# اشاره

# تجزیه و تحلیل فضایی خوشه‌بندی[1] براساس الگوریتم ژنتیک

مهدی مدیری
عضو هیأت علمی دانشکده مقشه‌برداری
mmodiri@ut.ac.ir

## چکیده

به منظور تجزیه و تحلیل فضایی خوشه‌بندی، از اصول و مشخصه‌های الگوریتم‌های ژنتیکی استفاده می‌شود. مقاله حاضر روش جدید تجزیه و تحلیل فضایی خوشه‌بندی براساس الگوریتم ژنتیکی را ارائه می‌نماید. نتایج تجربیات علمی و عملی نشان می‌دهد این روش می‌تواند مشخصه کلی توزیع را حفظ کرده و نتیجه مناسبی را اخذ نماید.

واژه‌های کلیدی: الگوریتم ژنتیک، خوشه‌بندی، تجزیه و تحلیل فضایی.

## مقدمه

تجزیه و تحلیل فضایی یک جنبه مشترک برای کارشناسان علوم زمین را فراهم می‌نماید. با پیدایش نقشه، افراد همیشه انواع راه‌حل‌های تجزیه و تحلیل فضایی را به طور خودآگاه یا ناخواسته در ذهن خود دارند. به عنوان مثال، برای اندازه‌گیری فاصله، آزیموت بین دو عارضه در نقشه و انجام تحقیقات و بررسی فنی و تصمیم‌گیری راهبردی با استفاده از نقشه انجام می‌یابد. برای اینکه بتوان تجزیه و تحلیل فضایی را انجام داد و اطلاعات فضایی را نیز انتقال داد، لازم است یک مشخصه عملیاتی اولیه را از سیستم اطلاعات جهانی به دست آورد. *(Girod, 2000)*

خوشه بندی یک جنبه با اهمیت در تجزیه و تحلیل فضایی می‌باشد به طوری که متغیرهای فضایی و مشخصه‌های عوارض فضایی را از نقطه نظر کلی و عمومی بررسی می‌نماید و اطلاعات توجیهی خوشه‌بندی را به نمایش می‌گذارد. هدف تجزیه و تحلیل ویژگی‌های بارز عوارض فضایی است و در خوشه‌بندی فضایی یک خوشه[2] به چندین خوشه مختلف تقسیم می‌شود تا از این طریق بتوان بعضی از مشخصه‌های جغرافیایی را استخراج نمود یا این که بتوان اصول تجزیه و تحلیل را انجام داد. اصولاً تجزیه و تحلیل خوشه‌ای یک نوع کلی از الگوریتم‌های بهینه‌سازی می‌باشد. الگوریتم‌های ژنتیک (GA)[3] براساس گزینش طبیعی یا متعارف[4] و نظریه ژنتیک می‌باشد که می‌توان در این زمینه تحقیقات گسترده و فراوانی را صورت داد. (ZHOU MING, 1999)

براساس تجزیه و تحلیل الگوریتم‌های موجود در تجزیه و تحلیل فضایی خوشه‌ای قوانین اصلی تقسیم و مشخصه‌های الگوریتم‌های ژنتیک در نظر گرفته می‌شود. در این جا روش جدید تجزیه و تحلیل فضایی خوشه‌ای براساس الگوریتم‌های ژنتیک ارائه می‌شود.

## معرفی روش‌های خوشه‌بندی فضایی

تجزیه و تحلیل خوشه‌بندی فضایی، یک نوع خوشه‌بندی می‌باشد که براساس موقعیت هندسی داده‌های فضایی است. بنابراین مشخصه خوشه‌بندی فضایی نقاط براساس این قالب است. معمولاً فاصله فضایی بین نقاط اصلی‌ترین مقدار آماری می‌باشد و می‌تواند مشخصات خوشه‌ای نقاط را مشخص سازد. نقاطی که در فاصله نزدیکی قرار دارند به عنوان یک خوشه و در غیر این صورت به عنوان خوشه‌های مختلف در نظر گرفته می‌شوند. براین اساس در خوشه‌بندی فضایی، قاعده‌ای که به صورت عمومی و مشترک استفاده می‌شود، اندازه فواصل است که در این بحث استفاده می‌گردد. (Bader, 1997)

در خوشه‌بندی فضایی می‌توان روشهای مختلفی راانتخاب و استفاده کرد که این روشها به سه دسته، خوشه‌بندی سیستماتیک،(۵) خوشه‌بندی تجزیه مرحله‌ای (۶)، خوشه‌بندی توزیعی(۷) تقسیم می‌شوند.

