アルゴリズム論

中野

Note 2 NP完全

2020.4.8作成 2021.4.16update 4.19update

今回の講義の概要

計算可能な問題は次の4種類に分類できる。

(1)　多項式時間アルゴリズム(計算時間が多項式のオーダーのアルゴリズム)が知られている問題 (クラス P)(Pは、polynomial多項式の意味)
ソート・サーチ・グラフの最短経路・平面グラフの判定・最小木・線形計画法[Karmarkar,1984]・......

(2)　指数関数時間かかることが判明しているもの
グラフの全ての部分木の発生・....

(3)　解の検証は多項式時間でできるが、解を求める多項式時間アルゴリズムは知られていないもの
(クラス NP完全 (NP-Complete))
充足可能性問題・独立点集合問題・ハミルトン閉路問題・.....

(4)　上記3つのどれにはいるか、現在不明のもの
グラフ同形問題(graph isomorphism)・素因数分解(prime factoring)・...

今回は、NP完全について解説する。
答が(Yes/No)のいずれかである問題を判定問題という。
多項式時間で検証できる判定問題の集合がクラスNPである。
クラスNPの問題のうち、一番ムズカシイ問題の集合がNP完全である。
もしNP完全に属する問題のひとつが、多項式時間で解けたら、クラスNPに属する問題は全て多項式時間で解ける。すなわちP=NPとなる。（のだけど、1971年以来世界中の研究者がトライしても、未解決なので、P≠NPだと予想されている。）

もし、あなたが、とりくんでいる問題がNP完全であるならば、これを解く、
高速なアルゴリズムを設計しようとすることは、とてつもなくムズカシイことなのである。
近似解を求める近似アルゴリズムや、
必ず解を求めるとは限らない確率的アルゴリズム・
発見的アルゴリズムを設計したほうが、現時点では、賢い選択といえる。

なぜ？多項式時間か

データがn個の問題を解くのに2種類のアルゴリズムがあり、
アルゴリズムAは2のn乗回の演算を用いて問題を解き、
アルゴリズムBはnの5乗の1000倍回の演算を用いて問題を解くとする。
　アルゴリズムAは指数関数時間アルゴリズムであり、
アルゴリズムBは多項式時間アルゴリズムである。
もし、1秒に1憶回の演算をする計算機で、問題を解くと、かかる時間は。。

	10	20	30	40	50	60	70	80	90
アルゴリズムA	10万分の1秒	100分の1秒	11秒	3時間	130日	371年	37万年	4憶年	4千憶年
アルゴリズムB	1秒	31秒	4分	17分	52分	2時間	5時間	9時間	16時間

なぜ？多項式時間か

このように、nの増加にしたがって、計算時間が増加するが、nが十分大きいと、
多項式時間よりも、指数関数時間のほうが必ず大きくなる。
つまり、小さな問題では、どちらが早く問題を解くかは、
あまり問題にならないことが多いが、(0.001秒 vs. 0.0012秒とか)
大きな問題では、多項式時間の方がきわめて早く問題を解くのである。

	10	20	30	40	50	60	70	80	90
アルゴリズムA	10万分の1秒	100分の1秒	11秒	3時間	130日	371年	37万年	4億年	4千億年
アルゴリズムB	1秒	31秒	4分	17分	52分	2時間	5時間	9時間	16時間

なぜ？多項式時間か

また、多くの研究者が開発している多項式時間アルゴリズムは、計算時間が
$O(n^2)$とか、$O(n^3)$とか、$O(n^2 \log n)$とかであり、
指数部の定数は3以下であることがほとんどであり、
$O(n^{10})$とかは、まずない。
このような意味でも多項式時間アルゴリズムは高速なのである。

	10	20	30	40	50	60	70	80	90
アルゴリズムA	10万分の1秒	100分の1秒	11秒	3時間	130日	371年	37万年	4億年	4千億年
アルゴリズムB	1秒	31秒	4分	17分	52分	2時間	5時間	9時間	16時間

