

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

ده الگوریتم برتر در داده کاوی

نویسنده: سینگ دانگ وا، ویپین کومار

ترجمه:

کیوان معقولی

استادیار دپارتمان مهندسی پزشکی - دانشگاه آزاد اسلامی واحد علوم و تحقیقات

نادر جعفرنیا دابانلو

ریاست مرکز تحقیقات مهندسی در پزشکی و بیولوژی

سنا دلاوری

مرتضی زنگنه سروش

دکتری مهندسی پزشکی بیوالکتریک و عضو باشگاه پژوهشگران جوان و نخبگان، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران



عنوان و نام پدیدآور

ده الگوریتم برتر در داده‌کاوی/ نویسنده [صحیح: ویراستار] سینگ‌دانگ وا، وپین کومار؛ ترجمه کیوان معقولی... [و دیگران].

قم: آثار نفیس، ۱۳۹۹.

۳۴۵ص.

978-600-7964-68-2

فیا

وضعیت فهرست نویسی

یادداشت

عنوان اصلی: The top ten algorithms in data mining, c2009.

ترجمه کیوان معقولی، نادر جعفرنیا دابانلو، سنا دلاوری، مرتضی زنگنه‌سروش.

داده‌کاوی

یادداشت

Data mining

موضوع

الگوریتم‌های کامپیوتری

موضوع

Computer algorithms

موضوع

وو، شین‌دونگ، ویراستار

شناسه افزوده

Wu, Xindong

شناسه افزوده

کومار، وپین، ۱۹۵۶ - م، ویراستار

شناسه افزوده

Kumar, Vipin

شناسه افزوده

معقولی، کیوان، ۱۳۵۱ - ، مترجم

شناسه افزوده

Q89/V6

رده بندی کنگره

۰۰۶/۳۱۳

رده بندی دیویی

۷۳۰۰۲۳۳

شماره کتابشناسی ملی

نام کتاب: ده الگوریتم برتر در داده‌کاوی

ترجمه: کیوان معقولی، نادر جعفرنیا دابانلو، سنا دلاوری، مرتضی زنگنه‌سروش (باشگاه پژوهشگران

جوان و نخبگان، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران)

چاپ اول: ۱۳۹۹

ناشر: انتشارات آثار نفیس

شابک:

ISBN- 978-600-7964-68-2

تیراژ: ۵۰۰ جلد

قیمت: ۱۰۰۰۰۰ تومان



کلیه حقوق برای مؤلف/ ناشر محفوظ می‌باشد.

■ آدرس پستی: قم - انتشارات آثار نفیس - صندوق پستی ۳۷۱۶۵/۱۶۵

■ تلفن: دفتر: ۰۲۵۳۲۸۰۴۶۳۴ مدیر مسئول: ۰۹۱۲۵۵۲۴۰۰۷

■ ایمیل: asarenafisbooks@gmail.com

■ فروشگاه اینترنتی: www.nafis114.com و www.nafis114.ir

فهرست

صفحه	عنوان
۱	مقدمه
۵	تقدیر و تشکر
۷	درباره مولفین
۱۱	همکاران
۱۳	فصل ۱: C4.5
۱۳	۱-۱ مقدمه
۱۶	۱-۲ توصیف الگوریتم
۲۴	۱-۳ خصوصیات و ویژگی‌های C4.5
۲۴	۱-۳-۱ هرس کردن درختی
۲۸	۱-۳-۲ استفاده بهبود یافته از ویژگی‌های پیوسته
۲۹	۱-۳-۳ پردازش شاخص‌های غایب
۳۲	۱-۳-۴ القای دسته قوانین
۳۳	۱-۴ بحث در مورد اجرای نرم افزار موجود
۳۴	۱-۵ دو مثال نمایشی
۳۴	۱-۵-۱ دسته داده‌های گلف
۳۵	۱-۵-۲ دسته داده‌های Soybean
۳۷	۱-۶ موضوعات پیشرفته
۳۷	۱-۶-۱ استخراج از روی انبار ثانوی
۳۸	۱-۶-۲ درخت تصمیم‌گیری قائم

