

مقدمة للأنتروبيا القصوى

An introduction to maximum entropy

عندما يتحدث الكريستالوجرافيون عن الأنتروبيا القصوى فإنهم يشيرون عادةً إلى تقنية لاستخلاص أكبر قدر ممكن من المعلومات من بيانات غير كاملة. إن البيانات المفقودة قد تكون أطوار عامل تركيب أو بعض شدات حيث يمكنها أن تتداخل في نموذج حيود المسحوق. سوف ننظر في هذا الفصل للأفكار ما وراء التقنية.

(١٠, ١) أنتروبيا Entropy

إن الأنتروبيا هو مفهوم مستخدم في الديناميكا الحرارية لوصف حالة ترتيب النظام. كم كبير بين الرياضيات قد نما حوله، وحيث إن نفس الرياضيات تحدث في أي مكان في العلم، فإن مفردات الأنتروبيا قد ذهبت معها. بعيداً عن الديناميكا الحرارية نفسها، فإن المجالات المفيدة من هذه الأفكار هي:

- ١- نظرية معلومات: تقيس الأنتروبيا كمية المعلومات في رسالة. بقدر انخفاض الأنتروبيا تكون المعلومات أكثر.
- ٢- نظرية احتمالية: يقيس الأنتروبيا التغير في الاحتمالية عند تغيير الظروف التي عندها تقدر الاحتمالية. تقابل قيمة أنتروبيا منخفضة احتمالية قصوى.
- ٣- إعداد صورة: تقيس الأنتروبيا كمية المعلومات في صورة.

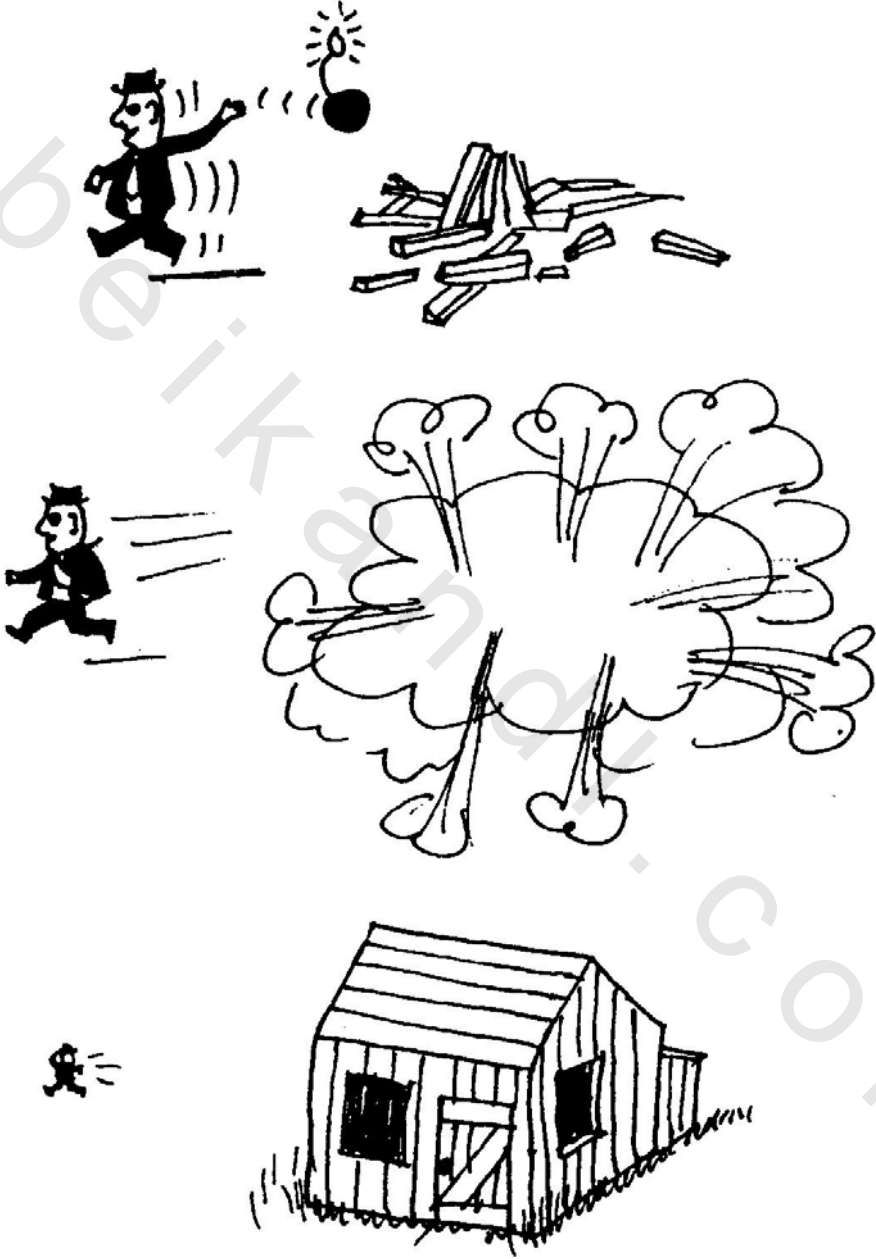
تعني الزيادة في أنثروبيا الذهاب من حالة أقل احتمالية إلى حالة أكثر احتمالية. أمثلة على تزايد أنثروبيا قد تكون في أن حرارتي جسمين تصبحا متساويتين عند تلامس حراري، تصبح الرسالة منحرفة قليلاً عند الانتقال أو تصبح الاحتماليات أقل إفراطاً بسبب أن المعلومات التي تعتمد عليها قد أصبحت قديمة. في كل حالة يجب أن تضيف شيء ما إلى النظام لعكس الاتجاه الطبيعي وهكذا تقلل الأنثروبيا. تزايد الأنثروبيا بطبيعتها.

