

Indexación espacial con R-tree sobre la base de datos de la Encuesta Nacional Agropecuaria 2024 del Perú

Autor: ILMA MAGDA MAMANI MAMANI

Escuela Profesional de Ingeniería Estadística e Informática, Universidad Nacional del Altiplano de Puno

September 30, 2025

1 Introducción y motivación

El presente trabajo aborda la **indexación espacial mediante estructuras R-tree**, aplicada a los microdatos de la Encuesta Nacional Agropecuaria (ENA) 2024 del Perú. Este tema resulta relevante porque los sistemas estadísticos enfrentan dificultades para procesar y consultar eficientemente grandes volúmenes de información georreferenciada de productores, cultivos y parcelas [de Estadística e Informática, INEI,I].

La pertinencia de la investigación radica en su impacto en el desarrollo rural, dado que un acceso ágil a la información agrícola favorece políticas públicas y la innovación tecnológica [Mori et al., 2022, San Emeterio de la Parte et al., 2023]. Además, la indexación R-tree ofrece ventajas sobre métodos tradicionales como B-tree y Quad-tree en consultas multidimensionales [Guttman, 1984, Mao et al., 2023, Kim et al., 2024].

Por tanto, se busca demostrar cómo la implementación de un índice R-tree en RStudio mejora los tiempos de consulta y la visualización de patrones espaciales en datos agrícolas nacionales [Gu et al., 2021, Xia et al., 2020, Gao et al., 2023].

2 Planteamiento del problema

La gestión de datos agropecuarios en el Perú ha sido predominantemente descriptiva, centrada en informes y tabulados, lo cual limita el análisis espacial [de Estadística e Informática, INEI,I]. Estas limitaciones dificultan operaciones como la detección de superposiciones de cultivos o la identificación de regiones con patrones productivos específicos [Sandonís-Pozo et al., 2022, Colaço et al., 2019a].

Si bien se han explorado técnicas como Quad-tree o Grid Index, estas muestran baja eficiencia en escenarios con datos masivos y multidimensionales [Mao et al., 2023, Zhou et al., 2017]. Este vacío justifica el uso de R-tree, cuya estructura basada en rectángulos mínimos optimiza búsquedas espaciales complejas [Guttman, 1984, Xia et al., 2020].

El caso de la ENA 2024 ofrece una oportunidad única: la disponibilidad de coordenadas, superficies de parcelas y atributos productivos que permiten implementar y evaluar consultas espaciales mediante R-tree [de Estadística e Informática, INEI, Mori et al., 2022].

3 Objetivos y preguntas de investigación

Objetivo general

Evaluar la eficiencia de la indexación espacial mediante R-tree en la ENA 2024, utilizando RStudio como entorno de desarrollo.

Objetivos específicos

- Preprocesar variables espaciales y productivas relevantes de la ENA 2024.
- Implementar un índice R-tree en RStudio usando librerías `sf` y `rgeos`.
- Medir el desempeño en consultas espaciales (rango, proximidad, superposición).
- Generar mapas temáticos interactivos de los patrones detectados.

Preguntas

- ¿Cómo mejora el R-tree los tiempos de consulta frente a métodos convencionales?
- ¿Qué patrones espaciales emergen de la ENA 2024 al aplicar consultas R-tree?

4 Metodología propuesta

La investigación adopta un enfoque **cuantitativo y experimental** [Kim et al., 2024, Gu et al., 2021].

Datos

Se emplearán los microdatos de la ENA 2024, que incluyen información georreferenciada y productiva a nivel de unidades agropecuarias, parcelas y cultivos [de Estadística e Informática, INEI]. De acuerdo

con el diccionario de datos, se priorizan variables como:

- **Ubicación:** LATITUD, LONGITUD, CCDD, CCPP, CCDI, NSEGM.
- **Identificación:** ID_PROD, UA.
- **Parcelas:** P105, P105_SUP_ha, P104_SUP_ha, P102 (número de parcelas).
- **Cultivos:** P115_COD, P115_TIPO, P117_SUP_ha.
- **Producción y riesgos:** P203, P210_SUP_ha, P224B (pérdidas).

Estas variables permiten construir geometrías espaciales y definir rectángulos mínimos para el R-tree.

Preprocesamiento

Incluye limpieza de inconsistencias, estandarización de coordenadas y validación topológica [Gao et al., 2023, Mori et al., 2022]. Para parcelas sin coordenadas, se aproximarán posiciones con el segmento NSEGM y superficies reportadas.

Implementación en R

Se utilizarán librerías `sf`, `sp` y `rgeos` para construir el índice R-tree. Se evaluará su desempeño en operaciones de búsqueda de rango, proximidad y solapamiento [Espinel et al., 2024, San Emeterio de la Parte et al., 2023, Guttman, 1984].