۳۹	 ۱-۶-۳ انتخاب ویژگی
۳۹	 ۱-۶-۴ روش های جمعی
۴۱	 ۱-۶-۵ قوانین دسته بندی
۴۲	 ۱-۶-۶ توصیف های مجدد
۴۳	 ۱-۷ تمرین
۴۶	 مراجع

فصل دوم: متوسط های k - ۵۱

۵۳	 ۱-۲ الگوریتم متوسط های k
۶۰	 ۲-۲ نرم افزار موجود
۶۲	 ۲-۳ مثال ها
۶۷	 ۲-۴ موضوعات پیشرفته
۷۱	 ۲-۵ خلاصه
۷۲	 ۲-۶ تمرین
۷۴	 مراجع

فصل ۳: ماشین های بردار پشتیبانی ۷۷

۷۷	 ۱-۳ کلاسه بندی بردار پشتیبانی (SVC)
۸۲	 ۲-۳ SVC با حاشیه ی نرم و بهینه سازی
۸۵	 ۳-۳ ترفند هسته ایی
۸۶	 ۱-۳-۳ تعریف (هسته ی محصول داخلی) [۷]
۸۷	 ۲-۳-۳ برهان
۹۰	 ۳-۳-۳ مثال (مثال تصویری)
۹۰	 ۴-۳-۳ مثال کاربرد واقعی - الگوریتم SVC
۹۳	 ۴-۳ نتایج نظری
۹۵	 ۱-۴-۳ برهان واپنیک و چروننکیس [۷]

۹۶	 ۳-۴-۲ تعریف (حاشیه).
۹۶	 ۳-۴-۳ تعریف (پوشش طبقه‌ی عبارتی) [۷].
۹۷	 ۳-۴-۴ برهان VC به همراه حاشیه [۷].
۹۸	 ۳-۴-۵ برهان مرز تعمیم SVC [۷].
۹۸	 ۳-۵- عقب گرد بردار پشتیبانی.
۹۹	 ۳-۵-۱ تعریف (عبارت غیر حساس به ϵ) [۷].
۱۰۱	 ۳-۶ اجرای نرم افزار.
۱۰۲	 ۳-۷ تحقیقات جاری و آیندم.
۱۰۲	 ۳-۷-۱ راندمان محاسبات.
۱۰۳	 ۳-۷-۲ انتخاب هسته.
۱۰۳	 ۳-۷-۳ تحلیل تعمیم.
۱۰۴	 ۳-۷-۱ تعریف (پیچیدگی رادماچر) [۳۵].
۱۰۴	 ۳-۷-۲ برهان تحلیل پیچیدگی [۳۵].
۱۰۵	 ۳-۷-۴ یادگیری SVM ساختاری.
۱۰۷	 ۸-۳ تمارین.
۱۰۹	 مراجع.

۱۱۳	 فصل ۴: Apriori
۱۱۳	 ۴-۱ مقدمه.
۱۱۴	 ۴-۲ توصیف الگوریتم.
۱۱۴	 ۴-۲-۱ استخراج الگوهای مکرر و قوانین وابسته سازی.
۱۱۵	 ۴-۲-۱-۱ Apriori.
۱۲۰	 ۴-۲-۲ AprioriTID.
۱۲۲	 ۴-۲-۳ استخراج الگوهای متوالی.
۱۲۶	 ۴-۲-۳-۱ All Apriori.