## خوشه‌بندی سیستماتیک

ابتدا نقاط به عنوان خوشه‌های n در نظر گرفته می‌شوند. سپس این نقاط به صورت مرحله‌ای به هم تلاقی داده می‌شوند. در چنین مراحل خوشه‌بندی، تعداد خوشه‌ها کمتر می‌گردد تا این که به یک مقدار مناسب برسند. این مراحل، خوشه‌بندی سیستماتیک نامیده می‌شوند. مرحله ایدئوگرافیک نیز ابتدایی‌ترین روش می‌باشد. به طوری که دو خوشه که فاصله بین آنها به هم نزدیک است در یک خوشه جمع می‌شوند. ابتدا خوشه‌های n-۱ به دست می‌آید. سپس فواصل بین این خوشه‌ها n-۱ باید مجدداً محاسبه گردد. از بین خوشه‌ها دو خوشه‌ای که فاصله آنها به یکدیگر بسیار نزدیک است، انتخاب می‌شود و بعد از آن می‌توان خوشه‌های n-۲ را به دست آورد. تا اینکه تعداد خوشه‌ها به یک مقدار کامل برسد. مقدار عددی و آماری خوشه‌هایی که در مراحل ترکیب هستند باید دوباره محاسبه گردند و فواصل بین سایر خوشه‌ها نیازی به محاسبه مجدد ندارد.

بنابراین کوانتم محاسبه بسیار کوچک است و فرآیندها نیز ساده می‌باشند. روش خوشه‌بندی، روش سنجیده و حساب شده‌ای است که به طور گسترده‌ای استفاده می‌شود. (Hongyan, 2007)

## خوشه‌بندی تجزیه مرحله‌ای

در روش خوشه‌بندی تجزیه مرحله‌ای، ابتدا نقاط n به عنوان یک خوشه فرض می‌شوند سپس می‌توان آنها را به صورت مرحله‌ای به یکدیگر تلاقی داد. بنابراین با این روش تعداد خوشه‌ها بیشتر و بیشتر خواهد شد تا این که به یک مقدار مناسب دست یافت. این مراحل خوشه‌بندی، تجزیه مرحله‌ای نامیده می‌شود. بسیاری از جنبه‌ها باید به عنوان مرحله به مرحله در نظر گرفته شود. این نوع روش به اندازه کافی کامل و تکامل یافته نیست. بنابراین به طور گسترده‌ای مورد استفاده قرار نمی‌گیرد. (ZHOU MING, 1999)

## خوشه‌بندی توزیعی

ابتدا باید آمپتی(۸) تعیین گردد، سپس فواصل بین هر نقطه و مراکز نیز باید تعیین شود که این روش خوشه‌بندی توزیعی نامیده می‌شود. مسئله اصلی این روش عبارت است از: اینکه چگونه مراکز را تعیین نمود؟ متخصصان زیادی پیرامون این روش بررسی ومطالعه کرده و به طور مداوم این روش‌ها را مورد تغییر و اصلاح قرار داده‌اند و روش خوشه‌بندی توزیعی را پذیرفته و به طور گسترده‌ای مورد بهره‌برداری قرار گرفته است. خوشه‌بندی سیستماتیک و تجزیه مرحله‌ای براساس بررسی‌های موردی استفاده می‌گردد.

در این مراحل مشخصه کلی توزیع در نظر گرفته نمی‌شود. مراحل خوشه‌بندی توزیعی، در صورتی که مراکز خوشه‌بندی مطابق با مشخصه‌های کلی توزیع باشند، خروجی مشخصه‌ها حفظ می‌شوند. اما این روش به انتخاب مراکز بستگی دارد. بنابراین باید روش جدیدی پیدا نمود تا بتواند مشخصه‌های کلی توزیع را داشته باشد و برای پیدا نمودن مراکز خوشه‌بندی، روش مناسب برگزید.