ك توضيح لهذا هو الحدث بعيد الاحتمال المصور في الشكل رقم (١٠,١). توجد طرق عديدة أكثر لترتيب كتل من الخشب لكي ننتج ركام غير مرتب عن كونها تستخدم لإنتاج سقيفة مفيدة. إن ترتيب عشوائي للخشب يكون لهذا أكثر احتمالاً لكي ينتج ركام بأنثروبيا عالي عن سقيفة بأنثروبيا منخفضة.

(١٠,٢) أنثروبيا قصوى Maximum entropy

عندما يكون الأنثروبيا قيمة عظمى يتضمن هذا أن كل المعلومات قد تمت إزالتها، لا تخبرنا الرسالة عن شيء، كل شيء له نفس الاحتمالية وتكون الصورة مسطحة تماماً. هذه حالة عديمة الفائدة تماماً أن تكون موجودة، لكن ليست هذه هي الطريقة التي تستخدم فيها أنثروبيا قصوى كتقنية عددية.

اعتبر تطبيق أنثروبيا قصوى لمساعدة في تكوين صورة تركيب بلوري، أي لإنتاج خريطة كثافة إلكترونية جيدة بقدر ما تسمح البيانات. عادةً ما تكون البيانات غير دقيقة (خطأ معلمي) وغير مكتملة (تحليل ضعيف، أطوار غير موجودة، انعكاسات متداخلة) معطية احتمالية لعدد لا نهائي من الخرائط التي تكون متوافقة مع نموذج الحيود المرصود.



الشكل رقم (١٠،١). توضيح لسناقص أتروبيا.

كيف يمكن أن تولد خريطة مقبولة خارج العدد اللاهائي من الاحتمالات؟ تحاول أنتروبيا قصوى أن تفعل هذا بطلب أن الخريطة تحتوي على معلومات قليلة بقدر الإمكان أي أن أنتروبيا الخاصة بها تكون عند الحد الأقصى، بالتعريض للقيود المفروضة بالبيانات العملية. يعني هذا إنه أياً كانت المعلومات التي تحتويها الخريطة تكون مطالباً بالبيانات ولا تتواجد كنتاج ثانوي بطريقة عددية أو افتراض محجوب. إنه بالإمكان أن لا تعمل أي افتراضات على الإطلاق عن المعلومات المفقودة وينتج من ثم تقدير عادل لخريطة حقيقية بقدر الإمكان.

(١٠,٣) حسابات بيانات غير مكتملة Calculations with incomplete data

لتوضيح كيف تتعامل مع بيانات غير مكتملة، دعنا نتخيل أن لدينا المعلومات الآتية:

- ١- ثلث جميع العلماء يتعاملون مباشرة مع بيانات بلورية- دعنا نطلق عليهم المهتمين بعلم البلورات.
- ٢- ربع جميع العلماء أعسر left-handed.

الجدول رقم (١٠,١). المشكلة وبعض الحلول الممكنة.

