



Sistema Clasificador por KNN de Vida Útil para Mermeladas de Fruta

Guillermo David Luna Martínez, Luis Armando García de la Rosa.
Academia de Sistemas Computacionales, Instituto Tecnológico Superior de
Guanajuato, Carretera Guanajuato-Puentecillas Km 10.5, Puentecillas,
Guanajuato, GTO, México.
email: lgarcia@itesg.edu.mx

Resumen

En el presente proyecto, se propone un clasificador para la vida útil de mermeladas (los días que aún le quedan al alimento antes de perder propiedades organolépticas), empleando el algoritmo de aprendizaje supervisado K-Vecinos Más Cercanos o KNN (*K-Nearest Neighbors*). El vector de características para la clasificación, se obtuvo del color (en espacio CIEL*a*b*) de la muestra, puesto que los cambios en color determinan cuantos días han pasado entre la fecha de fabricación y la fecha de la medición. Por lo que se aplicó visión por computadora para obtener esta información. Se recolectaron muestras (imágenes) de mermeladas artesanales se generó la base de conocimiento (90 muestras de 10 días de fabricación distinto, lo que en el algoritmo serían las clases). Con la finalidad de tener un patrón de comparación se midió el color de todas las muestras, empleando un colorímetro comercial. Posteriormente, se entrenó el algoritmo KNN, con la base de conocimiento generada (90% de la base de conocimiento para entrenamiento, 10% para prueba), encontrándose un nivel de fiabilidad de 96% de clasificación.

Introducción

Las condiciones del mercado actual exigen que los productos sean competitivos y que presenten la máxima calidad posible. En el caso de alimentos, esta calidad se refleja en que el producto tenga ciertas características que incline la balanza del



consumidor, tal es el caso de color, olor, etc., así como la capacidad de permanecer inocuo y en óptimas condiciones para su consumo. Por lo cual se requiere de sistemas que garanticen la mencionada calidad, tal es el caso de la determinación de la vida útil.

La vida útil de un alimento se define como, el periodo de tiempo en el cual el alimento conserva los atributos esperados (y deseados) por el consumidor y se logra su máxima comercialización. En específico el color indica en gran medida la mencionada vida útil.

Actualmente diversos sistemas computacionales se han empleado en la industria de alimentos, para clasificar los productos, enfocándose en identificar características relevantes y con ello realizar el proceso de selección

Objetivos

General: Desarrollar un sistema clasificador empleando el algoritmo de K-Vecinos más cercanos, para determinar de la vida útil de mermeladas de fresa, basándose en el color en espacio CIEL*a*b* como vector de características.

Específicos:

- Recolectar imágenes de mermeladas, para procesarlas y obtener el vector de características (color en espacio CIEL*a*b*).
- Generar una base de conocimiento de muestras mermeladas, para poder aplicarlos en el entrenamiento de algoritmos de aprendizaje supervisado y no supervisado.
- Entrenar el algoritmo de KNN, para identificar las distintas clases (días después de producción).
- Comprobar la eficiencia del algoritmo con muestras desconocidas.

Justificación

Los estudios de determinación de la vida útil son fundamentales en el sector alimentario. Se recurre a ellos para lanzar un nuevo producto y para evaluar cómo afectan los cambios de procesos de producción o las reformulaciones en la estabilidad de alimentos ya consumidos. La mayor o menor vida útil del producto



depende de la naturaleza del alimento en sí, pero también de otros factores como los procesos higienizantes y de conservación a los que se someta, el envasado y las condiciones de almacenamiento, como la temperatura y la humedad. La vida útil se establece tras someter el alimento a condiciones controladas de almacenamiento o en el caso de productos muy estables (como las mermeladas, las cuales pueden mantenerse en condiciones aptas para su consumo hasta por dos años), mediante procesos de deterioro acelerado, donde los datos que se obtienen, se extrapolan después para elaborar predicciones en situaciones reales de conservación. Para la evaluación de los productos se utilizan análisis físicos, químicos y microbiológicos, así como técnicas de evaluación sensorial. Todas estas pruebas por una parte pueden llegar a ser relativamente costosas y por otra parte afectan al producto ya que son en su mayoría destructivas. Es por ello que un sistema computacional basado en visión por computadora, sería un aporte muy significativo ya que podría ofrecer una forma de calcular la vida útil, sin afectar al producto, además de que se podría realizar en campo y dependiendo de las capacidades de cómputo, se pueden obtener resultados muy rápidamente.