۱۲۷	۴-۲-۴ مباحثه.....
۱۲۹	۴-۴ مباحثه در مورد اجراهای نرم افزاری موجود :.....
۱۳۱	۴-۵ دو مثال نمایشی.....
۱۳۱	۴-۵-۱ مثال های کارآمد.....
۱۳۱	۴-۵-۱-۱ استخراج دسته آیتم مکرر و استخراج قوانین وابسته سازی.....
۱۳۹	۴-۵-۱-۲ استخراج الگوهای متوالی.....
۱۴۰	۴-۵-۲ ارزیابی عملکرد.....
۱۴۶	۴-۶ موضوعات پیشرفته.....
۱۴۷	۴-۶-۱ بهبود در استخراج الگوی مکرر از نوع Apriori.....
۱۴۹	۴-۶-۲ استخراج الگوی مکرر بدون ایجاد منتخب.....
۱۵۱	۴-۶-۳ رویکرد تصاعدی.....
۱۵۲	۴-۶-۴ نمایش متراکم سازی شده: الگوهای بسته و الگوهای حداکثری.....
۱۵۵	۴-۶-۵ قوانین وابسته سازی کمی.....
۱۵۶	۴-۶-۶ استفاده از دیگر معیار اهمیت / جدایت.....
۱۵۸	۴-۶-۷ قوانین وابسته سازی طبقه.....
۱۵۹	۴-۶-۸ استفاده از عبارت ریچر: توالی ها، درخت ها و نمودارها.....
۱۶۲	۴-۷ خلاصه.....
۱۶۳	۴-۸ تمرین.....
۱۶۵	مراجع.....
۱۷۱	فصل ۵: امید ریاضی - پیشینه سازی.....
۱۷۱	چکیده.....
۱۷۱	۵-۱ مقدمه.....
۱۷۴	۵-۲ شرح الگوریتم.....
۱۷۶	۵-۳ پیاده سازی نرم افزاری.....

۱۷۸.....	۴-۵ مثال‌های تشریحی.....
۱۷۸.....	۴-۵ مثال ۵-۱: توزیع نرمال چندمتغیره.....
۱۸۴.....	۴-۵ مثال ۵-۲: آمیخته‌های تجزیه کننده عامل.....
۱۸۹.....	۵-۵ مباحث پیشرفته.....
۱۹۲.....	۵-۶ تمرینات.....
۲۰۳.....	مراجع.....

فصل ۶: رتبه صفحه ۲۰۷

۲۰۷.....	۶-۱ مقدمه.....
۲۰۹.....	۶-۲ الگوریتم رتبه صفحه.....
۲۱۷.....	۶-۳ افزونه: رتبه صفحه‌ی زمان‌بندی شده.....
۲۱۹.....	۶-۴ خلاصه.....
۲۱۹.....	۶-۵ تمرین‌ها.....
۲۲۱.....	مراجع.....

فصل ۷: ادابوست ۲۲۳

۲۲۳.....	۷-۱ مقدمه.....
۲۲۵.....	۷-۲ الگوریتم.....
۲۲۵.....	۷-۲-۱ نکات.....
۲۲۶.....	۷-۲-۲ یک روش بوستینگ عمومی.....
۲۲۸.....	۷-۲-۳ الگوریتم ادابوست.....
۲۳۳.....	۷-۳ نمونه‌های تصویری.....
۲۳۴.....	۷-۳-۱ حل مسئله XOR.....
۲۳۶.....	۷-۳-۲ عملکرد در داده حقیقی.....
۲۳۸.....	۷-۴ کاربرد حقیقی.....
۲۴۱.....	۷-۵ موضوع‌های سطح عالی.....

۲۴۱	۷-۵-۱ مسائل نظری
۲۴۷	۷-۵-۲ ادابوست چنددسته‌ای
۲۵۲	۷-۵-۳ موضوعات سطح عالی دیگر
۲۵۳	۷-۶ اجرای نرم‌افزاری
۲۵۳	۷-۷ تمرین‌ها
۲۵۵	مراجع

فصل ۸: k: kNN نزدیک‌ترین همسایه ۲۵۹

۲۵۹	۸-۱ مقدمه
۲۶۱	۸-۲ توصیف الگوریتم
۲۶۱	۸-۲-۱ توصیف سطح بالا
۲۶۲	۸-۲-۲ مسائل
۲۶۵	۸-۲-۳ اجرای نرم‌افزاری
۲۶۵	۸-۳ مثال‌ها
۲۶۸	۸-۴ موضوعات سطح عالی
۲۷۰	۸-۵ تمرین‌ها
۲۷۱	تأییدیه‌ها
۲۷۲	مراجع

فصل ۹: Naive Bayes ۲۷۵

۲۷۵	۹-۱ مقدمه
۲۷۷	۹-۲ شرح الگوریتم
۲۸۱	۹-۳ قدرت با وجود استقلال
۲۸۴	۹-۴ توسعه مدل
۲۸۸	۹-۵ اجرای نرم‌افزار
۲۸۹	۹-۶ مثال‌ها

