

Clustering basado en opsonización

Juan Esteban Rincón Marín^{1*}

^{1*}Universidad Politécnica de Madrid.

Abstract

En este trabajo se presenta un algoritmo de clustering descentralizado inspirado en el proceso de opsonización, implementado mediante un sistema multiagente que opera sobre un grid bidimensional. El enfoque separa explícitamente la toma de decisiones semánticas de la acción física, permitiendo la formación de clústeres a partir de interacciones locales y coordinación estigmérica. El algoritmo se evalúa sobre conjuntos de datos estándar de clustering y se compara con k-means utilizando métricas externas e internas de validación.

Keywords: Clustering, sistemas multiagente, inteligencia de enjambre, estigmérica, aprendizaje no supervisado

1 Introducción

El *clustering* no supervisado tiene como objetivo identificar estructuras latentes en los datos sin conocimiento previo de las etiquetas. Entre los métodos clásicos, *k-means* es ampliamente utilizado por su simplicidad, aunque presenta limitaciones en escenarios exploratorios debido a la necesidad de fijar el número de clústeres y a su sensibilidad a la inicialización [1].

Como alternativa, los algoritmos bioinspirados basados en inteligencia de enjambre explotan mecanismos de auto-organización mediante interacciones locales, permitiendo que la estructura global emerja de forma distribuida. Sin embargo, estos enfoques suelen presentar problemas de estabilidad y convergencia [2].

En este trabajo se propone un algoritmo de *clustering* inspirado en interacciones locales y en el concepto de *opsonización* [3]. El enfoque adopta una formulación espacial bidimensional y se evalúa en un contexto de *clustering* tradicional mediante comparación con *k-means* [4] utilizando métricas estándar.

Los métodos clásicos de clustering, como k-means, constituyen el principal punto de referencia en la literatura, mientras que los enfoques bioinspirados, en particular

los algoritmos de ant-based clustering, han sido propuestos como alternativas auto-organizadas con distintos problemas de estabilidad y convergencia [1, 2, 4].

2 Definición del problema y preprocesado

El problema abordado consiste en el *clustering* no supervisado de un conjunto de datos $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$, donde cada muestra $x_i \in \mathbb{R}^d$ está descrita por un vector de características numéricas. El objetivo es obtener una partición de los datos en grupos coherentes sin utilizar información de etiquetas durante el proceso de agrupamiento.

Para la evaluación experimental se emplean los conjuntos de datos *Iris* y *Wine*, ambos de uso estándar en la literatura de *clustering*. Las etiquetas reales se utilizan únicamente para la evaluación posterior mediante métricas externas.

Como preprocesado, todas las características se normalizan mediante estandarización (*z-score*) para evitar sesgos debidos a diferencias de escala en el cálculo de distancias. Adicionalmente, las muestras se proyectan sobre un espacio bidimensional discreto \mathcal{G} que actúa como entorno operativo del sistema. Esta proyección no persigue reducir dimensionalidad, sino proporcionar un soporte espacial para definir operaciones locales de interacción y reorganización.

3 Algoritmo propuesto

3.1 Visión general

Se propone *OIC-Grid*, un algoritmo de clustering descentralizado inspirado en el concepto de opsonización. Los datos se organizan sobre un grid bidimensional y se procesan mediante interacciones locales entre agentes *Marker*, encargados de la asignación semántica, y agentes *Transporter*, responsables de la reorganización espacial. Los clústeres se representan mediante *beacons* que almacenan prototipos dinámicos y permiten coordinación estigmérgica sin control global.

3.2 Representación de clústeres y similitud

Cada clúster k se representa mediante un *beacon* que mantiene un prototipo $\mu_k \in \mathbb{R}^d$, actualizado de forma incremental a partir de los objetos asignados. La similitud entre un objeto normalizado \tilde{x}_i y un clúster se evalúa mediante distancia euclídea. Un objeto se considera compatible con un clúster si

$$\|\tilde{x}_i - \mu_k\| \leq \tau,$$

donde τ es un umbral definido como percentil de las distancias observadas, lo que permite una adaptación relativa a la escala de los datos.

3.3 Regla de decisión de los agentes *Marker*

Los agentes *Marker* recorren el grid mediante movimientos aleatorios y aplican reglas de decisión locales al encontrar un objeto no etiquetado. Si no hay *beacons* visibles, se crea un nuevo clúster. En caso contrario, se selecciona el *beacon* más similar al objeto,

Algorithm 1 OIC-Grid (algoritmo de clustering propuesto)

Require: $\tilde{\mathcal{X}}$, parameters (τ, K_{local}) , agents, iterations T

```
1: Initialize grid and empty beacons  $\mathcal{B}$ 
2: for  $t = 1$  to  $T$  do
3:   for all Marker agents do
4:     Select nearby unlabelled object  $i$ 
5:     if no beacon visible then
6:       Create beacon and assign  $i$ 
7:     else
8:        $k^* \leftarrow$  most similar visible beacon
9:       if  $\|\tilde{x}_i - \mu_{k^*}\| \leq \tau$  or  $|\mathcal{B}_{\text{vis}}| = K_{\text{local}}$  then
10:        Assign  $i$  to  $k^*$  and update  $\mu_{k^*}$ 
11:       else
12:        Create beacon and assign  $i$ 
13:       end if
14:     end if
15:   end for
16:   for all Transporter agents do
17:     Relocate labelled objects toward their beacons
18:   end for
19: end for
20: return  $\ell(i)$ 
```

definido como aquel cuyo prototipo minimiza $\|\tilde{x}_i - \mu_k\|$. Si la distancia es menor o igual que el umbral τ , el objeto se asigna al clúster correspondiente; si es mayor y la capacidad local K_{local} lo permite, se crea un nuevo clúster; en caso contrario, el objeto se asigna al *beacon* más similar disponible. Esta regla garantiza progreso continuo y evita bloqueos en un proceso distribuido y asíncrono.