Metodología

Se diseñó un gabinete para la toma de las muestras, posteriormente, empleando madera MDF, se construyó el gabinete. Se usó un acrílico semitransparente como difusor de luz. En la figura 1a se muestra el prototipo ya armado. Con ayuda del gabinete, se tomaron muestras de 90 mermeladas de fresa (las cuales se obtuvieron de productores artesanales en Irapuato, Gto). Un ejemplo de esto se muestra en la figura 1b.



Figura 1. a) Prototipo del gabinete. b) Fotografía de una muestra de mermelada.



En seguida, se procedió a analizar estas muestras con un colorímetro comercial para poder tener un patrón para comparación. Enseguida, empleando Python se procesó la imagen para promediar el color de las muestras en el espacio $L^*a^*b^*$ (previamente, las muestras se preprocesaron con filtros digitales, para eliminar ruido y otros defectos de la imagen). Es importante señalar que se etiquetaron las muestras con la diferencia (en días) entre la fecha de fabricación y la fecha en que se realizó el análisis, con lo que se pudo separar las muestras en 10 clases (días) distintas.

Posteriormente se programó el clasificador KNN, igualmente en Python empleando las librerías Pandas, NumPy y Scikit-Learn. Se separaron los datos en forma aleatoria para utilizar 90% muestras el entrenamiento y 10% para la prueba. Se repitió el entrenamiento 10 veces y se promedió el desempeño.

Resultados

Como ya se mencionó, se comparó el color (promedio) de la muestra, con el que se obtiene con el colorímetro (esto para todas las muestras). Con lo cual se obtienen datos como los que se muestran en la tabla 1.

Tabla 1. Cinco muestras distintas comparadas con el patrón.

Muestra	Python			Colorímetro		
	L^*	a^*	b^*	L^*	a^*	b^*
1 (22 días)	40.85	41.93	23.84	40.98	41.88	23.98
2 (31 días)	34.44	37.66	21.59	34.54	37.44	21.87
3 (34 días)	39.94	39.75	23.69	39.88	39.98	23.76
4 (65 días)	39.69	34.69	22.81	39.76	34.76	22.98
5 (97 días)	39.63	27.87	23.90	39.66	27.98	23.98

Como se aprecia en esta tabla, los datos discrepan desde la primera cifra significativa, lo cual es tolerable (considerando que el sistema de captura es susceptible a mejoras).

Después de entrenar y probar el algoritmo KNN, se pudo comprobar que presenta un desempeño muy adecuado, ya que clasifica correctamente el 96% de las muestras usadas en la prueba. Finalmente se comprobó la capacidad

En la figura 2, se muestra en tres dimensiones la distribución espacial de todas las muestras. Por otro lado, en la figura 3, se muestra la distribución espacial de



clases (aleatorias) y tres elementos de prueba con su respectiva clasificación. En esta grafica se puede apreciar tres elementos desconocidos (con los marcadores +, × y *), siendo clasificados. Mientras que las muestras + y × se localizan muy cerca de la clase donde fueron ubicadas, la muestra *, se clasifica como perteneciente a la clase de 22 días, sin embargo se encuentra relativamente lejos de los elementos de dicha clase, sin embargo, el elemento × se clasifica como de 54 días y es evidente que queda muy lejos de la clase. Esto en realidad no es un error, ya que las características de dichos elementos se asemejan a los de las clases donde fueron ubicados. Lo que sucede es que no se cuenta con más información (clases) como para poder categorizarlos en forma más adecuada.