判定問題

問題には、いくつかの種類がある。

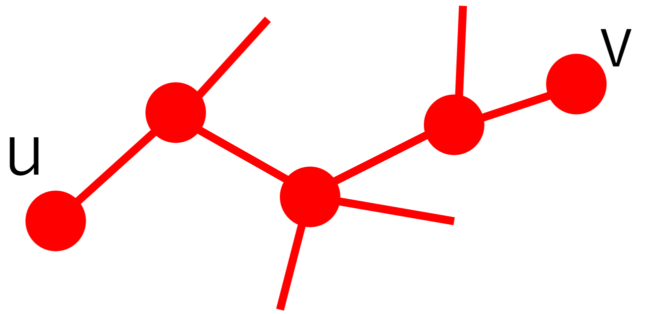
判定問題　答はYesかNOかのどちらか
　例　グラフGと2点u,vと定数kが与えられたとき、u,v間に長さk以下の道があるか?

最適化問題(Optimization problem)
　例　グラフGと2点u,vが与えられたとき、u,v間の最短距離を求めよ

数えあげ問題・列挙問題
　例　グラフGと2点u,vと定数kが与えられたとき、u,v間の長さkの道を列挙せよ
　(最近、流行っている研究分野のひとつである。)

探索問題　例　グラフGと2点u,vと定数kが与えられたとき、u,v間の長さk以下の道を一本求めよ

そのほかにも。。。



判定問題と最適化問題の関係

最適化問題を解くアルゴリズムは、判定問題も解ける

一般に、判定問題を解くアルゴリズムを繰り返し用いることにより、最適化問題も解ける。(2分探索で繰り返し解く)

つまり、判定問題の方が最適化問題より難しいことはない。

判定問題と最適化問題の関係

例
グラフの最短路問題で考えよう。
グラフGと2点u,vが与えられたとき、u,v間の最短距離を求めるアルゴリズムAがあるとする。
このアルゴリズムAを少し改造して、最短の道を求めたあと、
この最短の道の長さと定数kを比較して、判定問題の解を求めるアルゴリズムBが容易に作れる。

グラフGと2点u,vと定数kが与えられたとき、u,v間に長さk以下の道があるかどうか
を判定するアルゴリズムCがあるとする。Cを log m 回用いれば(2分探索で)
最適化問題を解くことができる。(ここでmはグラフの辺の本数)

つまり、
判定問題が多項式時間で解けるなら、最適化問題も多項式時間で解け、
最適化問題が多項式時間で解けるなら、判定問題も多項式時間で解ける。(ことが多い)

クラスP

多項式時間アルゴリズムが知られている判定問題の集合を
クラス P (complexity class P)という。

クラスNP

ホントは、言語とか非決定性チューリングマシンとかを使って厳密に定義する。
(decideとacceptの違いとか、co-NPとかは、計算理論を勉強しましょう。)
のだけど、この講義では、簡単に、
" 問題の解を与えられたとき、多項式時間で検証(verifies)できる判定問題の集合"
とします。

例　　グラフGが与えられたとき、各点をちょうど一回づつ通るサイクル
(ハミルトニ アンサイクル)がGに存在するかどうかを判定する問題を考えよう。
解(サイクル)が与えられれば、これが正しいかどうか検証することは、
多項式時間でできる。

クラスNP完全

(直観的な説明)
NP完全とはクラスNPの部分集合で、一番ムズカシイ問題からなるものです。

次のふたつを満たす問題Aの集合が、NP完全である。

1. 問題AはNPに属する。
2. もし、問題Aが多項式時間で解けたら、NPに属する問題はすべて多項式時間で解ける。（この後、解説します）

P=NP?