3.4 Regla de acción de los agentes *Transporter*

Los agentes *Transporter* se desplazan igualmente de forma aleatoria por el espacio y no toman decisiones semánticas. Su función es recoger objetos ya etiquetados y desplazarlos hacia la vecindad del *beacon* correspondiente, favoreciendo la consolidación espacial de los clústeres. Esta separación explícita entre decisión y acción distingue el enfoque propuesto de otros métodos de *ant-based clustering* y contribuye a la estabilidad del proceso.

3.5 Descripción del algoritmo

4 Evaluación experimental

La evaluación experimental tiene como objetivo comparar el algoritmo propuesto *OIC-Grid* con *k-means*, utilizado como método de referencia en clustering particional [4]. Dado que *OIC-Grid* incorpora hiperparámetros que influyen en su dinámica local,

Table 1 Resultados experimentales. Media y desviación estándar (std) para OIC-Grid y k-means.

2-9 Método / Dataset	Media				Std			
	ARI	NMI	Sil	DB	ARI	NMI	Sil	DB
OIC-Grid (Iris)	0.2815	0.4534	-0.1402	3.6486	0.0509	0.0327	0.0488	0.5467
k-means (Iris)	0.6201	0.6595	0.4599	0.8336	0.0000	0.0000	0.0000	0.0000
OIC-Grid (Wine)	0.2970	0.4886	-0.5146	13.9879	0.0472	0.0350	0.1285	2.1560
k-means (Wine)	0.8975	0.8759	0.2849	1.3886	0.0000	0.0000	0.0000	0.0000

se llevó a cabo una búsqueda en rejilla (*grid search*) sobre un conjunto discreto de configuraciones, seleccionando aquellas que maximizan las métricas externas de calidad de clustering.

Dado que *OIC-Grid* incorpora hiperparámetros que influyen en su dinámica local, se llevó a cabo una búsqueda en rejilla (*grid search*) sobre un conjunto discreto de configuraciones, seleccionando aquellas que maximizan las métricas externas de calidad de clustering.

La comparación entre ambos métodos se realiza utilizando exclusivamente métricas finales de calidad de clustering. En concreto, se emplean métricas externas (Adjusted Rand Index, ARI, y Normalized Mutual Information, NMI) y métricas internas (Silhouette Score y Davies Bouldin Index), ampliamente utilizadas en estudios comparativos de clustering [1, 5]. Por limitaciones de espacio, no se consideran métricas dinámicas u online.

5 Resultados

Los resultados muestran que *k-means* obtiene un rendimiento superior en ambos datasets, especialmente en las métricas externas. Además, presenta una desviación prácticamente nula entre ejecuciones, lo que refleja una convergencia estable al óptimo tras la estandarización de los datos. Este comportamiento es coherente con la naturaleza de cada enfoque: *OIC-Grid* prioriza un proceso de clustering distribuido, local y estigmérico, sin optimización explícita de una función global, mientras que *k-means* explota directamente la estructura geométrica del espacio de características mediante asignación centralizada a prototipos globales. Estos resultados permiten caracterizar con precisión las diferencias entre ambos enfoques dentro de un marco de clustering estándar.

6 Declaración sobre el uso de herramientas de Inteligencia Artificial

Durante el desarrollo de este trabajo se emplearon herramientas de inteligencia artificial basadas en modelos de lenguaje como apoyo a la redacción y estructuración del manuscrito, así como a la clarificación de explicaciones técnicas y la refactorización asistida de fragmentos de código.

Las herramientas de inteligencia artificial no se utilizaron para la generación de datos, la ejecución de experimentos ni la obtención de resultados. El diseño

metodológico, la implementación del algoritmo, la validación y la interpretación de los resultados fueron realizados por el autor.

7 Conclusiones

En este trabajo se ha presentado OIC-Grid, un algoritmo de clustering no supervisado basado en interacciones locales y coordinación estigmérgica, inspirado en el concepto de opsonización. Los resultados experimentales muestran que el método obtiene un rendimiento inferior a k-means en conjuntos de datos clásicos como Iris y Wine, especialmente en métricas externas, lo que indica que en problemas estáticos y geométricos la reorganización espacial mediante agentes *Transporter* no aporta una mejora significativa. No obstante, el enfoque resulta relevante como estudio de clustering emergente y puede ser de interés en escenarios distribuidos o físicos, como sistemas robóticos o multiagente.

Appendix A Repositorio y materiales adicionales

Se proporciona un repositorio público con la implementación del algoritmo, los resultados experimentales completos y una descripción detallada del método y su configuración experimental. Repositorio: https://github.com/Esk1z0/multiagents_opsonization_inspired_clustering.git

References

- [1] Jain, A.K., Murty, M.N., Flynn, P.J.: A survey of clustering algorithms. *ACM Computing Surveys* **31**(3), 264–323 (1999)
- [2] Vizine, A.L., Castro, L.N., Hruschka, E.R., Gudwin, R.R.: Towards improving clustering ants: An adaptive ant clustering algorithm. *Informatica* **29**(2), 143–154 (2005)
- [3] Wikipedia contributors: Antibody opsonization. https://en.wikipedia.org/wiki/Antibody_opsonization. Last edited: 2025-05-25. Accessed: 2026-01-05 (2025)
- [4] MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297. University of California Press, ??? (1967)
- [5] Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* **7**, 1–30 (2006)