به نام خداوند بخشنده ی مهربان

دانشگاه صنعتی شریف
دانشکده علوم ریاضی

عنوان ارائه:

پیچیدگی حالت میانگین

(فصل ۱۸کتاب computational complexity, Arora and Barak)

ارائه دهنده:

فاطمه احمدی موغاری

شماره دانشجویی:

۹۳۲۰۰۸۴۱

استاد مربوطه:

دکتر فروغمند اعرابی

بهار **۹۴**

مقدمه

در فصل های دیگر تنها به پیچیدگی در بدترین حالت پرداختیم. در پیچیدگی بدترین حالت، تمام ورودی ها را در نظر می گیریم در حالی که در پیچیدگی حالت میانگین تنها به ورودی هایی توجه می کنیم که در عمل رخ می دهند. گاهی ارزیابی الگوریتم در بدترین حالت، محک خوبی برای کارایی الگوریتم نیست؛ زیرا ممکن است بدترین حالت الگوریتم، احتمال رخداد

کمی در واقعیت داشته باشد. اما در پیچیدگی حالت میانگین به هر ورودی به میزان احتمال وقوع آن در واقعیت، بها می دهیم.

برای مثال در گراف های n راسی که احتمال حضور هر یال با انداختن یک سکه متعادل مشخص می شود بسیاری از مسائل NP-complete بسیار راحت حل می شوند. برای مثال مسئله ی سه رنگ پذیری با احتمال بالایی در زمان خطی و مسائل خوشه و مجموعه ی مستقل راسی در زمان قابل حل هستند.

از انگیزه های مطالعه ی پیچیدگی حالت میانگین می توان به موارد زیر اشاره کرد:

- احتمال کم رخ دادن ورودی های بد
- تولید ورودی هایی سخت برای مسائل رمزنگاری
- یافتن کاراترین الگوریتم برای یک مسئله از بین چند الگوریتم با پیچیدگی یکسان

در واقع از دیدگاه پیچیدگی حالت میانگین الگوریتمی کارا است که روی ورودی های بد، احتمال کمی داشته باشد

1. مسائل توزیعی و کلاس distP

برای بررسی پیچیدگی حالت میانگین باید بدانیم ورودی ها با چه احتمالی رخ می دهند پس تعریف مسئله، به تعریف توزیع وابسته است. علاوه بر این با توجه به تاکید پیچیدگی حالت میانگین بر واقعی بودن ورودی ها، تنها توزیع هایی در این مقوله مطرح هستند که در جهان اطراف ما قابل رخ دادن باشند یعنی حداقل بتوان در زمان چند جمله ای از توزیع نمونه گیری کرد

تعریف: (مسئله ی توزیعی)
یک مسئله ی توزیعی زوج است که L یک زبان و دنباله ای توزیع های (توزیع روی ورودی های به طول n) است.

همانند پیچیدگی بدترین حالت، در پیچیدگی حالت میانگین نیز دو کلاس و distNP معادل Pو NP تعریف می کنیم:

تعریف: (چندجمله ای در حالت میانگین و)
زمان اجرای الگوریتم A روی ورودی است و مسئله ی توزیعی را عضو کلاس می نامیم هر گاه الگوریتم A برای پذیرش زبان L و ثابت های وجود داشته باشد که .

2. توزیع های دنیای واقعی

توزیع های قابل رخداد در طبیعت دو نوع هستند:

- *P-computable*: توزیع هایی که با یک ماشین تورینگ قطعی در زمان چند جمله ای بتوان توزیع تجمعی آن را محاسبه کرد.

- *P-samplable*: توزیع هایی که با یک ماشیتن تورینگ احتمالاتی در زمان چند جمله ای بتوان نمونه های آن را تولید کرد..

هر توزیع P-computable، یک توزیع P-samplable نیز هست اما عکس آن صحیح نیست (مگر آن که).

بسیاری از توزیع های طبیعی، P-computable هستند اما اغلب، خروجی توابع مولد شبه تصادفی، قابل محاسبه در زمان چند جمله ای نیستند اما P-samplable هستند.

