

## Alcohol detection in “lager” beers using NIR

D. Brais Castiñeiras Galdo

Actualmente existe una gran rivalidad en cuanto a las técnicas de medición destructivas y a las técnicas de medición no destructivas. Mientras unas optan por la destrucción de la muestra para que el resultado de dicho proceso pueda ser analizado por un experto, las otras optan por un análisis empírico y repetible, que otorga la posibilidad de realizar numerosos tipos de análisis sobre una misma muestra.

En cuanto a las técnicas de medición no destructiva destaca la espectroscopia de infrarrojo cercano, técnica que se encarga de analizar el espectro electromagnético en busca de los enlaces de carbono pertenecientes a la sustancia a examinar. Ésta recibe ese nombre debido a que la región del espectro que analiza se sitúa muy próxima a la luz visible, concretamente entre los 700 nm y los 2500 nm (1).

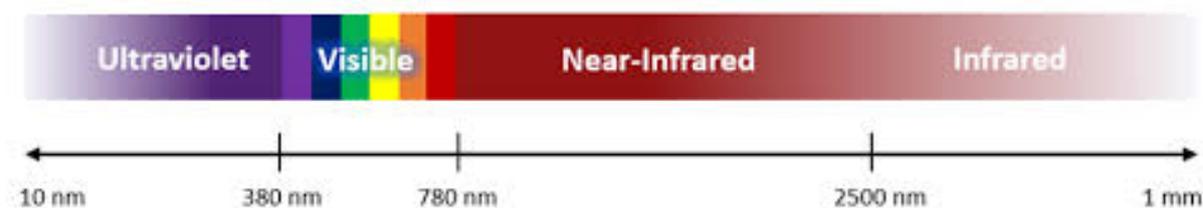


Figura 1: Regiones de análisis de la espectroscopia de infrarrojo cercano

Los dispositivos usados para llevar a cabo dichas mediciones son los espectrofotómetros. Estos dispositivos tienen un tamaño y coste considerables, por lo que su uso está reducido a localizaciones como los laboratorios.

Para solucionar este problema, se han creado los espectrofotómetros portátiles, que tienen un tamaño y coste reducido. El mayor inconveniente es que el software para lidiar con ellos es inexistente o está desactualizado.

El objetivo del proyecto es la creación de una aplicación para dispositivos móviles que permita la comunicación con un espectrofotómetro portátil. A modo de objetivos secundarios se han establecido los siguientes:

1. Posibilidad de almacenar las mediciones realizadas.
2. Adaptación a distintos dispositivos: móviles y tabletas.
3. Integración con técnicas de aprendizaje máquina para poder realizar estimaciones acerca de los datos en tiempo real.

#### 4. Disponibilidad en varios idiomas para alcanzar un mayor número de usuarios.

Se ha realizado un estudio comparativo acerca de las posibles soluciones, cuyo resumen se puede ver en la tabla 1. Se ha optado por la solución de *Texas Instruments: NIRScan NANO EVM* debido a su relación calidad precio y a la propiedad completa de los datos, punto que se considera de vital importancia dado que, de este modo, se tiene un control absoluto sobre los mismos.

Comparativa de todas las soluciones					
Característica	WiSci	Pasco	Scio	Changhon	NIRscan
Calidad	Media	Buena	Muy buena	Mala	Excelente
Proporción de un API	Sí	Sí	Sí	No	Sí
Documentación	Parcial o incompleta	Parcial o incompleta	Parcial o incompleta	-	Completa
Soporte técnico	Sí	Sí	Sí	Sí	Sí
Propiedad de los datos	Sí	Sí	No	-	Sí
Compatibilidad con Android	Sí	Sí	Sí	Sí	Sí
Conexiones	USB, Bluetooth	USB, Bluetooth	USB, Bluetooth	Integrado	USB, Bluetooth
Peso	48 g	57 g	35 g	-	84 g
Precio	\$250,00	\$399,00 + \$69,00	\$950,00	\$250,00	\$999,00

Tabla 1: Comparativa de las soluciones

Se han detectado 4 grandes módulos a la hora de llevar a cabo la toma de requisitos del sistema a desarrollar:

1. Comunicación con el espectrofotómetro. Es el módulo central del sistema. Se encarga de gestionar toda la conexión del teléfono móvil-espectrofotómetro. Incluye tareas como gestionar la conexión o cambiar la configuración del dispositivo además de realizar mediciones. En este momento se decide que el sistema operativo a usar es Android.
2. Almacenamiento de las mediciones. Es el encargado de almacenar las mediciones realizadas con la aplicación. Incluye la gestión de problemas, que son una forma de agrupar las mediciones en conjuntos bajo unas mismas etiquetas. Esto otorga la posibilidad de crear una base de datos para realizar tareas de aprendizaje máquina haciendo uso de los datos recogidos.
3. Integración de técnicas de aprendizaje máquina. Realiza estimaciones acerca de las muestras recogidas. Puede ser en tiempo real, con una medición recién tomada o haciendo uso de los datos almacenados previamente en el teléfono móvil.
4. Internacionalización. Se encarga de gestionar el idioma de la aplicación. Se han considerado los idiomas de inglés y español debido a su gran número de hablantes.

A continuación se realiza una demo...

Se ha llevado a cabo una prueba práctica para comprobar el correcto funcionamiento del sistema. Esta estaba situada en el ámbito de las cervezas. Consistía en reconocer si una cerveza de tipo "lager" tenía alcohol o no a partir de una muestra.

Se ha llevado a cabo un estudio del arte en el que se han detectado numerosos estudios acerca del tema. La gran diferencia es que se ha hecho uso de un espectrofotómetro de laboratorio, lo que otorga mucha más precisión y alcanza un mayor rango de medición que el dispositivo usado para llevar a cabo este experimento.

Aún con lo mencionado anteriormente, es de destacar que el alcohol se encuentra presente en los rangos comprendidos entre los 1100 nm y 1300 nm, y entre los 2110 nm y 2400 nm. El rango de medición del dispositivo del que se dispone se sitúa entre los 900 nm y 1700 nm, por lo que solo se puede acceder al primer rango.

Para establecer el marco metodológico se ha contado con la ayuda de un experto en espectrofotometría de la Universidade da Coruña. Éste ha otorgado unas pautas acerca de cómo llevar a cabo las mediciones. Entre estas se pueden destacar:

1. Mismas condiciones de temperatura. La tecnología NIR se ve muy afectada por este parámetro.
2. Adquisición de celdas de espectrofotometría. Para llevar a cabo las mediciones de una forma estándar.
3. Mismo tiempo en el exterior de la botella. Para tratar de extraer el gas en las mejores condiciones.
4. Distancia de realización de la medición menor que 2 mm. Para que exista el menor ruido posible en la realización de las mismas.
5. Mismo ángulo de medición. Al igual que sucede con la temperatura, este parámetro influye de una forma sustancial en la tecnología NIR.
6. Uso de filtros para extraer la mayor cantidad posible previa a la realización de la muestra.

El problema generado en la aplicación contenía el nombre de la cerveza por medición y como etiquetas se establecían la graduación de la cerveza en cuestión y un identificador de la misma. De este modo se pueden reutilizar los datos para realizar otro tipo de problemas relacionados con la graduación de la misma.

En cuanto a la configuración del espectrofotómetro destaca que tiene 2 tipos:

1. Column. Es de tipo lineal.
2. Hadamard. Está basado en correlaciones de matrices. Este es mejor en el 99% de los casos.

Para este problema se ha usado la configuración de tipo Hadamard y se ha repetido la medición 30 veces para evitar la mayor cantidad de ruido posible. Además, se ha usado todo el rango de acción del dispositivo.

Se ha creado un soporte casero para situar la muestra y el espectrofotómetro siempre en la misma posición y se ha tapado la celda con cartulina negra para simular una celda de luz con el objetivo de evitar que la radiación se escapara de la misma.

En cuanto a la base de datos generada, se han tenido en cuenta 36 cervezas de distintas graduaciones y marcas. Tan solo se han usado cervezas “puras”, es decir, cervezas que carecían de sustancias como limón o tequila, ya que la composición de las mismas podría afectar a la medición. A mayores, se han realizado 3 mediciones por cerveza para hacer la base de datos más robusta.

En referencia al tratamiento de las mediciones, se han usado redes de neuronas y *TensorFlow Lite*, esto es debido a que el sistema desarrollado está integrado con *TensorFlow Lite* y, de este modo, se puede generar el modelo para realizar estimaciones en tiempo real.