## اصول و مشخصه‌های اصلی GA

الگوریتم‌های ژنتیکی ابتدا توسط پروفسور ت، هُلند[۹] مطرح شد؛ زمانی که او و دانشجویانش سیستمهای تطبیقی را در اواخر دهه ۱۹۵۰ و اوایل ۱۹۶۰ میلادی مورد مطالعه قرار دادند و بعد از آن‌ها نیز این مسئله به عنوان روش بهینه مؤثر مورد مطالعه و بررسی قرار گرفت. GA اصول ارزیابی بیولوژیکی، فناوری بهینه‌سازی و فناوری رایانه را با هم ادغام می‌نماید و یک رشته کاربردی جدید را فراهم می‌سازد. GA سه اپراتور اصلی، انتخاب، گذر و جهش را دارد که به عنوان الگوریتم‌های بهینه‌سازی عددی نامگذاری می‌گردند. این الگوریتم‌ها بر بعضی از معایب الگوریتم‌های سنتی غلبه می‌نماید. GA شامل کلیه قوانین و پیشنهادهای بیولوژیکی و وراثت است که دارای مزایای قابل توجهی می‌باشد به طوری که از الگوریتم‌های بهینه‌سازی سنتی مجزا می‌گردند. (ZHOU MING 1999)

۱) GA از کدبندی مقادیر قابلیت تصمیم‌گیری که به عنوان عملکرد پدیده می‌باشد، استفاده می‌نماید و این مطلب می‌تواند برای بعضی از مشکلات بهینه‌سازی که مفاهیم عددی را ندارد یا برای استفاده از آنها مشکل دارند، مفید باشد.

۲) GA عملکرد اصلی را به عنوان اطلاعات مختصر، به طور مستقیم استفاده می‌نماید.

۳) GA اطلاعات چند منطقه را استفاده می‌کند.

۴) GA از فناوری احتمالات استفاده می‌نماید.

GA می‌تواند یک ساختار چندمنظوره را ارائه نماید که برای حل سیستم‌های پیچیده استفاده می‌گردد و هیچ گونه بستگی به زمینه ایدئوگرافیک مشکلات ندارد و مشخصه انتقالی بارزی را دارد. بنابراین در بسیاری از موضوعات به طور گسترده‌ای استفاده می‌شود. در این مقاله GA با خوشه‌بندی توزیعی ترکیب می‌شود و روشی را ارائه می‌کند تا بتوان تجزیه و تحلیل خوشه‌بندی را براساس الگوریتم ژنتیکی انجام داد.

## تجزیه و تحلیل فضایی خوشه‌بندی براساس GA

با ترکیب عناصر GA و خوشه‌بندی توزیعی می‌توان راه اصلی را برای خوشه‌بندی توزیعی و انتخاب مراکز خوشه‌بندی تعیین نمود و از GA مشخصه‌های کلی را برای تحقیق استفاده کرد. بنابراین GA برای پیدا کردن مراکز خوشه‌بندی مورد استفاده قرار می‌گیرد که می‌تواند مشخصه‌های کلی را به طور اتوماتیک در خود نگه دارد. سپس از روش خوشه‌بندی توزیعی برای بررسی سایر نقاط استفاده نمود. این مراحل بایستی تا مرحله پایانی انجام گیرد که با مشخصه‌های کلی توزیع مطابقت نماید. (Weibel. 1995)

چندین مشکل دیگر نیز وجود دارد که باید برطبق موارد ذیل مرتفع گردد.

### کدگذاری

کدگذاری اولین مشکلی است که باید با استفاده از GA حل گردد و یک مرحله اصلی برای طراحی