3. کلاس distNP و مسائل distNP-Complete

تعریف: (کلاس distNP)
مسئله ی توزیعی را عضو کلاس distNP می نامیم هر گاه و D یک توزیع P-computable باشد

حال به تعریف تحویل مسائل توزیعی می پردازیم:

تعریف: (تحویل حالت میانگین)
مسئله ی توزیعی به مسئله ی توزیعی تحویل حالت میانگین می یابد هرگاه تابع محاسبه پذیر چندجمله ای و توابع چندجمله ای وجود داشته باشند که در شروط زیر صدق کنند:
1- (صحت) برای هر : 2- (قاعدگی طول) برای هر : 3- (غلبه) برای هر و هر عدد طبیعی :

شرط اول، صحت تحویل و شرط دوم، خاصیت تعدی تحویل را تضمین می کنند و علت وجود شرط سوم این است که از پیش آمدن شرایطی که ورودی هایی از که احتمال کمی دارند با احتمال زیادی در با توجه به توزیع D طاهر شوند، جلوگیری کند.

قضیه: اگر و باشد آن گاه .

3.1.یک مسئله ی distNP-Complete

مسئله ی توزیعی را distNP-Complete گویند هر گاه یک مسئله ی distNP باشد و هر مسئله ی عضو کلاس distNP به آن تحویل شود ().

تعریف: (وجود یک مسئله ی distNP-Complete)
زبان را مجموعه ی هایی در نظر بگیرید که ماشین تورینگ غیر قطعی روی ورودی بعداز t گام خروجی 1 دهد.
توزیع بر روی ورودی های به طول را به صورت زیر در نظر بگیرید:
کدینگ به صورت تصادفی از رشته های به طول حداکثر $logn$ و t به صورت تصادفی از مجموعه ی و x به صورت تصادفی از انتخاب شود.
این توزیع در زمان چند جمله ای قابل محاسبه است.
بنابراین، مسئله ی یک مسئله ی distNP-Complete است.

ایده ی اثبات: مسئله ی با تحویل بدیهی یک زبان NP-Complete است اما لزوما این تحویل در شرایط تحویل میانگین صدق نمی کند (به دلیل عدم برآورده ساختن شرط سوم). در واقع نباید دارای قله باشد(منظور از قله ورودی هایی با طول و احتمال بالاتر از است که وقتی به تحوریل می یابند در احتمالی کمتر از داشته باشند). بنابراین، ابتدا باید قله ها را رفع کنیم.
برای هر توزیع P-computable می توان تابع محاسبه پذیر چند جمله ای را تعریف کرد که یک به یک باشد و و باشد.
اصطلاحا به این تابع، تابع رفع قله می گویند.
حال با استفاده از تابع رفع قله ، ماشین تورینگ غیر قطعی را به این صورت تعریف می کنیم: روی هر ورودی ، رشته ی را حدس بزن که و سپس را به عنوان خروجی بده.

برای تحویل هر عضو کلاس distNP به هر ورودی از را به نگاشت می هیم که k یک چند جمله ای برحسب طول توصیف و طول است.

نکته: اگر الگوریتمی کارا در حالت میانگین برای مسئله ی وجود داشته باشد، آن گاه برای هر زبان و هر توزیع D که P-samplable باشد الگوریتمی کارا کارا در حالت میانگین برای مسئله ی وجود دارد.

3.2.توزیع های P-samplable

مجموعه ی sampNP را مجموعه ی مسائل توزیعی ای در نظر بگیرید که گاه و D یک توزیع P-samplable باشد. مسئله ی را sampNP-Complete گویند هر گاه یک مسئله ی sampNP باشد و هر مسئله ی عضو کلاس sampNP به آن تحویل شود ().

قضیه: اگر یک مسئله ی distNP-Complete باشد آن گاه sampNP-Complete نیز هست.

