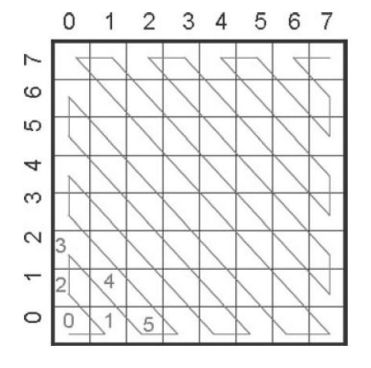
## 2-7- تشكيل الشبكة

يتم في هذه المرحلة ترتيب العناقيد ضمن شبكة وفق قيمة تابع الملائمة الخاصة بكل منها. يجب ترتيب العناقيد ضمن الشبكة بحيث يتشابه كل عنقود مع العناقيد المجاورة له. لتحقيق ذلك تم استخدام منحني ملء الفضاء space filling curve، والذي يهدف لزيارة كل النقاط في فضاء متعدد الأبعاد. تم استبعاد كل من العبور الأفقي والعمودي، والاعتماد على العبور القطري المحيطي كما هو موضح في الشكل 22 وذلك لتسريع وصول الوكيل للمكافأة أثناء تدريب خوارزمية Q-Learning. حيث أنه بتطبيق ذلك، ستكون العناقيد المشابهة لكل عنقود محيطة به، ومنه سيتمكن الوكيل من الوصول إليها بشكل أسرع.

شكل 22: آلية ترتيب العناقيد ضمن الشبكة

## 2-8- تعريف تابع Q-Learning

### 2-8-1- فضاء الحالة State Space

يتكون فضاء الحالة من n2 حالة مميزة، بحيث أن كل حالة هي عنقود مكون من مجموعة مستخدمين ومجموعة عناصر. يمكن اختيار أي حالة على أنها حالة البداية، ولكن تم الحرص على اختيار جميع الحالات على أنها حالة البداية وذلك لضمان الوصول إلى جميع العناقيد الموجودة ضمن الشبكة.

### 2-8-2- فضاء الحركات Action Space

يتضمن الفضاء 4 حركات يمكن القيام بها: خطوة واحدة للأعلى، للأسفل، لليمين، ولليسار.

### 2-8-3- تابع الانتقال Transition Function

يتبع تابع الانتقال الذي اعتمدنا عليه سياسة epsilon greedy، بحيث يتم الانتقال للخطوة التالية إما بشكل عشوائي بنسبة epsilon، أو باتجاه الحالة التي تؤدي لتعظيم المكافأة بنسبة 1-epsilon.

### 2-8-4- تابع المكافأة Reward Function

يعتمد تابع المكافأة R(st , at , st+1) الذي تم اختياره على الحالة الحالية، الحركة التي سيتم القيام بها ابتداءً من الحالة الحالية، والحالة القادمة. تم اختيار مؤشر Jaccard ليكون تابع المكافأة. يعبر هذا المعيار عن عدد المستخدمين المشتركين بين حالتين، بحيث أن مزيد من التداخل (التشارك) يعطي مكافأة أفضل. يمكن تعريف مؤشر Jaccard على النحو التالي:

### 2-8-5- الهدف

يهدف الوكيل لتعظيم المكافأة قبل أن يصل للحالة النهائية، حيث يتم التوقف في حال خرج الوكيل من الشبكة أو لم يتم اكتشاف أي أفلام جديدة بعد الانتقال للحالة الجديدة.

## 2-9- تدريب خوارزمية Q-Learning

خرج الخوارزمية بعد عملية التدريب سيكون Q-Table والذي يحدد قيمة Q لكل حركة ضمن كل حالة. يتم تحديث قيم Q-Table خلال كل خطوة بناء على معادلة Bellman الموضحة سابقاً في الفقرة 1-6-1-1. ضمن الآتي pseudo code يوضح الخوارزمية:

**Input**: State Space S, Action Space A, Transition Function T, Agent

**Output**: Policy π

For i = 1 to *trials*:

Select a random state s,

While *step* < *max steps per episode* or No *game over:*

a = ξ-greedy with action

execute action a

s’ = T(s, a)

Qnew = (1-α)\* Qold + α \* ( r + γ \* maximum Q value in s’ )

if No *new movie found* or *Agent moves off the grid* then:

game over = True

else

s = s’

end

end

end

## 2-10- توليد التوصيات

يتم توليد التوصيات باقتراح مجموعة الأفلام الموجودة ضمن كل حالة يتم المرور عليها ابتداءً من حالة البداية وانتهاءً بحالة النهاية، وبحيث أن كل حركة تتم باتجاه قيمة Q الأعلى وفق السياسة التي تم تعلمها. لتحديد حالة البداية، قمنا باتباع منهجيتين موضحتين في الفقرات الآتية، وتختلفان فيما إذا كان المستخدم جديداً ضمن النظام.

### 2-10-1- حالة مستخدم موجود ضمن النظام

يتم تحديد حالة البداية بالبحث عن العنقود الأكثر تشابهاً مع المستخدم تبعاً لتقييمات الأفلام، وذلك باستخدام تابع التشابه Cosine Similarity وفق الخطوات الآتية:

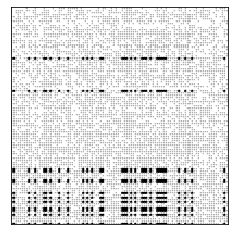
* إنشاء شعاع يتضمن متوسط تقييمات الأفلام، أي متوسط التقييمات التي تم إعطاؤها من قبل كل المستخدمين لكل فيلم موجود ضمن العنقود على حدى.
* إنشاء شعاع خاص بالمستخدم يتضمن تقييماته على نفس الأفلام الموجودة في العنقود.
* حساب قيمة التشابه بين المستخدم وكل من العناقيد باستخدام تابع التشابهcosine similarity بين شعاع المستخدم والشعاع الممثل لهذا العنقود.
* اعتبار العنقود ذو قيمة التشابه الأكبر مع المستخدم على أنه العنقود الممثل لحالة البداية.

