

الخوارزميات و الخيارات - ٢

د. خليل السعيد

تعقيباً على مقالنا الأول نود التذكير بحقيقتين هامتين ينبغي أخذهما بعين الاعتبار في القراءة اللاحقة: فالحقيقة الأولى مفادها أن إيجاد الخيار الأمثل من بين حلول مسألة ما يتطلب استنفاد سائر تلك الحلول. أما الحقيقة الثانية فهي أن عدد الحلول قد يكون من الضخامة بحيث أن التعقيد الأساسي في المسألة يأتي مترتباً على تلك الضخامة.

و الآن دعونا نوضح بعض المصطلحات التي سنستخدمها لاحقاً. فعندما نقول **الحل الأمثل** فنحن نقصد الحل الذي يلبي **معياراً محدداً و واضحاً للمثالية** يعبر عنه ضمن شروط المسألة لفظياً أو بواسطة **دالة** يتوخى تحقيق **مؤشر كمي أصغري أو أعظمي لها**. و قد لا تتطلب بعض المسائل إيجاد حل أمثل بالضرورة، بل **إثبات وجود حل ما أو استحالة**.

عادة ما نقصد **بحجم مسألة ما عدد عناصر مدخلاتها**. فعلى سبيل المثال في مسألة الأحجار يكون حجم المسألة هو عدد الأحجار. كما **نقصد بدرجة تعقيد المسألة عدد الحلول الناتجة** و التي علينا اختيار الأمثل منها وفق الشروط الواردة في المسألة. و يكون عدد الحلول على صورة دالة يتمثل المتغير المستقل فيها بحجم المسألة. فإذا كانت الدالة على شكل كثير حدود قلنا بأن درجة التعقيد على شكل كثير حدود، و إذا كانت الدالة أسية قلنا بأن درجة التعقيد أسية.

و الآن لنتناول عدداً من الأمثلة الشبيهة بالتى أسلفنا ذكرها و لكن ببعض من التعقيد.

المسألة الأولى مسألة التاجر الرحالة (Travelling Salesman Problem)

لنفرض أن تاجراً يعتزم القيام بجولة في مجموعة من المدن عددها n شريطة أن يزور كل منها مرة واحدة و يعود في نهاية المطاف إلى المدينة التي ابتداء منها رحلته و بأدنى مجموع لقيم تذاكر السفر، علماً أن الرحلات الجوية يتم تسييرها بين كافة المدن و قيمة التذكرة من المدينة i إلى المدينة j (على سبيل المثال) قد تختلف عن قيمة التذكرة من المدينة j إلى المدينة i . بعبارة أخرى: على التاجر

اختيار المسار (أو ربما المسارات) الأدنى تكلفة لرحلته من بين مجموعة من المسارات كافة.

إذاً، لدينا المصفوفة $A = \|a_{ij}\|$ ذات المقياس $n \times n$ حيث a_{ij} هي تكلفة التذكرة بين i و j ، و الدالة C التي تحدد تكلفة المسار الذي يشتمل على العدد n من عناصر المصفوفة بحيث:

$$C = \sum_{k=1}^n (a_{ij})_k \rightarrow \min$$

نجد هنا بأن عدد خيارات مدينة الانطلاق يساوي n أما عدد خيارات المدينة الثانية فيساوي $n-1$ و الثالثة $n-2$ و هكذا إلى أن يتبقى الخيار الأوحده هو عودة التاجر إلى حيث انطلق. إذاً فإن عدد المسارات (الخيارات) يساوي $n \times (n-1) \times (n-2) \times \dots \times 2 \times 1 = n!$

أي يمكننا القول بأن درجة التعقيد النظرية لحل هذه المسألة تساوي $O(n!)$ و هل تعلم يا عزيزي القارئ بأن حل مسألة من هذا النوع لمجموعة تتضمن ٢٥ مدينة فقط سيتطلب ما يزيد على ٤٠ عاماً من العمل المتواصل لأحدث الحواسيب في وقتنا المعاصر؟ و إذا لم تصدق فحاول حساب قيمة $٢٥!$