Evaluación

Se medirán métricas de tiempo de respuesta y accesos a disco en consultas estratificadas por regiones naturales (Costa, Sierra, Selva) [Mao et al., 2023, Colaço et al., 2019b].

5 Impacto y conclusión

El proyecto contribuirá a modernizar la infraestructura de datos agrícolas en el Perú, incrementando la eficiencia de análisis en encuestas nacionales [?]. Académicamente, llenará un vacío en la literatura nacional sobre uso de R-tree en información agropecuaria [San Emeterio de la Parte et al., 2023, Espinel et al., 2024]. En la práctica, permitirá que el INEI y el MIDAGRI optimicen la explotación de sus bases, facilitando evidencia para políticas públicas [de Estadística e Informática, INEI,I].

En conclusión, la implementación de R-tree en RStudio representa un paso inicial hacia el uso de técnicas avanzadas de gestión espacial en el ámbito estadístico nacional, con potencial de replicarse en otros sistemas de información y encuestas masivas [Gu et al., 2021, Zhou et al., 2017].

References

- A. F. Colaço et al. Spatial variability in commercial orange groves. part 1. *Precision Agriculture*, 2019a. doi: 10.1007/s11119-018-9612-3. URL <https://link.springer.com/article/10.1007/s11119-018-9612-3>.
- A. F. Colaço et al. Spatial variability in commercial orange groves. part 2. *Precision Agriculture*, 2019b. doi: 10.1007/s11119-018-9615-0. URL <https://link.springer.com/article/10.1007/s11119-018-9615-0>.
- Instituto Nacional de Estadística e Informática (INEI). Encuesta nacional agropecuaria 2024 – microdatos, 2024a. URL <https://datosabiertos.gob.pe/dataset/encuesta-nacional-agropecuaria-ena-2024a>.
- Instituto Nacional de Estadística e Informática (INEI). Encuesta nacional agropecuaria 2024 – principales resultados, 2024b. URL <https://www.gob.pe/institucion/inei/informes-publicaciones/6879473-productores-agropecuarios-principales>.
- J. Espinel et al. Artificial intelligence in agricultural mapping: A review. *Agriculture*, 14(7):1071, 2024. doi: 10.3390/agriculture14071071. URL <https://www.mdpi.com/2077-0472/14/7/1071>.
- Y. Gao et al. Research on efficient indexing of large-scale geospatial data based on multi-level geographic grid. *ISPRS Annals*, 2023. doi: 10.5194/isprs-annals-X-1-W1-2023-73-2023. URL <https://isprs-annals.copernicus.org/articles/X-1-W1-2023/73/2023/>.
- T. Gu et al. A reinforcement learning based r-tree for spatial data indexing in dynamic environments. *arXiv preprint*, 2021. URL <https://arxiv.org/abs/2103.04541>.
- Antonin Guttman. R-trees: A dynamic index structure for spatial searching. *Proceedings of the 1984 ACM SIGMOD Conference*, pages 47–57, 1984. doi: 10.1145/971697.602266. URL <https://dl.acm.org/doi/10.1145/971697.602266>.
- J. Kim et al. Sgir-tree: Integrating r-tree spatial indexing as subgraph for graph dbmss. *ISPRS International Journal of Geo-Information*, 13(10):346, 2024. doi: 10.3390/ijgi13100346. URL <https://www.mdpi.com/2220-9964/13/10/346>.
- Q. Mao, M. A. Qader, and V. Hristidis. Comparison of lsm indexing techniques for

- storing spatial data. *Journal of Big Data*, 2023. doi: 10.1186/s40537-023-00734-3. URL <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00734-3>.
- G. Meza Mori et al. Spatial analysis of environmentally sensitive areas in amazonas, peru. *Sustainability*, 14(22):14866, 2022. doi: 10.3390/su142214866. URL <https://www.mdpi.com/2071-1050/14/22/14866>.
- I. San Emeterio de la Parte et al. Big data and precision agriculture: a novel spatio-temporal semantic iot data management framework. *Journal of Big Data*, 2023. doi: 10.1186/s40537-023-00729-0. URL <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00729-0>.
- L. Sandonís-Pozo et al. Satellite multispectral indices to estimate canopy parameters. *Precision Agriculture*, 2022. doi: 10.1007/s11119-022-09956-6. URL <https://link.springer.com/article/10.1007/s11119-022-09956-6>.
- J. Xia et al. Dapr-tree: a distributed spatial data indexing scheme with r-tree. *International Journal of Digital Earth*, 2020. doi: 10.1080/17538947.2020.1778804. URL <https://www.tandfonline.com/doi/full/10.1080/17538947.2020.1778804>.
- Y. Zhou et al. Spatial indexing for data searching in mobile sensing environments. *Sensors*, 17(6):1427, 2017. doi: 10.3390/s17061427. URL <https://www.mdpi.com/1424-8220/17/6/1427>.