GA می‌باشد. علاوه بر روشهای کدگذاری که شکل آرایشی هر کروموزم را نشـان مـی‌دهد. روش کدگذاری مجدد را معرفی می‌نماید که از ژن فضای تحقیقی به فضای توضیحی و تشریحی انتقال یافته‌اند و بر روش اپراتورانتخاب، جهش و گذر اثر می‌گذارند. زمانی که مراکزخوشه‌بندی از نظر تعاملی بررسی شود که به انجام تجزیه و تحلیل خوشه‌بندی نیاز دارد یک نقطه می‌تواند هم مرکز باشد یا مرکز نباشد. بنابراین روش کدگذاری برای ارائه خروجی استفاده می‌شود. در صورتی که سـاختار کـدگذاری بسـیار پیچیده باشد یعنی تعداد نقاط بسیار زیاد می‌باشد و محاسبه نیز بسیار پیچیده می‌گردد. بنابراین در مراحل کدگذاری که براساس اصل ساده‌سازی و آسان‌سازی می‌باشد از کد دوتایی استفاده می‌شود که می‌تواند به راحتی کدبندی و کدگذاری مجدد شود. روش کدگذاری ایدئوگرافیک نیز به شرح ذیل معرفی می‌گردد به طوری که هر چه تعداد نقاط بیشترباشد طول کدگذاری نیز بر همین رول مشخص می‌گردد که نشان می‌دهد نقطه در مرکزخوشه‌بندی انتخاب شده است. در صورتی که مقدار ۱ باشد، نقطه انـتخاب مـی‌شود در صورتی که ۰ باشد، نقطه انتخاب نمی‌شود. به عنوان مثال: در صورتی که مجموعه نقاط P {P1, P2, P3, P8, P5, P4, P1...P8 } باشد، نشان می‌دهد که P8, P5, P4, P1 مراکزخوشه‌بندی هستند. (Axelsson, 1999)

## - انطباق عملکرد

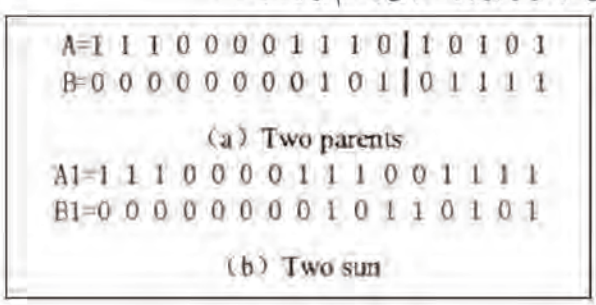
در GAاحتمال ژنتیکی فرد به وسیله تطبیق فردی قطعی می‌گردد. یعنی در صورتی که تطبیق وسیع باشداحتمال ژنتیکی نیز زیاد می‌شود. چگونه می‌توان عملکرد تطبیق را قطعی ساخت تا تأثیر زیادی را بر قابلیت GA داشته باشد؟ جواب این پرسش به این صورت است که SUM (P) تعداد مـراکـزانـتخابی خوشه‌بندی می‌باشد و می‌توان تابع تطبیق را به صورت زیر طراحی نمود.

$$F(t) = SUM\ (P) \qquad (۱)$$

که SUM(P) تعداد مراکزانتخابی خوشه‌بندی می‌باشد و با در نظر گرفتن دقت، انتظار است که تعداد SUM (P) کمتر شود.

## اپراتورهای GA

اصولاً اپراتورهای GA، انتقال، گذر و جهش می‌باشد. درحال حاضر، یک اپراتورانتقال و گذر (نگاره ۱) تطبیق می‌یابد. اما نمی‌توان اپراتور جهش را تطبیق داد ولی از گزینه بهینه‌سازی کلی استفاده می‌شود. این عملکرد نشان می‌دهد که به هر کروموزوم باید تنها یک راه حل منطقی تعلق گیرد. در این صورت مـی‌توان تـضمین نـمود کـه هـر نـقطه‌ای در مـرکزیت بـه یکـی از نـقاطی کـه در مـرکزخـوشه‌بندی است، محول می‌گردد و فاصله بین دو مرکز نیزنباید خیلی به هم نزدیک باشد.




نگاره ۱

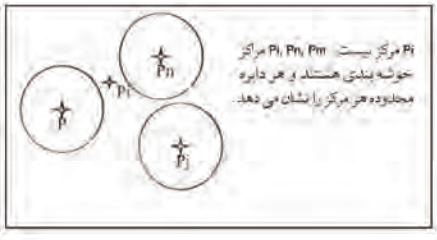
## آرایش کدگذاری مجدد کروموزوم

ابتدا کروموزمی که از طریق گزینش بهینه‌سازی کلی شده است کدگذاری مجدد می‌شود. مراکز خوشه‌بندی نیز از نقاط اصلی انتخاب می‌گردد.

## انتخاب مرکز خوشه‌بندی

بر طبق موقعیت واقعی هر مرکزخوشه‌بندی، فقط باید یک مرکزخوشه‌بندی وجود داشته باشد (نگاره ۲). با فرض Pi و Pj که هر دو مراکزخوشه‌بندی می‌باشندچنانچه فاصله بین آنها کوتاهتر ازحد مجازباشد، نشان می‌دهد که Pi و Pj بسیار به یکدیگر نزدیک هستند. تعداد Pi از ۱ به ۰ تغییر می‌نماید. و Pi مرکز خوشه‌بندی نخواهد بود.