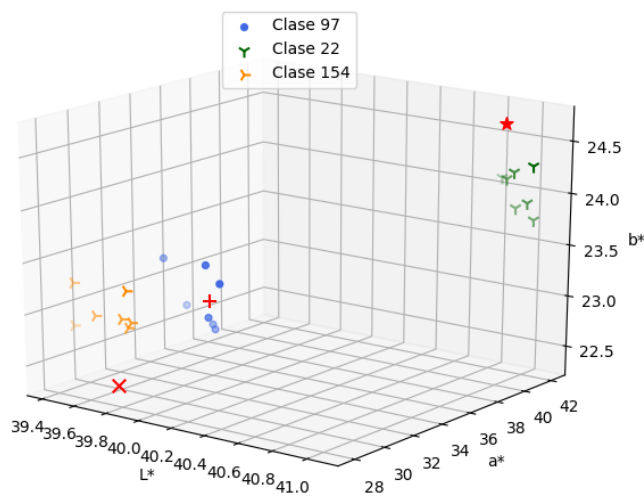
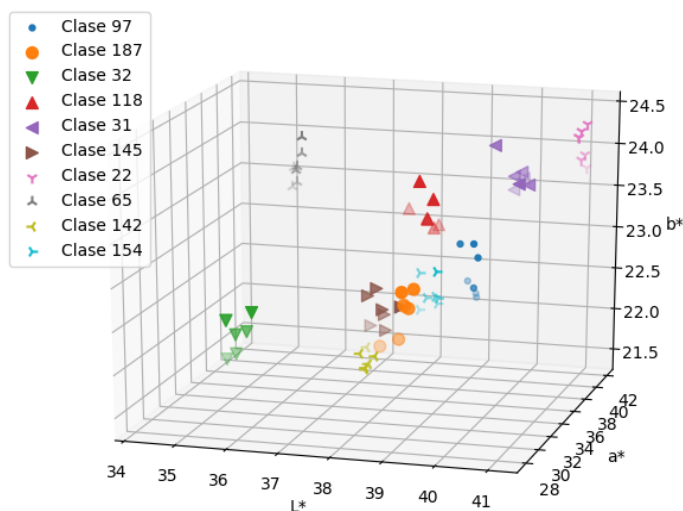




Figura 3. Distribución espacial de tres clases, con tres muestras clasificadas por KNN.

Conclusiones

Se desarrolló un sistema clasificador por medio de KNN, que es capaz de clasificar mermeladas artesanales en días después de producción, basándose en su característica de color (medido por fotografía digital). Se programó y entreno el algoritmo de aprendizaje supervisado K-Vecinos Más Cercanos o KNN. Determinándose una eficiencia del 96% de aserción, tras 10 entrenamientos de forma aleatoria. Este resultado permite suponer que se cuenta con un algoritmo de clasificación con la capacidad de determinar los días que tiene de haberse producido la mermelada de fresa y con ello poder establecer mediante una extrapolación el lapso de tiempo en donde la mermelada aún conserva su color, es decir la vida útil.

Queda como trabajo a futuro, probar otros algoritmos supervisados de clasificación, Redes Neuronales Artificiales, e inclusive no supervisados como el K-Means, pero para estos últimos, es recomendable aumentar la base de conocimientos para tener mayor diversidad de muestras apropiadamente. De igual forma, se requiere desarrollar un sistema (computacional) que en forma automática determine la vida útil de las mermeladas de fresa, empleando una imagen.

Referencias

Bhargava, A. y Bansal, A. (2018). *Fruits and vegetables quality evaluation using computer vision: A review*, *Journal of King Saud University - Computer and Information Sciences*, pp 1-31. <https://doi.org/10.1016/j.jksuci.2018.06.002>.

Carrillo-Inungaray, M. L. y Reyes-Munguía, A. (2014). *Vida útil de los alimentos*, *Revista Iberoamericana de las Ciencias Biológicas y Agropecuarias*, Vol. 2, Núm. 3, pp. 1-25. ISSN 2007-9990.

Corradini, M. G., (2018). *Shelf Life of Food Products: From Open Labeling to Real-Time Measurements*, *Annual Review of Food Science and Technology*.

Duda, R. O., Hart, P., E. Stork, D. G., (2000). *Nonparametric Techniques in Pattern Classification 2nd Edition* (pp. 161-214), New York USA, Wailey.



Mercado-Flores, J.; López-Orozco, M; Martínez-Soto G. y Abraham-Juárez, M. R. (2016). *Estimación de la Vida de Anaquel Mediante Pruebas Aceleradas en Fresa Entera en Bolsa de Polietileno y Pulpa de Fresa Congelada, Investigación y Desarrollo en Ciencia y Tecnología de Alimentos*, Vol. (1), No.1, pp. 636-640.

Mitchell, T. M., (1997). *Instance-Based Learning in Machine Learning* (pp. 230-248), McGraw-Hill Science/Engineering/Math, USA.

Montoya-Holguin, C.; Cortés-Osorio, J. A. y Chaves-Osorio, J. A., (2014). *Sistema automático de reconocimiento de frutas basado en visión por computador, Ingeniare*. Revista chilena de ingeniería, Vol. (22), No. 4, pp. 504-516.

Labuza, T. P., (1982). *Chapter 1: Shelf Life in Shelf-Life Dating of Foods, Food and Nutrition Press*. USA.