۲۸۹	۱-۶-۹ مثال ۱
۲۹۱	۲-۶-۹ مثال ۲
۲۹۳	۷-۹ مباحث پیشرفته
۲۹۵	۸-۹ تمرین ها
۲۹۷	مراجع
۲۹۹	فصل ۱۰: درخت های طبقه بندی و رگرسیون (CART)
۳۰۰	۱-۱۰ پیش نویس های CART
۳۰۱	۲-۱۰ مروری اجمالی
۳۰۲	۳-۱۰ یک مثال متداول
۳۰۷	۴-۱۰ الگوریتم به طور مختصر تثبیت شد
۳۰۸	۵-۱۰ قواعد شکافندگی
۳۱۲	۶-۱۰ احتمالات قبلی و توازن کلاس
۳۱۵	۷-۱۰ کنترل مقدار
۳۱۸	۸-۱۰ اهمیت مشخصه
۳۲۰	۹-۱۰ ساختار ویژگی دینامیک
۳۲۰	۱۰-۱۰ یادگیری حساس به هزینه
۳۲۲	۱۱-۱۰ قاعده توقف، هرس کردن، توالی های درخت و انتخاب درخت
۳۲۴	۱۲-۱۰ درختان احتمالات
۳۲۷	۱۳-۱۰ مبانی نظری
۳۲۸	۱۴-۱۰ تحقیقات مرتبط با مرحله بعد از CART
۳۳۰	۱۵-۱۰ در دسترس بودن نرم افزار
۳۳۱	۱۶-۱۰ تمرین
۳۳۳	مراجع

مقدمه

در تلاش برای تعیین چند الگوریتم تاثیر گذار که بطور گسترده‌ای در داده‌کاوی بکار گرفته می‌شوند، کنفرانس بین‌المللی IEEE در زمینه داده‌کاوی (ICDM, <http://www.cs.uvm.edu/~icdm/>), ده الگوریتم برتر در داده‌کاوی را برای ارائه در ICDM'06 هنگ کنگ مشخص کرد. این ده الگوریتم برتر داده‌کاوی عبارتند از: C4.5 و SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naïve Bayes, CART و k-Means که در این کتاب مورد بررسی قرار می‌گیرند.

در اولین گام از چنین تشخیصی، در سپتامبر ۲۰۰۶ برندگان جایزه‌ی نوین ACM, KDD و محققین IEEE ICDM را دعوت کردیم تا ده الگوریتم برتر در داده‌کاوی را نام ببرند. همه به غیر از یکی از آن‌ها به دعوت ما پاسخ دادند. از هر کدام خواستیم تا اطلاعات زیر را فراهم سازد:

الف/ نام الگوریتم

ب/ دلایل مختصر

ج/ یک مرجع چاپ شده به عنوان دلیل

همچنین توصیه نمودیم که هر الگوریتم کاندید شده باید به‌طور گسترده توسط دیگر محققان ارجاع گردیده باشد و در این زمینه بکار گرفته شده باشد و اگر الگوریتم‌های انتخاب شده توسط هر یک از این افراد را به‌عنوان یک گروه در نظر بگیریم، این گروه باید از زمینه‌های مختلف در داده‌کاوی نمایشی منطقی به همراه داشته باشد.