نگاره ۲




نگاره ۳

## بررسی سایر نقاط

در نظر گیرید که هر نقطه‌ای که در مرکز نباشد، در این صورت توضیحی را برای خود دارد. در صورتی که Pi متعلق به هر مرکزخوشه‌بندی نباشد سپس بر طبق قانون فاصله، Pi مرکزخوشه‌بندی می‌شود و مقدار کروموزوم نیز به ۱ تغییر می‌نماید (نگاره ۳)

## جایگزینی کروموزوم

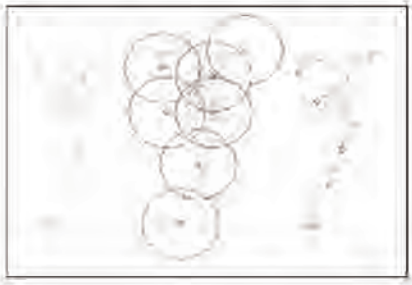
کروموزوم قدیمی از طریق کروموزوم بازسازی شده جایگزین می‌گردد و در صورتی که GA با گزینه بهینه‌سازی کلی ارتباط داده شود، در این صورت کارایی GA به طور قابل توجهی افزایش می‌یابد و زمان اتلافی نیز در این راه حل وجود ندارد و می‌تواند به طور قابل توجهی بر GA تأثیر داشته باشد. (Hongyan,2007)

## پارامترهای GA

پارامترهای متعارف GA میزان جمعیت (M)، مقدار گذر Pc و احتمال جهش Pm را شامل می‌شود.

این پارامترها تأثیر زیادی بر GA دارد و بایستی با دقت زیادی انتخاب شوند. اصولاً این پارامترها از طریق آزمایش و تجربه کسب می‌گردند.

در این بررسی، اندازه جمعیت = 50 ,Pc=0 ,Pm=0 منظور شده است. (Bader, 1997)



نگاره ۴

## اپراتورهای انتخاب

در طی تجزیه و تحلیل بیولوژیکی وراثت و تکامل طبیعی، نمونه‌هایی که می‌توانند به تکامل منطبق شوند، احتمال بیشتری را برای وراثت خواهند داشت. در غیر این صورت آنها احتمالات کمتری را دارند. GA از اپراتور انتخاب برای انجام مراحل استفاده می‌نماید که این روش بهترین روش در شرایط فعلی می‌باشد که برای تولید مجدد و بقاء تطبیق یافته است.

اپراتورهای انتخاب GA مختلف و معمولاً اپراتورهایی هستند که اپراتور تصادفی، اپراتور دوره‌ای و غیره می‌باشند. در این مقاله از اپراتور تورنمنت استفاده شده است. این روش انتخاب باعث گردید فردی که تطبیق‌سازی بیشتری دارد شانس بیشتری را نیز برای بقا و زندگی داشته باشد و فقط از مقدار تطبیق استفاده نموده که با اندازه مقدار انتخابی تناسب ندارد. (Hongyan, 2007)

## شرایط تکمیل نشده در GA

در اینجا دو شرط تکمیل نشده در GA وجود دارد. یکی از شرایط تعداد تولید ژنتیک است که نشان می‌دهد در صورتی که GA به تولید ژنتیک برسد، GA متوقف می‌شود و کار ناتمام می‌ماند. شرط دیگر، این مسئله می‌باشد که میانگین تطبیق با بزرگترین تطبیق مقایسه می‌گردد. در صورتی که اختلاف بین میانگین تطبیق و بزرگترین تطبیق در یک حد مجاز مشخص شده باشد، در این صورت GA متوقف می‌گردد و در صورتی که مراحل ادامه یابد GA نمی‌تواند خروجی بیشتری را به دست آورد. ترکیب دو شرط بالا نیز می‌تواند تأثیر مثبتی را برروی GA داشته باشد. برطبق مراحل فوق این اصول محوری و کلیدی در ساختار GA حل می‌گردد. بر طبق این قاعده که هر نقطه مرکز خوشه‌بندی نمی‌باشد، مراحل خوشه‌بندی بر طبق قاعده فاصله انجام می‌گیرد.