المسألة الثانية مسألة التقسيم لتكن لدينا المجموعة A التي تضم العدد n من العناصر بحيث يتمتع كل عنصر بمؤشر عددي قيمته a_i و المطلوب تقسيم هذه المجموعة إلى مجموعتين جزئيتين A_1, A_2 بحيث يكون $\{A_1 \cap A_2\} = \emptyset$ و كذلك بحيث أن مجموع المؤشرات العددية في هاتين المجموعتين يساوي $\sum_{i=1}^n a_i / 2$ و إذا لم نتمكن من تقسيم المجموعة وفق هذه الشروط فعلينا إثبات استحالة الحل من خلال التحقق من كافة عمليات التقسيم. نلاحظ هنا أن مسألة الأحجار السالف ذكرها هي عبارة عن مسألة تقسيم.

إن عدد خيارات التقسيم هنا يساوي 2^n

المسألة الثالثة مسألة الجعبة لنفرض أن لدينا مجموعة تضم العدد n من الأشياء الثمينة، و يتميز كل من تلك الأشياء بحجم معين a_i و ثمن معين b_i و

المطلوب ملئ الجعبة ذات الحجم المحدود B بعدد من هذه الأشياء بحيث يكون ثمنها مجتمعة أعظماً. أي:

$$\sum_{i=1}^m b_i \leq B$$

$$\sum_{i=1}^m a_i \rightarrow \max$$

عدد خيارات الانتقاء هنا أيضاً يساوي 2^n

المسألة الرابعة شبكة الطرق لنفرض أن محافظة ما تضم العدد n من المدن و النواحي و القرى المطلوب ربطها بشبكة من الطرق المعبدة بحيث يمكن الانتقال بين أي نقطة و أخرى ضمن المحافظة، و في ذات الوقت يكون المجموع الكلي لطول الطرق في الشبكة أصغرياً.

يمكن بناء النموذج الرياضي لهذه المسألة على غرار مسألة التاجر الرحالة، و يمكننا التأكد من أن عدد الخيارات فيها يساوي 2^{n-2}

إن التمتع في هذه المسائل يدفعنا إلى استنتاج مفاده أنها جميعها مسائل ذات طبيعة توافقية (combinatorial) لأنها تتطلب إثبات الوجود أو الاستحالة و تتضمن عدداً من الخيارات الواجب البحث فيها وفقاً لشروط معينة.

إن أية مسألة تتمتع بمثل هذه المزايا يجب و سوف تكون ضمن مجال اهتمامنا لو أردنا الولوج إلى كنه نظرية الخوارزميات.

لن نتوقف كثيراً عند طبيعة المسائل التوافقية و البحث في مزاياها لأن ذلك يتطلب بحد ذاته بحثاً منفرداً.

هناك ثمة لغة تستخدم للتعبير المبسط و الأنيق عن مسائل الخيارات و خوارزميات حلولها تسمى نظرية البنى (Graph Theory) و تلقى رواجاً كبيراً في شتى فروع العلوم التطبيقية.

ترتبط المسائل التوافقية بتنظيم العمليات الحسابية و المنطقية على الحاسوب و كذلك بنى المعطيات المختلفة و ترتيبها و البحث فيها و اشتقاق

البنى الأخرى منها. هذا و قد أفضت جهود العلماء و الباحثين إلى عدد كبير من الخوارزميات الحاذقة و البارعة في حل المسائل التوافقية.

الخوارزميات المستنفذة و درجة تعقيد الخوارزمية

إن مصطلحي "المسائل التوافقية" و "مسائل الخيارات" هما في واقع الأمر مرادفان لمفهوم يقصد به ذات الدرجة من التعقيد، إذ يقتضي ذلك المفهوم استنفاد كافة الحلول . و تختلف مسائل الخيارات عن المسائل التوافقية كون الأولى هي مسائل توافقية تضاف إليها شروط الحل الأمثل. و تسمى **الخوارزميات المستنفذة** تلك الخوارزميات التي **تستنفذ سائر حلول المسألة الواحد تلو الآخر بالبحث و المقارنة بغية اختيار الأمثل منها**.

كما أسلفنا الذكر فإن ظهور الحاسوب قد سمح بحل مسائل كانت منذ عهد ليس ببعيد شبه مستحيلة. فلو أخذنا على سبيل المثال مسألة التقسيم، فقد كان من الممكن قبل ظهور الحاسوب حل تلك المسألة بالخوارزميات المستنفذة يدوياً لمجموعة تضم ١٠ عناصر، بينما تسنى بظهور الحاسوب حلها لمسألة تضم ٣٠ عنصراً.