بعد از ارائه الگوریتم‌ها در مرحله ۱، میزان ارجاع به هر یک از این الگوریتم‌ها توسط جستجوی Google در اواخر اکتبر ۲۰۰۶ مورد تأیید قرار گرفت و الگوریتم‌های کاندید شده‌ای که کمتر از ۵۰ ارجاع داشتند را حذف کردیم. سپس بقیه الگوریتم‌های باقیمانده (۱۸ الگوریتم) در ۱۰ موضوع: تحلیل انجمنی^۱، کلاسه‌بندی و یا کلاسه‌بندی^۲، خوشه‌بندی^۳، یادگیری آماری^۴، bagging و boosting، الگوهای دنباله‌ای^۵، کاوش تلفیقی^۶، مجموعه‌های rough^۷، کاوش ارتباطات^۸ و کاوش گرافی^۹ سازمان دهی شدند. برای برخی از این ۱۸ الگوریتم، مثل k-means، مقاله معرف لزوماً مقاله اصلی برای معرفی الگوریتم نبوده است بلکه مقاله‌ای جدید که اهمیت تکنیک را مشخص می‌کند. این مقالات معرف در وب سایت ICDM^{۱۰} وجود دارند. در مرحله سوم فرآیند معرفی، با چالش بزرگ‌تری در جامعه تحقیقاتی روبرو گردیدیم. اعضای کمیته برنامه‌ریزی KDD-06 (کنفرانس بین‌المللی 2006 ACM SIGKDD درباره‌ی کشف دانش و داده‌کاوی)، ICDM06 (کنفرانس بین‌المللی 2006 IEEE درباره‌ی داده‌کاوی‌ها) و SDM06 (کنفرانس بین‌المللی 2006 SIAM درباره‌ی داده‌کاوی‌ها) و همینطور برندگان ACM KDD و برندگان تحقیق IEEE ICDM را دعوت کردیم تا از بین ۱۸ الگوریتم ۱۰ الگوریتم برتر را انتخاب کنند. نتیجه‌ی این رای‌گیری در میزگرد ICDM06 به صورت ۱۰ الگوریتم برتر داده‌کاوی، آورده شده است.

¹. Association analysis

². Classification

³. Clustering

⁴. Statistical learning

⁵. Sequential patterns

⁶. Integrated mining

⁷. Rough sets

⁸. Link mining

⁹. Graph mining

¹⁰. <http://www.cs.uvm.edu/~icdm/algorithms/CandidateList.shtml>

در میزگرد ICDM06 در ۲۱ دسامبر ۲۰۰۶ امکان اعلام رای آزادی را برای همه ۱۴۵ شرکت کننده به منظور معرفی ۱۰ الگوریتم برتر از بین ۱۸ الگوریتم موجود فراهم نمودیم و نتیجه این رای گیری با نتیجه حاصل از ۳ گام فوق یکسان بود. جلسه ی آخر کنفرانس ICDM06 به صورت یک میزگرد سه ساعته تنظیم گردید که به طور همزمان جلسات ارائه هفت مقاله از کنفرانس های هوش وب^۱ (WI'06) و تکنولوژی عامل هوشمند^۲ (IAT'06) در همان مکان ارائه می گردید و با این حال مورد توجه ۱۴۵ نفر از شرکت کنندگان قرار گرفت.

بعد از ICDM'06، از نویسندگان اصلی و برخی از افرادی که ۱۰ الگوریتم را در این میزگرد ارائه نموده اند، دعوت به عمل آوردیم تا یک مقاله ژورنال برای توصیف هر الگوریتم بنویسند و اهمیت الگوریتم را توضیح دهند و تحقیقات جاری و آتی پیرامون الگوریتم را بررسی کنند. مقاله در ژانویه ۲۰۰۸ در مجله ی دانش و سیستم های اطلاعاتی^۳ [۱] به چاپ رسید. این کتاب گسترشی بر مبنای آن مقاله نوشته شده است که دارای ساختار یکسانی در هر بخش برای هر یک از الگوریتم ها به صورت توصیف الگوریتم، نرم افزار موجود، مثال ها و کاربردهای نمایشی، موضوعات پیشرفته و تمیزها است. هر بخش کتاب توسط دو بازبین مستقل از هم و یکی از ویراستارهای کتاب بازبینی شده است. بعضی از بخش ها قبل از پذیرش نهایی طبق بازبینی ها کاملاً تجدید نظر شده اند.