	(أ) كشف عام		(ب) باستخدام المعلومات		(ج) النهاية العظمى الأصغر	
	أيسر	أيمن	أيسر	أيمن	أيسر	أيمن
مهتمين بعلم البلورات	a	b	a	$\frac{1}{3} - a$	0	$\frac{4}{12}$
غير مهتمين بعلم البلورات	c	d	$\frac{1}{4} - a$	$\frac{5}{12} + a$	$\frac{3}{12}$	$\frac{5}{12}$
	(د) الانتروبيا الأكبر		(هـ) التباين الأدنى		(و) الانتروبيا القصوى	
مهتمين بعلم البلورات	$\frac{3}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{7}{24}$	$\frac{1}{12}$	$\frac{3}{12}$
غير مهتمين بعلم البلورات	0	$\frac{8}{12}$	$\frac{5}{24}$	$\frac{11}{24}$	$\frac{2}{12}$	$\frac{6}{12}$

الآن نفرض السؤال: ما هي النسبة من كل العلماء تكون للكريستالوجرافيون الأعرس؟

المعلومات المعطاة عن حالة الأعرس وعلماء الكريستالوجرافيا غير كافية للإجابة عن السؤال بدقة، دعنا نرى إلى أي مدى يمكن أن نذهب إلى إجابة معقولة. يمكن وضع المشكلة كما في الجدول رقم (أ، ١٠)، حيث (أ) النسبة للكريستالوجرافيون أعرس، (د) هي نسبة علماء آخرون أيمن وهكذا. تخبرنا المعلومات بأن:

$$(١٠, ١) \quad a+b = \frac{1}{3} \quad a+c = \frac{1}{4} \quad \text{و} \quad a+b+c+d = 1$$

باستخدام هذه المعادلات الثلاث يمكن أن نحذف b ، c و d ونضع كل شيء في حدود المتغير المفرد a في الجدول رقم (ب، ١٠).

لو أننا رفضنا بشكل معقول علماء بالسالب، فإن أي قيمة لـ a بين 0 و $\frac{1}{4}$ سوف تكون إجابة محتملة للسؤال. على سبيل المثال، تعطي $a = 0$ حل واحد لأقصى درجة، مبين في الجدول رقم (ج، ١٠)، الذي لا يوجد فيه كريستالوجرافيون أعرس على الإطلاق. الطرف الأقصى الآخر بـ $a = \frac{1}{4}$ (الجدول رقم د، ١٠)، حيث يكون كل غير الكريستالوجرافيون أيمن. حيث أن أي من هذا غير محتمل، فنحن بحاجة إلى معيار معقول ليطبق لكي ينتج إجابة مقبولة.

دعنا نرى فيما لو كان من المعقول أن تطلب حل المربعات الصغرى أي إيجاد قيمة a التي تقلل الفرق في المدخلات في الجدول رقم (ب، ١٠) إلى الحد الأدنى. مثل هذا المعيار سوف يجنب بالتأكيد الحلول القصوى التي نظرنا إليها بالفعل. يعطي الفرق حول المتوسط بـ:

$$(١٠,٢) \quad V = \left(a - \frac{1}{4}\right)^2 + \left(\frac{1}{12} - a\right)^2 + a^2 + \left(\frac{1}{6} + a\right)^2$$

ويقع الحد الأدنى V عندما تكون $dV/da = 0$ ، معطية $a = \frac{1}{24}$. هذه النتائج مبيّنة في جدول ١٠,١ (هـ). يبدو من هذا أن $\frac{1}{8}$ من كل الكريستالوجرافيون يكونوا أعرس ($\frac{1}{3} \times \frac{1}{8} = \frac{1}{24}$)، لكن نرى أن $\frac{5}{16}$ من غير الكريستالوجرافيون يكون أيضاً أعرس ($\frac{2}{3} \times \frac{5}{16} = \frac{5}{24}$). لماذا ينبغي أن يوجد هذا الاختلاف؟ إن المعلومة الأصلية لا تشير إلى هذا، ويبدو أننا قد عملنا بعض فرض خفي قد تسبب فيها. إنها قد نشأت بسبب التقنية غير الملائمة المستخدمة للحصول على النتيجة. ليس هذا سبب جيد لماذا كل الاحتمالات ينبغي أن تكون قريبة لبعضها قدر الإمكان. ما تحتاجه فعلياً هو طريقة للحصول على تقدير عادل لعدد الكريستالوجرافيون العسر. حيث الكريستالوجرافيون الأعرس لا نتوقع أن يكونوا أي شيء مختلف عن باقي العلماء في هذا الصدد، فإن من المقبول أو المعقول أن نفترض أن $\frac{1}{4}$ منهم يكونوا أعرس مثل باقي التعداد العلمي. في غياب أي معلومات أكثر، فإن النتيجة الأكثر مصداقية أو قبولاً هي تلك المبينة في الجدول رقم (١٠,١) مع $a = \frac{1}{12}$.