و كما أسلفنا أيضاً فإن إمكانيات الحاسوب محدودة، و يبين الجدولان الواردان أدناه تلك المحدودية.

يبين الجدول الأول تأثير التطور التقني للحواسيب على الخوارزميات كثيرة الحدود و الأسية.

درجة التعقيد	الحجم الأقصى لمدخلات المسألة الممكن حلها خلال ساعة واحدة على		
	الحواسيب المعاصرة	على حواسيب أسرع بـ ١٠٠ مرة من الحواسيب المعاصرة	على حواسيب أسرع بـ ١٠٠٠ مرة من الحواسيب المعاصرة
n	N_1	$100N_1$	$1000N_1$

$31,6N_2$	$10N_2$	N_2	n^2
$10N_3$	$4,64N_3$	N_3	n^3
$3,98N_4$	$2,5N_4$	N_4	n^5
$9,97 + N_5$	$6,64 + N_5$	N_5	2^n
$6,29 + N_6$	$4,19 + N_6$	N_6	3^n

و يبين الجدول الثاني الحجم الأقصى لمدخلات المسألة الممكن حلها خلال الزمن المبين

زمن حل المسألة						درجة التعقيد
ثانية واحدة	١٠ ^٢	١٠ ^٤	١٠ ^٦	١٠ ^٨	١٠ ^{١٠}	
ثانية واحدة	ثانية واحدة	ثانية واحدة	ثانية واحدة	ثانية واحدة	ثانية واحدة	
(1,7)	(2,7)	(١٢)	(٣)	(٣٠٠)		
(دقيقة)	(ساعة)	(يوم)	(سنوات)	(سنة)		
١٠ ^٣	١٠ ^٥	١٠ ^٧	١٠ ^٩	١٠ ^{١١}	١٠ ^{١٣}	1000 n
١٤٠	7,7.10 ³	5,2.10 ⁵	3,9.10 ⁷	3,1.10 ⁹	2,6.10 ¹¹	1000 n log n
١٠ ^٢	١٠ ^٣	١٠ ^٤	١٠ ^٥	١٠ ^٦	١٠ ^٧	100 n ²
٤٦	2,1.10 ²	١٠ ^٣	٦.10 ³	2,1.10 ⁴	١٠ ^٥	10 n ³
٢٢	٣٦	٥٤	٧٩	١١٢	١٥٦	n log n
٥٩	٧٩	٩٩	١١٩	١٣٩	١٥٩	2 ^{n/3}
١٩	٢٦	٣٣	٣٩	٤٦	٥٣	2 ⁿ
١٢	١٦	٢٠	٢٥	٢٩	٣٣	3 ⁿ

تسمى درجة تعقيد الخوارزمية تلك الدالة التي تعبر عن عدد خطوات الحل بالاعتماد على حجم مدخلات المسألة. فإذا كانت لدينا على سبيل المثال:

$$f(n) = 3n^4 + 5n^3 \log n + 6n + 4$$

و كانت n عدد عناصر المجموعة قيد البحث، فإننا نقول بأن $f(n)$ هو عدد خطوات الحل للخوارزمية، و نقول بأن هذه الخوارزمية ذات درجة تعقيد كثيرة الحدود و تساوي $O(n^4)$

أو $O(n^5)$ و تعتبر درجة التعقيد هذه نظرية و هي المعتمدة عادة في تقييم الخوارزميات. و تعتبر $O(n^4)$ تقييماً من الأسفل للخوارزمية، بينما $O(n^5)$ تقييماً من الأعلى.

و إذا كانت لدينا:

$$g(n) = 2^{2n} + 2^{n-4} + 3n^3$$

فتكون درجة تعقيد الخوارزمية نظرياً $O(2^n)$

تسمى $f(n)$ بدالة كثيرة الحدود بينما $g(n)$ دالة أسية. و تعتبر الخوارزمية فعالة (جيدة) إذا كانت من نوع $f(n)$ و يقال عنها بأنها غير فعالة (سيئة) إذا كانت من نوع $g(n)$ و هذا واضح من خلال الجدولين السابقين.

