

Generación de Datos Sintéticos para Machine Learning en RMN-Hahn

Matias Nuñez, Analia Zwick

22 de septiembre de 2023

Objetivo

Generar datos sintéticos para entrenar un modelo de aprendizaje automático, utilizando el contexto de la Resonancia Magnética Nuclear con secuencia de control Hahn (RMN-Hahn). Visualización con ML sin supervisar.

Contexto

En RMN, la señal total $S(t)$ está relacionada con la distribución de tamaños de compartimentos $P(l)$ y la magnetización proveniente de los espines confinados en cada compartimento $M_l(t)$ (ver Referencia [1]). La relación es:

$$S(t) = \sum_l P(l)M_l(t) \quad (1)$$

El objetivo es generar datos para un modelo de ML que relacione S con P .

Instrucciones

1. **Generación de la Distribución de Tamaños de Compartimentos $P(l)$:**

- Decida cuántos compartimentos l desea considerar. Esto debería ser un parámetro de la función de generación.
- La distribución $P(l)$ tiene la forma de una función log-normal [1]

$$P(l) = \frac{1}{l \ln(\sigma) \sqrt{2\pi}} \exp \left(-\frac{(\ln(l) - \ln(l_c))^2}{2\ln(\sigma)^2} \right). \quad (2)$$

donde l_c es el tamaño medio del compartimento y σ es el ancho de la distribución de probabilidad con respecto a su media.

2. Generación de la Función $M_l(x)$:

Consideremos la secuencia de control de RMN "Hahn", $M_l = M_{l,\text{Hahn}}$ con

$$M_{l,\text{Hahn}}(t) = \exp \left\{ -\gamma^2 G^2 D_0 \tau_c^2 \left[t - \tau_c \left(3 + e^{-\frac{t}{\tau_c}} - 4e^{-\frac{t}{2\tau_c}} \right) \right] \right\}. \quad (3)$$

donde $\tau_c = l_c/(2D_0)$ está asociado a la longitud de correlación l_c que define el tamaño de la cavidad que restringe el proceso de difusión que experimentan los espines. γ es el factor giromagnético del spin nuclear del proton, G el gradiente externo aplicado y D_0 el coeficiente de difusión.

- Testear que la Magnetización esta bien simulada comparando con Fig. 2(a) del paper [2] que muestra el comportamiento de M vs $(\gamma^2 G^2 D_0)^{1/3} t$ en el rango $[1, 100]$. La figura muestra varias curvas, cada una esta asociada a un tamaño distinto de cavidad l_c determinado por el tiempo de correlación τ_c dado en el caption.
- Para cada valor de t en este rango, genere valores para $M_{l,\text{Hahn}}(t)$.

3. Cálculo de $S_{\text{Hahn}}(t)$:

- Utilice la Eq. (1) para calcular $S_{\text{Hahn}}(t)$ para cada valor de t en su rango.
 - Para $P_l(t)$ de Eq. (2) considerar un único valor $l_c = 3,7\mu m$ y generar varias distribuciones posibles variando σ .
 - Para $M_{l,\text{Hahn}}$ considerar un único valor de τ_c correspondiente a $l_c = 3,7\mu m$ y $D_0 = 10^{-5} cm^2/s$ (ojo con el cambio de unidades).

4. Creación del Conjunto de Datos:

- Ahora tiene un conjunto de pares de datos de entrada y salida. Cada valor de t y su correspondiente $S_{\text{Hahn}}(t)$ es una entrada, y el vector $P(l)$ es la salida correspondiente. O puede verse en forma inversa también. La idea será encontrar un mapeo que relacione ambos conjuntos de datos.

Visualización de los datos generados:

- Investigue sobre métodos de reducción de dimensionalidad tipo PCA, t-SNE o UMAP (Machine learning no supervisado) para poder visualizar los datos generados en 2D. Ver Ref. [3]

Desafío Adicional:

- Introduzca algún ruido en sus datos para simular las imperfecciones reales de los experimentos.
- Varíe el número de compartimentos y vea cómo afecta la complejidad de sus datos, esto puede verse en los mapas generados con PCA/tsne/UMAP.

Preguntas

1. ¿Cómo afecta el ruido introducido a la distribución de $S_{\text{Hahn}}(x)$?
2. ¿Cómo espera que se comporten diferentes modelos de aprendizaje automático con este conjunto de datos?

Referencias

- [1] Shemesh, N., Álvarez, G. A., & Frydman, L. Size Distribution Imaging by Non-Uniform Oscillating-Gradient Spin Echo (NOGSE) MRI. PLoS ONE 10(7):e0133201 (2015). doi:10.1371/journal.pone.0133201
- [2] Zwick, A., Suter, D., Kurizki, G. & Álvarez, G. A.. Precision limits of tissue microstructure characterization by Magnetic Resonance Imaging, Phys. Rev. Applied 14, 024088 (2020) doi.org/10.1103/PhysRevApplied.14.024088

- [3] Matias Nuñez, Exploring materials band structure space with unsupervised machine learning, Computational Materials Science, Volume 158, 2019, Pages 117-123, ISSN 0927-0256, <https://doi.org/10.1016/j.commatsci.2018.11.002>. (<https://www.sciencedirect.com/science/article/pii/S092702561830716X>)