Debido al tamaño de la base de datos, se ha decidido que el número de capas ocultas de las redes de neuronas sea una. Se han realizado 50 repeticiones del entrenamiento. El criterio de parada ha sido que la precisión supere en 90 % y no exista mejora en 5 epochs o que el número de epochs sea igual a 200. Como técnica de validación se ha usado la técnica de *K-fold Record-wise cross-validation*, que permite asilar las 3 mediciones de una misma cerveza para realizar validación con las mismas sin que influyan en los datos utilizados en la fase de entrenamiento. Algunos de los resultados obtenidos se pueden apreciar en la tabla 2.

Identificador	Neuronas en la capa 1	Función de activación de la capa 1	Precisión entrenamiento	Desviación típica en entrenamiento	Precisión en test	Desviación típica en test
0	4	Sigmoid	0.864	0.145	0.68	0.455
1	32	Sigmoid	0.917	0.047	0.814	0.388
2	16	Sigmoid	0.89	0.059	0.793	0.400
3	64	Sigmoid	0.910	0.043	0.807	0.390
4	128	Sigmoid	0.912	0.041	0.81	0.386
5	4	Relu	0.744	0.195	0.535	0.495
6	32	Relu	0.901	0.036	0.806	0.389
7	16	Relu	0.902	0.061	0.795	0.400
8	64	Relu	0.908	0.030	0.817	0.386
9	128	Relu	0.913	0.030	0.812	0.389
10	8	Relu	0.852	0.157	0.710	0.451

Tabla 2: Redes de neuronas entrenadas

En un futuro se tratará de mejorar estos resultados y crear una base de datos con un tamaño mucho mayor para tener representación del mayor número de patrones posible.

Existen una infinidad de aplicaciones posibles haciendo uso de esta tecnología. A continuación se mostrarán algunas de las muchas posibilidades que ofrece:

1. Sector alimentario. Existen numerosas oportunidades en este sector. Además está presente en el día a día de todas las personas, por lo que se considera un sector muy atractivo para llevar a cabo algún estudio relacionado con el mismo. Algunas de las posibilidades que puede ofrecer son:
  - a) Búsqueda de elementos nocivos en agua.
  - b) Búsqueda de bacterias en lácteos.
  - c) Estimación del porcentaje de cacao en chocolate.
  - d) Búsqueda de setas venenosas.
  - e) Conocer el estado actual de las frutas.
  - f) Búsqueda de las propiedades del vino.
2. Sector de la metalurgia. El análisis de los elementos propios del sector puede otorgar muchas facilidades a la hora de realizar el trabajo. Se podría detectar fácilmente las sustancias presentes en alguna estructura, lo que otorgaría facilidades a la hora de detectar las propiedades de la misma. A modo de ejemplo, algunas de las oportunidades pueden ser:
  - a) Detección de la pureza de las joyas.
  - b) Detección del tipo de metal.
3. Análisis ambiental. Otro gran sector en el que realizar tareas de investigación es en el análisis ambiental. El análisis del entorno es muy importante para poder evitar la polución o las sustancias nocivas que pueden provocar grandes pérdidas. Algunas de las posibilidades pueden ser:
  - a) Concentración de ciertas composiciones en el aire.
  - b) Detección de sustancias nocivas en el aire.
  - c) Detección de algunas concentraciones en distintos gases.
  - d) Detección de sustancias en el suelo.
4. Aplicaciones específicas en medicina. Este es un sector muy relevante que puede facilitar considerablemente la toma de decisiones de los médicos. En base a un análisis acerca de un órgano de un paciente se podrían detectar algunas de las enfermedades que pueda padecer, lo que hace que este sector sea muy relevante en cuanto a las posibles aplicaciones de este sistema. Algunas de las oportunidades pueden ser:
  - a) Detección de células cancerígenas.
  - b) Detección de pólipos.
5. Otros. Existen otros grandes sectores en los que este sistema sería de gran utilidad. Algunas de las posibles aplicaciones pueden ser:
  - a) Detección de fugas de gas.
  - b) Análisis e identificación de pigmentos en pinturas.
  - c) Supervisión de las concentraciones de  $CO_2$  en invernaderos.
  - d) Análisis forense: detección de alcohol en sangre.
  - e) Geología: detección de diversos materiales en rocas.