من الجدير ذكره أن كافة مسائل الخيارات يمكن تقسيمها إلى صنفين: مسائل ذات خوارزميات جيدة و مسائل ذات خوارزميات سيئة، أو -كما يعتبر البعض- مسائل لا يوجد لها خوارزميات جيدة حتى اليوم.

نود التنويه هنا إلى أن مصطلح "درجة التعقيد" يمكن تأويله بأشكال مختلفة. فهناك درجة التعقيد النظرية، و هناك درجة التعقيد المتوسطة التي تقاس من خلال عدة نماذج لمسألة محددة ذات حجم مدخلات n و تقسيم زمن الحل على عدد تلك النماذج. كما يمكن قياس درجة التعقيد من الأسفل، أي أن زمن الحل لا يقل عما تشير إليه الدالة، و القياس من الأعلى يعني أن الزمن لا يزيد عما تشير إليه الدالة. و تأخذ درجة التعقيد بالاعتبار أسوأ الأحوال، أي أن الحل الأمثل يمكن أن يكون ترتيبه الأخير ضمن مجموعة الحلول قيد البحث. و تعد درجة التعقيد النظرية تجريدية إلى حد كبير و هي قد لا تعكس في بعض الأحيان السلوك التطبيقي للخوارزمية. فعلى سبيل المثال تعمل خوارزمية الترتيب السريع لمجموعة n من العناصر و التي درجة تعقيدها $O(n^2)$ بصورة أكثر فعالية من

خوارزمية الترتيب الهرمي ذات الدرجة $O(n \log_2 n)$ مع أن درجة تعقيد الأخيرة أدنى من الأولى. و تعمل كلتا الخوارزميتان بدرجة تعقيد متوسطة تساوي $O(n \log_2 n)$

هذا و تعتمد درجة تعقيد الخوارزمية ليس على حجم المسألة فقط بل على شروطها أيضاً، فتغيير بسيط في الشروط قد يؤدي إلى تبسيط أو تعقيد المسألة. فعلى سبيل المثال توجد خوارزمية فعالة لحل مسألة شبكة الطرق، أي أنه بإمكاننا إيجاد الحل الأمثل بواسطة خوارزمية كثيرة الحدود، و لكن خوارزمية كثيرة الحدود قد لا يكون بمقدورها حل المسألة فيما لو طلب منا إيجاد كافة الحلول المثلى. فعدد الشبكات يساوي 2^{n-2} و ليس من المستبعد أن تكون كلها مثلى.

كما تعتمد درجة التعقيد أيضاً على حجم الذاكرة المستخدمة، فعلى سبيل المثال لو كانت لدينا سلسلة طولها n من الأرقام ٠ و ١ و طلب منا إيجاد عدد الأرقام ١ في هذه السلسلة. نجد ببساطة انه ينبغي المرور بكافة أرقام السلسلة و حساب عدد الأرقام ١ فيها، أي بواسطة خوارزمية تعقيدها $O(n)$ و يتطلب ذلك n من وحدات الذاكرة التي تخزن فيها السلسلة. أما لو تم تخزين كافة السلاسل ذات العدد n من الأرقام فإن رقم كل سلسلة يحتاج لوحدة ذاكرة واحدة فقط و عدد تلك الوحدات هو 2^n ، سوى أن خوارزمية إيجاد ذلك العدد تتطلب الوصول إلى وحدة الذاكرة التي تخزن السلسلة المطلوبة و ذلك بخطوة واحدة فقط، أي $O(1)$

و قبل الانتقال إلى البحث المفصل في بعض أنواع الخوارزميات (و هو ما سيكون موضوع المقال القادم) نود التنويه إلى ما سنستخدمه من خوارزميات يستحسن معرفة درجة تعقيدها سلفاً. فحسب عدد بطول n يتطلب خوارزمية بتعقيد $O(n \log n \log \log n)$ ، و جداء مصفوفتين يتطلب $O(n^{2.2})$ ، و حل مسائل البرمجة الخطية يتطلب $O(n^{3.5})$ ، و التعرف على الأعداد الأولية $O(\log \log n)$ ، و تحويلات فورييه السريعة $O(n \log n)$