もしNP完全に属する問題のひとつが、多項式時間で解けたら、
クラスNPに属する問題は全て多項式時間で解ける。
すなわちP=NPとなる。(のだけど、1971年以来世界中の研究者がトライしても、未解決なので、P≠NPだと予想されている。)

(この問題の解決には100万ドルの賞金もかかっています.
http://www.claymath.org/millennium-problems/millennium-prize-problems　および
https://www.claymath.org/millennium-problems
参照)

NP完全問題の例　Satisfiability Problem(SAT)

充足可能性問題(Satisfiability)(SAT)
(1971年に一番最初にNP完全であることが証明された問題)

論理変数　x_1, x_2 ,….
リテラル　x_j, $\bar{x}_j$　(j=1,2,….)
$\bar{x}$はｘの否定（ほんとはxの真上にバー）
リテラルの和の節C_i (i=1,2, ..., m)の集合が与えられたとき、
この節すべてを1にするような論理変数への0,1の割り当てがあるかどうか判定せよ。

NP完全問題の例　Satisfiability Problem(SAT)

例1

$C_1=x_1+x_2+\bar{x}_3$

$C_2=\bar{x}_2+\bar{x}_3$

$C_3=x_3+x_4$

は $x_1=1, x_2=0, x_3=0, x_4=1$

とすれば、3つの節のすべてを1にできる。

例2

$C_1=x_1$

$C_2=\bar{x}_1$

は、$x_1=1$と0のどちらでも、いずれかの節が0になってしまう。

つまりどうしても2つの節すべてを1にできない。

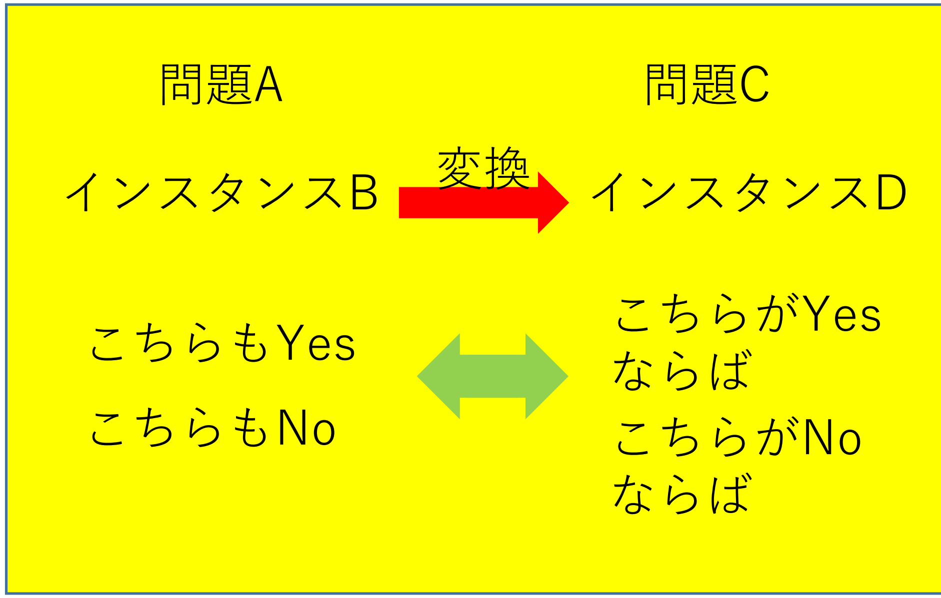
なぜ、SATがNP完全問題であるのかは、本講義では証明しない。

(計算理論を勉強しましょう。)

NP完全問題であることを証明するには

SATがNP完全であることを利用すると、ある問題がNP完全であることを証明することは比較的容易である。

問題Aの任意のインスタンスBを、問題CのインスタンスDに変換する多項式時間アルゴリズムがあり、
このとき、BがYESのときは、必ずDがYESであり、BがNOのときは、必ずDがNOであるならば、
問題Aは問題Cに多項式時間帰着可能である、という。



問題Aが問題Cに多項式時間帰着可能であるとき、問題Cが多項式時間アルゴリズム
を持てば、問題Aも多項式時間アルゴリズムを持つことになる。
(もし問題Aが多項式時間アルゴリズムを持ってもCが多項式時間アルゴリズムを
持つかどうかはワカラナイことに注意)

(NP完全問題である)SATが, (NPに属する)問題Xに多項式時間帰着可能ならば、
XもNP完全である。
(つまり、もし、問題Xが多項式時間アルゴリズムを持てば、
(NP完全問題である)SATも多項式時間アルゴリズムを持ち,
NPに属する問題はすべて多項式時間で解けることになる.)