## نتیجه‌گیری

این مقاله روش تجزیه و تحلیل فضایی را براساس GA فراهم می‌سازد و به صورت تئوری احتمال و امکان این نظریه را تجزیه و تحلیل می‌نماید و از نظر فنی آن را اعتبار می‌بخشد. تجربیات نشان می‌دهد که تجزیه و تحلیل فضایی خوشه‌بندی که براساس الگوریتم‌ها ژنتیکی می‌باشد خصوصیات کلی توزیع را نیز

حفظ می‌نماید. (نگاره ۴)

فقط تجزیه و تحلیل فضایی یک جنبه بسیار کوچک از کاربردهای GA در تجزیه و تحلیل فضایی می‌باشد. بسیاری از مشکلات تجزیه و تحلیل فضایی باید بیشتر مورد بررسی قرار گیرد و یا به شکل بهینه تغییر یابد. در این زمینه به تحقیقات گسترده‌ای نیاز می‌باشد.

## منابع و مآخذ


1- Axelsson, p., 1999. Processing of laser scanner data - algorithms and application. Journal of photogrammetry and Remote sensing 54 (2-3).

2- Bader, M. and Weibel, R., 1997, Detection and resolving size and proximity.

3- Girod, B., Greiner, G., Niemann, H, 2000. Principles of 3D Image Analysis and Synthesis. Kluwer Academic publishers, Dordrecht - Netherland.

4- Hongyan DENG, Fang WU, Chang YIN., 2007. The Spatial Analysis of Clustering based on Genetic Algorithms. Institute of surveying and Mapping, University, Zhengzhou, China.

5- Kim Lowell, Gold C, 1995. Using a Fussy surface - based cartographic Representation to Decease Digitizing Efforts for Natural Phenomena. Cartography and Geographic Information systems, vol. 22, No. 3, 1995.

6- Sagi Filin., 2004. Surface classification from airborne laser scanning data, Computers & Geosciences 30 (2004).

7- Weibel, R., 1995. Map generalization in the context of digital system. Cartography and Geographic Information System. Vol. 22, No. 4.

8- ZHOU, M, Sun Zedong, 1999. Genetic Algorithms Theory and Applications, Beijing, China.

9- ZHANG Wenliane, LIANG Yi, 2000. The Mathematic Base of Genetic Algorithms , xi'an , China.


## پی‌نوشت

۱) خوشه‌بندی به عنوان روش طبقه‌بندی، امکانی را فراهم می‌نماید که در یک روش بسیار طبیعی، تأکیداتی را در فرآیند دخالت داد. خوشه‌بندی را می‌توان به عنوان ترکیبی از دو فرآیند دانست:

الف) شناسایی الگوها در داده‌های مبتنی بر ویژگیهایی و

ب) گروه‌بندی داده‌ها در خوشه‌ها.

ویژگی اطلاعات را گردآوری می‌کند و در داده‌ها قرار داده و تفکیک و جدایی میان طبقه‌ها را اطمینان می‌دهد. خوشه‌بندی داده‌ها را می‌توان به عنوان یک تقسیم‌بندی جامع به قطعات مجزا، و هر یک با ویژگی همگن تعریف نمود. خوشه‌بندی داده‌ها بیشتر در طبقه‌بندی عوارض و پوشش زمین در تصاویر ماهواره‌ای یا تصویربرداری لیزری استفاده می‌شود. خوشه‌بندی داده‌ها را می‌توان به عنوان یک تقسیم‌بندی جامع به قطعات مجزا، هر یک با ویژگی همگن بکار برد فرض شود $R = \{Pi \mid i=1,2,...N\}$ داده‌ها با Pi نقاط اندازه‌گیری شده و N تعداد نقاط در مجموعه داده باشد و $Si = \{S1, S2, ..., Sk\}$ خوشه‌های داده‌ای باشد که در آن $Si = \{Pi1, Pi2, ..., PiNi\}$ قسمتهایی است که با گردآوری نقاط مشخص شده‌اند.

2) Cluster
3) Genetic Algorithms (GA)
4) Natural Selection
5) Systemic Clustering
6) Stepwise Decomposition
7) Distinguish Clustering
8) Umpty
9) T.Holland استاد دانشگاه میشیگان