سابقاً كان متطلباً أن تعطى أنتروبيا قصوى تقدير عادل من بيانات غير مكتملة أي لا نعمل افتراضات خفية حول معلومة مفقودة. هل حل أنتروبيا قصوى لهذا يقابل حلنا الأكثر قبولاً في الجدول رقم (١٠,١)؟ سوف تشتق الصيغة لأنتروبيا الكميات في الجدول رقم (ب) ١٠,١ فيما بعد (انظر معادلة ١٠,٩)، لكن هذا لا يوقفنا عن استخدامها هنا. بتطبيق الصيغة لمشكلتنا يعطي:

$$S = -a \log(a) - \left(\frac{1}{3} - a\right) \log\left(\frac{1}{3} - a\right) + \left(\frac{1}{4} - a\right) \log\left(\frac{1}{4} - a\right) - \left(\frac{5}{12} + a\right) \log\left(\frac{5}{12} + a\right)$$

(١٠,٣)

وتحدث الأنتروبيا القصوى عند $dS/da = 0$ ، معطية $a = \frac{1}{12}$. لا يعطي أنتروبيا قصوى هنا الحل الأكثر قبولاً للمشاهد بالفعل في الجدول رقم (١٠,١).

(١٠,٤) تكون الصورة Forming images

حلت أنتروبيا قصوى المشكلة بطريقة أكثر إقناعاً، لكن هل يمكنها بالفعل أن تنتج خرائط كثافة إلكترونية؟ إن كل ترتيب في الجدول رقم (١٠,١) يمكنه بالضبط أن يمثل خريطة كثافة إلكترونية بسهولة مثل احتماليات حالة الأعرس بين التعداد العلمي. على كل، فإن مشكلة الكريستالوجرافي الأعرس قد اختبرت لتوضيح الطريقة بسبب أن كل القيود تكون في نفس الفراغ المكان كعرض رقم مع خريطة الكثافة الإلكترونية تكون المعلومات عن السعات في الفراغ المعكوس. يكون هذا بالتالي على الجانب الخطأ من تحول فورير بقدر ما يتم اعتبار الخريطة، التي تُدخل تعقيدات في الرياضيات. لقد أبقينا على هذا لك حتى يمكنك أن ترى بشكل أكثر وضوحاً كيف تتصرف الطريقة.

(١٠,٥) أنتروبيا واحتمالية Entropy and probability

جزء من تعريف الأنتروبيا هو أن الأنتروبيا لنظام معقد يكون مجموع أنتروبيات أجزائه المنفصلة، يكون حالة ترتيب نظام المقاسة بالأنتروبي احتمالية معينة للحدوث. ذلك بأن لكل قيمة أنتروبيا يكون هناك قيمة احتمالية مقابلة. يمكن أن يكتب هذا: $S = f(P)$ ، حيث S هي أنتروبيا، P احتمالية و f هي الدالة بينهما.

دعنا الآن ندرس نظام من جزأين بأنثروبيا S_1 و S_2 واحتمالين مقابلين P_1 و P_2 . لو أن هذه الحالات تكون غير معتمدة على بعضها البعض، تكون احتمالية النظام المتحد هو $P_1 \times P_2$ بينما يكون الأنثروبيا له $S_1 + S_2$. هكذا يكون أنثروبيا الكل:

$$S = S_1 + S_2 = f(P_1) + f(P_2) = f(P_1 \times P_2) \quad (١٠,٤)$$

قارن هذه العلاقة بخاصية لوغاريتمية أساسية:

$$\log(a) + \log(b) = \log(a \times b) \quad (١٠,٥)$$

واضح من هذا أنه يمكننا أن نكتب للأنثروبيا $S = \log(P)$ ونهمل أي عوامل ثابتة قد تضاعف دالة اللوغاريتم.