NP完全問題いろいろ

3SAT

CLIQUE

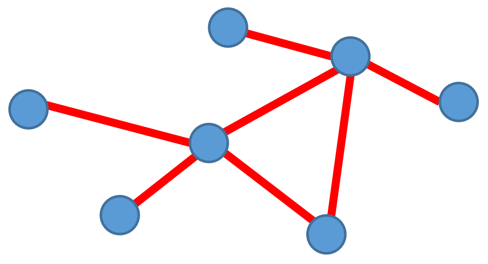
VERTEX-COVER

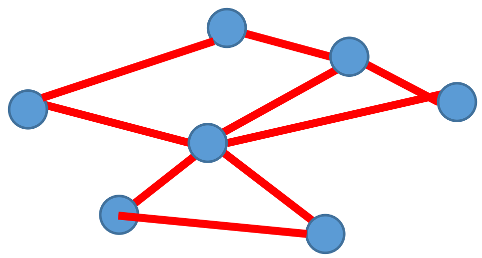
SUBSET-SUM　　　11　25　41　47　69　49

HAM-CYCLE

TSP

など..........(数千の問題がある)





NP完全であることの証明の例

3SAT問題 (NP完全問題であることを次週証明します)

論理変数　x_1,　x_2 ,....

リテラルをx_j, x^-_j (j=1,2,....)とする。

ちょうど3つのリテラルの和であるような節の集合

C_i (i=1,2, ...,)が与えられたとき、

この節すべてを1にするような

論理変数への

0,1の割り当てがあるかどうか判定せよ。

$C_1=x_1+x^-_2+x^-_3$

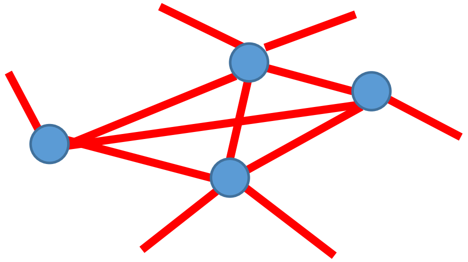
$C_2=x^-_1+x_5+x_3$

$C_3=x_1+x_2+x_4$

$C_4=x_3+x_7+x_4$

CLIQUE問題 (今からNP完全問題であることを証明します)

グラフがと整数kが与えられたとき、
サイズkのクリーク(完全部分グラフ)を持つか判定せよ
(サイズkのクリークとは、k個の点の集合で、
そのk点中のどの2点間にも辺があるもの、です。)



