmo con pagerank basado en links visuales rentre imagenes para mejorar la busqueda de imagenes de productos. tambien se ha estudiado una aplicacion cliente servidor en la que el usuario pueda hacer una foto de un producto y buscarlo con su movil.

Hay dos aproximaciones, una mediante el ranking de los resultados y otra mediante un clustering que encuentra un grupo de las imagenes mas representativas.

Primero Wang propuso un algoritmo que podia coger la relevancia y la diversidad explorando el contenido y los tags de las imagenes. Deselaers saco los algoritmos para optimizar la relevancia y la diversidad de los resultados de busqueda. En el 2 paso Leuken saco unas ligeras tecnicas de clustering en combinacion con una funcion de ponderacion de las caracteristicas visuales que captura diferentes aspectos de los resultados. Liu propuso un metodo para sumarizar los resultados cogiendo la relevancia y calidad. Jia hizo otro algoritmo para encontrar los ejemplares que representan mejor los resultados. Nuestro trabajo difiere del de Liu en 2 aspectos: El suyo solo mira resultados en google, el nuestro introduce amazon. El mira la info de ranking de forma heuristica y nosotros estimamos la relevancia del ranking y lo explora con nuestro metodo. por eso el nuestro mola mas.

# Image set summarization

Siempre se ha utilizado la propagacion de afinidad AP convencional. nostro usamos una nueva.

# Conventional AP

Tiene 2 caracteristicas que no puede controlar el AP convencional.

# discussion

Siempre se ha usado la AP con el ranking de las imagenes. Heuristicamente se pillan y se convierten en puntuaciones de relevancia o confianza. los experimentos demuestran que nuestro metodo es mejor.

El proceso se hace rapido ya que amazon da pocas imagenes y de google solo se recogen los primeros resultados.

# 5 ranking based relevance estimation for google

Se utiliza google para hacer una estimacion de relevancia con 100 productos.

# experimental settings

SE pillan 300 productos de la seccion de electronicas y muebles de amazon. Estos productos no tienen interseccion con los usados antes.Las categorias son rigidas pej no son en rollo ropa y sombreros. 1 porque no tiene mucho sentido para la visualizacion de categorias no rigidas y luego porque se puede coger los ejemplos de las categorias rigidas que se ajusten a esas.

Se cogen imagenes de google y amazon para cada producto. Para amazon se quitan los duplicados (parecidos) de forma simple. quedan de 1 a 7 imagenes por producto. de google se pillan las primeras 200 y se quitan las repetidas (parecidas) con otro metodo mas complejo. Para este metodo se utiliza un parametro α. Para ajustarlo se hacen 30 busquedas y se generan resultados con difentes valores de alfa. 5 usuarios observan los resultados y eligen la mejor imagen en terminos de visualizacion para un producto. Ahora se compara la aproximacion con 4 metodos que utilizan google, amazon y AP. (/poner nombres)

Despues de hacer la aproximacion Se comparan los metodos en terminos de relevancia y diversidad, para cada imagen la relevancia se marca segun la votacion de 3 voluntarios.

# resultados

Se haran pruebas de los 4 metodos sobre 300 productos. amazon+google\_top es el mejor. amazon+google\_cap es un poco peor que el otro y que google\_top porque este tiene un proceso de diversificacion y los otros no.

el metodo heuristico se va de los valores reales y el amazon+google\_ap no tiene en cuenta el ranking google.

se evalua la diversidad de lo ejemplares. mas diversidad hace que se cubran mas apariencias visuales del producto. se definen niveles de similaridad si es baja hay mucha. diversidad.amazon+googleap, heuristic y amazon+googlecap la tienen baja debido a al proceso AP. amazon+googlecap es que la tiene mas baja. aunque es un poco peor que las otras da mucha mas diversidad en los resultados.

con nuestro metodo no hay redundancia.ahora con 15 personas que controlan de busqueda de productos e imagenes se les deja hacer 300 busquedas y se comparan con los resultados del metodo amazon+googlecap y los otros. Luego se les pregunta por que metodo es mejor. como no se aclaran se aplica un test estadistico para analizar las comparaciones.

# conclusion

se propone una solucion nueva para ver productos con 2 motores a la vez. se propone utilizar esta mezcla para recoger informacion. muchas imagenes de google se seleccionan con el metodo de clustering complementando a amazon. los resultados tienen diversidad y relevancia. los experimentos demuestran que nuestra aproximacion es efectiva. en el futuro se integraran mas recursos web y mas motores de busqueda.