امیدواریم تعیین ۱۰ الگوریتم برتر باعث شود که داده کاوی دارای کاربرد وسیع تری در دنیای واقعی گردد و محققان بیشتری را در این زمینه به تحقیقات افزون تر پیرامون این الگوریتم ها، تاثیرات آنها و مسائل تحقیقاتی جدید تشویق نماید. این ۱۰ الگوریتم شامل طبقه بندی، خوشه بندی، یادگیری آماری، تحلیل مشارکتی و کاوش ارتباطات می شود که همه در بین مهم ترین موضوعات در تحقیقات و پیشرفت های

¹. Web Intelligence

². Intelligent Agent Technology

³. Knowledge and Information Systems

مرتبط با داده کاوی می باشند که برای طراحی دوره های درسی مرتبط با داده کاوی، یادگیری ماشینی و هوش مصنوعی نیز بکار می روند.

www.Ketab.ir

تقدیر و تشکر

تعیین ۱۰ الگوریتم برتر داده‌کاوی در ماه می ۲۰۰۶ در بحثی بین دکتر جیانونگ کائو در دپارتمان حساب دانشگاه پلی تکنیک هنگ کنگ و دکتر زین دانگ وو شروع شد. زمانی که دکتر وو در حال ارائه سمیناری درباره‌ی ۱۰ مسئله‌ی نزاع برانگیز درباره‌ی تحقیقات داده‌کاوی در دانشگاه پلی تکنیک بود [۲]. دکتر وو و دکتر ویپین کومار این بحث را در KDD 06 در آگوست ۲۰۰۶ با افراد گوناگونی ادامه دادند و حمایت‌های بسیار جالبی در یافت کردند.

نایلا الیوت در دپارتمان علوم کامپیوتر و مهندسی در دانشگاه مینسوتا، گفته‌های مرتبط با الگوریتم را به همراه آراء آنها در فرآیند تعیین ۳ مرحله‌ای جمع‌آوری کردند. یانگ زنگ در دپارتمان علوم کامپیوتر در دانشگاه ورمانت تاییدیه‌های ۱۰ مرحله را که به شکل‌های متفاوتی بودند به فرمت یکسان LaTeX درآورد که این فرآیند منجر به جلوگیری از اتلاف زمان شد.

سینگ دانگ وا، ویپین کومار

۱۵ سپتامبر ۲۰۰۸

مراجع:

- [۱] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg, Top 10 algorithms in data mining, Knowledge and Information Systems, 14(2008), 1: 1-37.

- [۲] Qiang Yang and Xindong Wu (Contributors: Pedro Domingos, Charles Elkan, Johannes Gehrke, Jiawei Han, David Heckerman, Daniel Keim, Jiming Liu, David Madigan, Gregory Piatetsky-Shapiro, Vijay V. Raghavan, Rajeev Rastogi, Salvatore J. Stolfo, Alexander Tuzhilin, and Benjamin W. Wah), 10 challenging problems in data mining research, *International Journal of Information Technology & Decision Making*, 5, 4(2006), 597–604.

www.Ketab.ir

درباره مولفین

زین دانگ وو استاد دانشگاه و عضو هیئت علمی گروه علوم کامپیوتر دانشگاه ورمانت ایالات متحده است. او دکترای هوش مصنوعی از دانشگاه ادینبورگ بریتانیا دارد. مقالاتش شامل داده کاوی، سیستم‌های دانش محور و تحقیقات دانش وب می‌باشد. بیش از ۱۷۰ مقاله در این موضوعات در مقالات و کنفرانس‌های مختلف مشتمل بر IEEE TKDE, TRAMI, ACMTOIS, DMKD, KAIS, IJCAI, AAAI, ICML, KDD, ICDM, WWW ارائه کرده است. همچنین ۱۸ کتاب و مجموعه مقالات کنفرانس^۱ داشته است و برنده‌ی جایزه بهترین مقاله IEEE ICTAL 2005 و برنده مسابقات بهترین مقاله الگوریتم و نظریه IEEE ICDM 2007 می‌باشد.