サイズ4のクリークの例

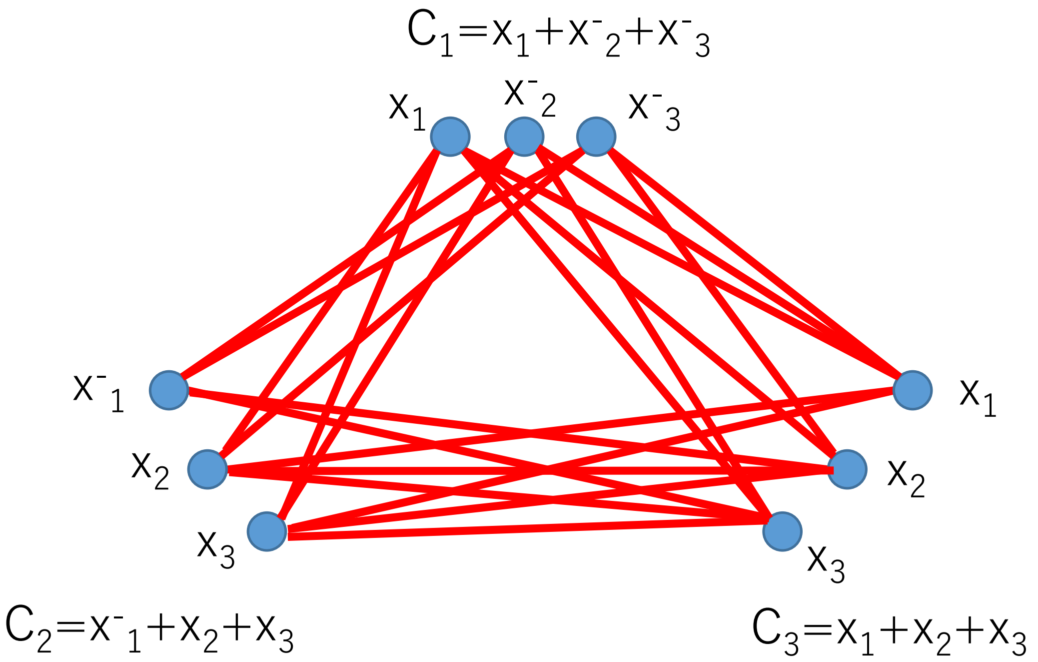
NP完全であることの証明の例

CLIQUE問題はNPに属していることは明らか。(答を多項式時間で検証できるから。)
あとは、(NP完全問題である)3SAT問題がCLIQUE問題に多項式時間帰着可能であることを示せばよい。

任意の3SATのインスタンスを、CLIQUEのインスタンスに次のように変換する。
各節の各リテラルに対応して点を置く。(節の個数をkとします。)
各2点のペアx,yが、次のふたつの条件を満たすとき、x,y間に辺を置く。
(1)　x,yは異なる節に属する
(2)　xはyの否定ではない。