### 2-10-2- وجود مشكلة البداية الباردة (حالة مستخدم غير موجود ضمن النظام)

يتم تحديد حالة البداية بالبحث عن العنقود الأكثر تشابهاً مع المستخدم تبعاً للمعلومات الديموغرافية، وذلك باستخدام تابع التشابه Cosine Similarity وفق الخطوات الآتية:

* إنشاء الشعاع الممثل لكل مستخدم من المستخدمين انطلاقاً من معلوماته الديموغرافية بعد معالجتها وترميزها كما هو موضح في الفقرة 2-4-1.
* حساب قيم التشابه بين المستخدم وكل مستخدم ضمن العنقود بحسب تابع التشابه Cosine Similarity.
* اعتبار متوسط قيم التشابه بين المستخدم وكل من المستخدمين الموجودين ضمن العنقود على أنها قيمة تشابه المستخدم مع العنقود.
* اعتبار العنقود ذو قيمة التشابه الأكبر مع المستخدم على أنه العنقود الممثل لحالة البداية.

# الفصل الثالث: التجارب والنتائج

تم تطبيق خوارزمية Bimax الموضحة في الفقرة 2-5 على مصفوفة تقييمات تتألف من 100 مستخدم و100 فيلم، بحيث أن كل مستخدم في هذه المجموعة قام بتقييم 50 فيلم على الأقل وكل فيلم تم تقييمه 50 مرة على الأقل، وذلك لكي تكون المجموعة مليئة بالتقييمات (يكون للمستخدمين سلوك واضح يمكن للخوارزمية التعلم منه). تم اختيار العتبة 3 لتحويل مصفوفة التقييمات إلى مصفوفة ثنائية، بحيث أن التقييم يأخذ القيمة 1 في حال كان أكبر أو يساوي 3 و0 في حال كان غير ذلك. يوضح الشكل 23 أكبر عنقود تم الحصول عليه (العنقود ممثل باللون الأسود وبقية المعطيات ذي القيمة 1 ممثلة باللون الرمادي).

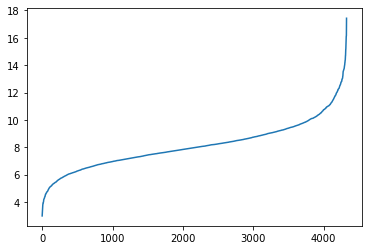
شكل 23: أكبر عنقود تم الحصول عليه بعد تطبيق خوارزمية bimax (العنقود ممثل باللون الأسود وبقية المعطيات ذي القيمة 1 ممثلة باللون الرمادي).

تم إجراء فلترة باستخدام الطريقتين الآتيتين على الترتيب لاختيار العناقيد الأفضل:

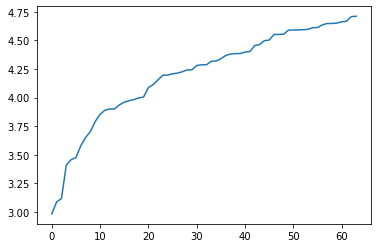
1. فلترة على الحجم: تم اختيار العناقيد التي تحوي على الأقل 30 مستخدم و30 فيلم.
2. فلترة بحسب تابع الملائمة الموضح في المعادلة (2.6). يوضح الشكل 24 مخطط يقابل كل عنقود من العناقيد الناتجة بقيمة تابع الملائمة الخاصة به بعد إجراء عملية ترتيب لهذه القيم.

سنختار حجم الشبكة التي ستمثل فضاء الحالة لخوارزمية Q-Learning ليكون 8x8، ولذلك تم اختيار أفضل 64 عنقود وفق نوعي الفلترة الموضحين أعلاه. يوضح الشكل 25 قيم تابع الملائمة لهذه العناقيد.

قمنا بتدريب خوارزمية Q-Learning على العناقيد الناتجة بعد وضعها ضمن الشبكة المذكورة وذلك لـ 150 محاولة وعدد أعظمي للخطوات ضمن المحاولة قدره 10، وبالإبقاء على القيم الافتراضية للموسطات المترفّعة الخاصة بها. قمنا باختيار حالة البداية بشكل عشوائي في كل محاولة بحيث يتم المرور على كل الحالات، وقمنا باختيار قيمة epsilon = 0.3 لسياسة epsilon-greedy المشروحة في الفقرة 2-8-3 والتي يتنقل الوكيل ضمن تجربتنا وفقها. قمنا بإعطاء قيم عشوائية ضمن المجال [-1, 1] لقيم Q-Table في البداية. يوضح الشكل 26 منحني التعلم الذي تم الحصول عليه، والشكل 27 نفس المنحني ولكن منعّم بأخذ الوسطي على نافذة ذات حجم 10. نلاحظ أنه في البداية يتزايد المنحني ومن ثم يبدأ بالاستقرار، وهذا يدل على اكتمال عملية التعلم.



شكل 24: قيم تابع الملائمة للعناقيد الناتجة



شكل 25: قيم تابع الملائمة لأفضل 64 عنقود

# 

شكل 27: منحني تعلم خوارزمية Q-Learning المنعّم بأخذ الوسطي على نافذة ذات طول 10

شكل 26: منحني تعلم خوارزمية Q-Learning