4.مفاهیم فلسفی و عملی

Impagliazzo پنج سناریو برای جهان واقع ارائه کرده است:

- *Algorithmica* : جهانی که باشد یا تقریبا معادل باشند . در این جهان الگوریتمی جادویی برای حل مسئله ی SAT در زمان خطی وجود دارد. این جهان، مدینه ی فاضله ی محاسباتی است که می توان در آن، خلاقیت را اتوماتیک کرد و طرح های رمزنگاری را شکست.

- *Heuristica* : جهانی که باشد اما باشند. در این جهان الگوریتم های خوبی برای حل حداکثری مسائل هست یعنی برای هر مسئله ی NP می توان روی اغلب داده ها، خروجی را در زمان چند جمله ای بدست آورد؛ ممکن است ورودی هایی باشند که این الگوریتم ها زمان طولانی ای برای خروجی دادن صرف کنند اما این نوع ورودی ها به ندرت اتفاق می افتند.
  در این جهان بعضی ویژگی های برقرار است از جمله شکستن طرح های رمزنگاری و یافتن کوتاهترین اثبات ها. اما بعضی از ویژگی های نیز برقرار نیست از جمله آن که PH به P فرو نمی ریزد.

- *Pessiland* : بدترین جهان ممکن است که در آن نه در هستند و نه تابع یک طرفه وجود دارد. یعنی نه الگوریتم های جالب برای حل کامل یا حل حداکثری مسائل NP هست و نه می توان رمزنگاری کرد.

- *Minicrypt* : جهانی که در آن تابع یک طرفه وجود دارد (بنابراین) و مسائل خوش ساخت NP از جمله *factoring* در زمان چند جمله ای قابل حل هستند. چون تابع یک طرفه وجود دارد پس هیچ طرح رمزنگاری با کلید عمومی وجود ندارد اما از طرفی بسیاری ار کاربردهای رمزنگاری با کلید اختصاصی از جمله امضاهای دیجیتالی، تابع های مولد شبه تصادفی با تابع یک طرفه قابل انجام هستند.

- *Cryptomania* : جهانی که در آن مسئله ی *factoring* روی ورودی های بزرگ حتی در حالت میانگین، نمایی است. بسیاری از محققان بر این باورند که این جهانی است که ما در آن زندگی می کنیم و برای

حل مسائل مهم *NP* به روش های تقریبی، هیوریستیکی و خلاقانه و نیاز داریم. این جهان میزبان طرح ها و کابردهای جالب و فراوان رمزنگاری است.

کم کردن تعداد این جهان های احتمالی، دغدغه ی اصلی پیچیدگی محاسباتی است.

5.تاریخچه و افراد فعال در زمینه ی پیچیدگی در حالت میانگین

مطالعه ی پیچیدگی حالت میانگین از سال 1970 و در خلال مطالعات مربوط به رمزنگاری و رویکردهای هیوریستیک برای زبان های *NP* و به هدف یافتن پاسخ برای سوالات حوزه ی رمزنگاری آغاز شد. از آن جایی که امنیت پروتکل های رمزنگاری برمبنای آن است که مسائلی مانند *factoring* حتی در حالت میانگین هم الگوریتم کارامدی ندارند، بنابراین مفاهیم مربوط به پیچیدگی حالت میانگین در مقاله های رمزنگاری، شبه تصادفی، الگوریتم های هیوریتیک و ... نیز به چشم می خورد.

برای اولین بار Knuth در سال 1972 مفهوم پیچیدگی حالت میانگین را به طور رسمی در جلد سوم "The Art of computer programming" بیان کرد و Leonid Levin ساختاری کلاسیک برای آن طراحی کرد، کلاس های مختلف آن را تعریف و مسئله ی distNP-Complete معرفی کرد. Impagliazzo نیز در توسعی مفاهیم پیچیدگی حالت میانگین تلاش های بسیاری کرده است. امروزه نیز افرادی نظیر Fortnow, Bogdanov, Antunes, Trevisan و ... در این حوزه پژوهش می کنند.

یکی از مسائل باز در این حوزه که هم چنان حل نشده است یافتن مسئله ای طبیعی (و نه ساختگی) است که distNP-Complete باشد.