例
$C_1=x_1+x^{-}_2+x^{-}_3$
$C_2=x^{-}_1+x_2+x_3$
$C_3=x_1+x_2+x_3$

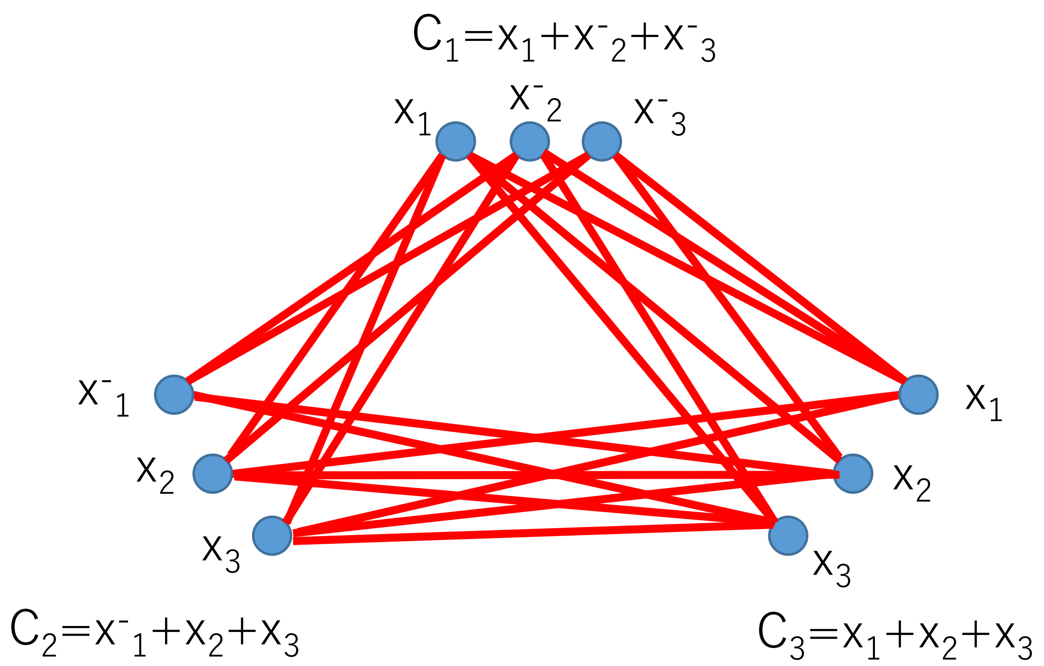
から9個の点と21本の辺を持つグラフが描ける。



NP完全であることの証明の例

もし、3SATのインスタンスがYES なら, 対応するCLIQUEのインスタンスもYESであることを示そう。

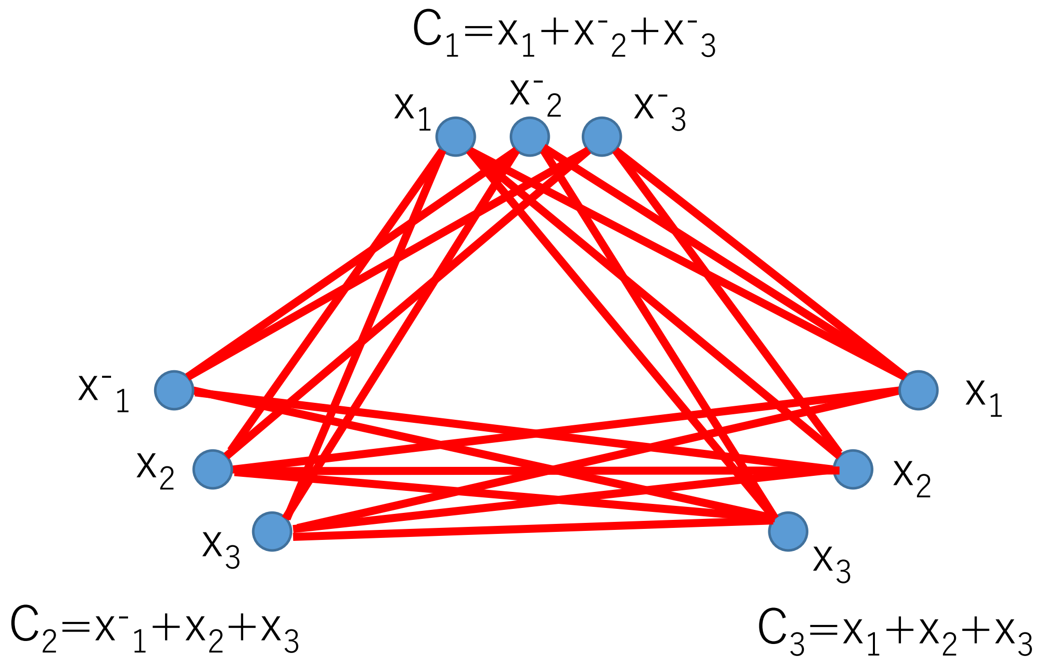
3SATのインスタンスがYESであるので、
すべての節を1とする変数への0/1の割り当てがある。
この割り当ての際に、各節には1のリテラルが
少なくとも1個はある。
各節から1のリテラルをひとつずつ選ぶと、
対応するグラフではk個の点が選べる。
グラフの作りかたからこのk点はクリークである。
なぜならば、
各節からリテラルをひとつずつ選んだこと、および、
k点は3SATの答えをもとに構成したので、
条件(1)と(2)より選択したk点中の任意の2点を結ぶ辺が
あるからである。



NP完全であることの証明の例

逆に、もし
CLIQUEのインスタンスがサイズkのクリークを持つ(YES)ならば、
対応する3SATのインスタンスはYESであることを示そう

CLIQUEがサイズkのクリークを
持つとしよう。
このとき、このクリークのk点は、
各節に対応した3点のうち、
ちょうど1点だけを含む。
(同じ節に属する2つのリテラル間には、辺がないため)
また、互いに否定となるリテラル間には辺がない。
よって、このクリークのk点は
3SATのすべての節を1にするような
変数への0,1の割り当てに対応する。



今回の講義の概要

計算可能な問題は次の4種類に分類できる。

(1)　多項式時間アルゴリズム(計算時間が多項式のオーダーのアルゴリズム)が知られている問題 (クラス P)(Pは、polynomial多項式の意味)
ソート・サーチ・グラフの最短経路・平面グラフの判定・最小木・線形計画法[Karmarkar,1984]・......

(2)　指数関数時間かかることが判明しているもの
グラフの全ての部分木の発生・....

(3)　解の検証は多項式時間でできるが、解を求める多項式時間アルゴリズムは知られていないもの
(クラス NP完全 (NP-Complete))
充足可能性問題・独立点集合問題・ハミルトン閉路問題・.....

(4)　上記3つのどれにはいるか、現在不明のもの
グラフ同形問題(graph isomorphism)・素因数分解(prime factoring)・...