(١٠,٦) خرائط كثافة إلكترونية Electron density maps

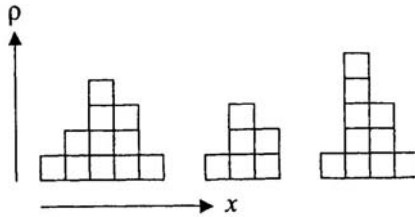
لو بالإمكان أن نشغل على احتمالية كثافة إلكترونية واقعة، فإن هذا سوف يعطي في الحال قياس للأنثروبيا الخاصة بها. دعنا نتخيل بناء خريطة في بعدين على صينية باستخدام حبات الرمل. يكون الحوض المسطح مقسماً إلى صناديق صغيرة مقابلة إلى نقاط شبكة الإحداثيات المرجعية التي تحسب عندها الخريطة عادة، وتمثل الكثافة بعدد حبات الرمل المكدسة في كل صندوق. إن نثر الرمل داخل الحوض المسطح عشوائياً سوف ينتج خريطة، رغم أنها عادة لا تكون طريقة جيدة جداً. من ناحية ثانية فإن عدد الطرق التي يمكن بها ترتيب حبات الرمل لإنتاج الخريطة هو مقياس كم تكون مرجحة أن تحدث. لو تكون هناك طريقة واحدة لترتيب الرمل لإنتاج خريطة، فأهنا سوف تحدث نادراً جداً بالرمي العشوائي.

نحن الآن في حاجة إلى أن نشتغل على كم طريقة يمكن أن يرتب بها الرمل ليعطي خريطة معينة. يبين الشكل رقم (١٠,٢) كيف أن خريطة البعد الواحد يمكن عملها من حبوب منفردة. افترض أن الرمل يتكون من عدد N من حبات متماثلة وأن الخريطة تعمل من وضع حب منفردة في موقع عند زمن ما. تكون للحبة الأولى عدد N اختيار أو فرصة من الأماكن لتذهب فيه. يكون للحبة الثانية عدد $N-1$ من الفرص، من ثم يمكن وضع الاثنتين في الخريطة في $N \times (N-1)$ من طرق مختلفة. يكون للحبة الثالثة $N-2$ مكان للذهاب، معطية $N \times (N-1) \times (N-2)$ اتحادات من الأماكن للحبات الثلاث. هكذا يمكن ملاحظة أن كل الحبات N يمكن أن ترتب في $N!$ طرق معاً.

حيث أن الحبات تكون متماثلة، فليس من الأهمية كيفية ترتيبها في الحوض المسطح؛ عدد الحبات فقط في الحوض المسطح هو الذي يؤثر على شكل الخريطة. لو أن هناك n_1 حبة في الصندوق الأول، يمكن أن ترتب في $n_1!$ من طرق مختلفة داخل الصندوق بدون التأثير على الخريطة. على سبيل المثال، يوضح الشكل رقم (١٠,٣) الترتيبات الممكنة ($3!=6$) لثلاث حبات. لذا، فإن اتحادات $N!$ لا بد من اختزالها بهذا العامل لتترك $N!/n_1!$ اتحادات. كل صندوق يمكن أن يعامل بهذه الطريقة، ومن ثم يكون الرقم النهائي لاتحادات الموضع المختلفة لحبات الرمل هو:

$$(10,6) \quad \frac{N!}{n_1!n_2!\dots n_m!}$$

حيث يوجد m نقاط شبكة إحداثيات مرجعية grid في الخريطة.



الشكل رقم (١٠,٢). تمثيل تخطيطي لخريطة في بعد واحد.

1	3	2	3	1	2
2	1	3	2	3	1
3	2	1	1	2	3

الشكل رقم (٣، ١٠). طرق ترتيب ثلاثة أشياء في صندوق.

سوف يكون هذا متناسباً مع احتمالية الحدوث للخريطة. يمكننا لهذا أن نحصل على قياس للأنتروبيا S بأخذ لوغاريتم هذا التعبير:

$$(١٠,٧) \quad S = \log\left(\frac{N!}{n_1!n_2!\cdots n_m!}\right)$$

يتم تبسيط هذا بشكل كبير بأخذ تقريب ستيرلنج Stirling لمضروب الأعداد الكبيرة:

$$(١٠,٨) \quad \log(N!) \approx N \log(N) - N$$

بمعاملة كل المضروبات بهذه الطريقة وتذكر أن $\sum n_i = N$ يؤدي إلى الصيغة للأنتروبيا الخريطة:

$$(١٠,٩) \quad S = -\sum_{i=1}^m n_i \log(n_i)$$

لو أنك تقيس الكثافة الإلكترونية باستخدام إلكترونات/ Å^3 بدلاً من عد حبات الرمل، ينبغي لصيغة الأنتروبي أن تتغير إلى:

$$(10, 10) \quad S = \sum_{i=1}^m \rho_i \log\left(\frac{\rho_i}{q_i}\right)$$

حيث ρ_i هي الكثافة المصاحبة لـ i th نقطة شبكية إحدائيات مرجعية و q_i هي الكثافة المتوقعة عند النقطة. بداية سوف تكون q_i بالضبط متوسط الكثافة في الخلية، لكن يمكن تحديثها كلما حصلنا على معلومات أكثر.