دکتر وو سر ویراستار ژورنال دانش و مهندسی داده IEEE^۲ (TKDE)، انجمن کامپیوتر IEEE) موسس و رئیس کنونی کنفرانس بین‌المللی داده کاوی IEEE^۳، موسس و سر ویراستار افتخاری سیستم‌های اطلاعاتی و دانش (KAIS) توسط اسپرینگر^۴، رئیس (۲۰۰۲، ۲۰۰۶) کمیته فنی جامعه انفورماتیک هوشمند (TCII) انجمن کامپیوتر^۵ و ویراستار سری کتاب‌های پردازش پیشرفته اطلاعات و دانش (AI&KP)^۶ می‌باشد. او به عنوان رئیس کمیته برنامه‌ریزی ICDM'03 (کنفرانس بین‌المللی IEEE ۲۰۰۳ درباره‌ی داده کاوی) و کمک سرپرست کمیته برنامه‌ریزی

^۱. Conference proceedings

^۲. IEEE Transactions on Knowledge and Data Engineering

^۳. IEEE International Conference on Data Mining (ICDM)

^۴. Knowledge and Information Systems (KAIS, by Springer)

^۵. IEEE Computer Society Technical Committee on Intelligent Informatics (TCII)

^۶. Advanced Information and Knowledge Processing (AI&KP)

KDD-07 (سیزدهمین کنفرانس بین‌المللی ACM SIGKDD درباره‌ی کشف دانش و داده کاوی) برگزیده شده است. او برنده جوایز خدمات ACM SIGKDD سال ۲۰۰۴، خدمات فوق العاده IEEE ICDM در سال ۲۰۰۶، و استاد راهنما پروژه تحقیقاتی در Changjiang (و یا Yangtze River) سال ۲۰۰۵ در دانشگاه فنی هی فی^۱ برگزیده شده که توسط وزارت آموزش و پرورش چین و موسسه لی کا شینگ^۲ حمایت مالی شده بود. او به عنوان سخنران دعوت شده به کنفرانسهای بین المللی بسیاری مانند: NSF-NGDM07, PAKDD-07, IEEE EDOC06, IEEE ICTAI04, IEEE/WIC/ACM WI04/IAT04, SEKE 2002, PADD-97 می‌باشد.

ویشن کومار هم اکنون استاد دانشگاه ویلیام نوریز^۳ و سرپرست گروه مهندسی و علوم کامپیوتر و استاد دانشگاه مینسوتا^۴ می‌باشد. او کارشناسی مهندسی برق و ارتباطات را از دانشگاه فنی هند، رورکی (سابق بر این دانشگاه رورکی^۵) در سال ۱۹۷۷ دریافت کرد، کارشناسی ارشد مهندسی برق را از موسسه‌ی بین‌المللی فیلپس^۶ در ایند هاون هلند در سال ۱۹۷۹ گرفت و دکترای علوم کامپیوتر را از دانشگاه ماری لند^۷، کالج پارک، در سال ۱۹۸۲ دریافت کرد.

تحقیقات کنونی کومار شامل داده کاوی، بیو انفورماتیک و محاسبات با کارایی بالا^۸ می‌باشد. تحقیقات او باعث گسترش مفهوم معیار کارایی یکسان^۹ گردید که برای ارزیابی قابلیت افزایش مقیاس^{۱۰} الگوریتم‌های موازی همانند الگوریتم‌های موازی با

¹. Hefei

². Li Ka Shing

³. William Norris

⁴. University of Minnesota

⁵. University of Roorkee

⁶. Philips International Institute

⁷. University of Maryland

⁸. High-performance computing

⁹. Isoefficiency metric

¹⁰. Scalability

بازده بالا و نرم افزار فاکتورگیری ماتریس نامتراکم^۱ (PSPASES) و تقسیم بندی گرافی^۲ (METIS, ParMetis, hMetis) می باشد. او بیش از ۲۰۰ مقاله تحقیقاتی نوشته است و در نگارش ۹ کتاب مشارکت داشته که شامل دو کتاب بسیار پر کاربرد "مقدمه ای بر محاسبات موازی"^۳ و "مقدمه ای بر داده کاوی" می باشد که هر دو توسط انتشارات آدیسون ولسلی^۴ منتشر شده اند.

کومار به عنوان رئیس و نائب رئیس بسیاری از کنفرانس ها و اتاق کارها درباره داده کاوی و محاسبات موازی برگزیده شده که برخی از آن ها عبارتند از: کنفرانس بین المللی IEEE درباره داده کاوی^۵ (۲۰۰۲)، سیمپوزیوم بین المللی پردازش های موازی و توزیعی^۶ (۲۰۰۱) و کنفرانس بین المللی SIAM درباره داده کاوی^۷ (۲۰۰۱). کومار نائب رئیس کمیته برنامه ریزی کنفرانس بین المللی SIAM درباره داده کاوی و عضو کمیته برنامه ریزی کنفرانس بین المللی IEEE درباره داده کاوی و کنفرانس بین المللی IEEE درباره بيو انفورماتیک و بيو مديسين می باشد.

وی سرپرست ویراستاران مجله تحلیل آماری و استخراج داده^۸، سر ویراستار بولتن انفورماتیک IEEE^۹ و ویراستار سری کتاب های کشف دانش و داده کاوی^{۱۰} است که توسط انتشارات CRC Press/Chapman Hall به چاپ رسیده است. کومار عضو برد تخصصی ویراستاران داده کاوی و کشف دانش^{۱۱}، دانش و سیستم های اطلاعاتی^{۱۲}،

¹. Sparse matrix factorization

². Graph partitioning

³. Introduction to Parallel Computing

⁴. Addison-Wesley

⁵. IEEE International Conference on Data Mining

⁶. International Parallel and Distributed Processing Symposium

⁷. SIAM International Conference on Data Mining

⁸. Journal of Statistical Analysis and Data Mining

⁹. IEEE Intelligent Informatics Bulletin

¹⁰. Data Mining and Knowledge Discovery Book Series

¹¹. Data Mining and Knowledge Discovery

¹². Knowledge and Information Systems

بولتین هوش محاسباتی IEEE^۱، بررسی سالانه‌ی انفورماتیک هوشمند^۲، محاسبات موازی^۳، مجله محاسبات توزیعی و موازی^۴، مجله مهندسی داده و دانش IEEE^۵ (۱۹۹۷-۱۹۹۳) هم زمانی IEEE^۶ (۱۹۹۷-۲۰۰۰) و تکنولوژی توزیعی و موازی IEEE^۷ (۱۹۹۷-۱۹۹۵) می‌باشد. او عضو IEEE, AAAS, SIAM و ACM می‌باشد. کومار به جهت مشارکت در طراحی و تحلیل الگوریتم‌های موازی^۸، تقسیم‌بندی گرافیکی^۹ و داده‌کاوی^{۱۰}، جایزه دستیافت فنی علوم کامپیوتر IEEE سال ۲۰۰۵ را دریافت نموده است.

¹. IEEE Computational Intelligence Bulletin

². Annual Review of Intelligent Informatics

³. Parallel Computing

⁴. Journal of Parallel and Distributed Computing

⁵. IEEE Transactions of Data and Knowledge Engineering

⁶. IEEE Concurrency

⁷. IEEE Parallel and Distributed Technology

⁸. Design and analysis of parallel algorithms

⁹. Graph-partitioning

¹⁰. Data mining

همکاران

Songcan Chen, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Joydeep Ghosh, University of Texas at Austin, Austin, TX

David J. Hand, Imperial College, London, UK

Alexander Liu, University of Texas at Austin, Austin, TX

Bing Liu, University of Illinois at Chicago, Chicago, IL

Geoffrey J. McLachlan, University of Queensland, Brisbane, Australia

Hiroshi Motoda, ISIR, Osaka University and AFOSR/AOARD, Air Force Research

Laboratory, Japan

Shu-Kay Ng, Griffith University, Meadowbrook, Australia

Kouzou Ohara, ISIR, Osaka University, Japan

Naren Ramakrishnan, Virginia Tech, Blacksburg, VA

Michael Steinbach, University of Minnesota, Minneapolis, MN

Dan Steinberg, Salford Systems, San Diego, CA

Pang-Ning Tan, Michigan State University, East Lansing, MI

Hui Xue, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Qiang Yang, Hong Kong University of Science and Technology, Clearwater Bay,

Kowloon, Hong Kong

Philip S. Yu, University of Illinois at Chicago, Chicago, IL

Yang Yu, Nanjing University, Nanjing, China

Zhi-Hua Zhou, Nanjing University, Nanjing, China