

بررسی روش های متن کاوی

مصطفی بریندزاده و سعید چراغی فر

عضو هیئت علمی دانشگاه پیام نور



انتشارات مینوفر
Minufar.ir

مشهد مقدس - ۱۳۹۵



- | | |
|---------------------|---|
| سرشناسه | : بیرومندزاده،، مصطفی، ۱۳۵۷ - |
| عنوان و نام پدیدآور | : بررسی روش‌های متن کاوی/مصطفی بیرومندزاده و سعید چراغی فر. |
| مشخصات نشر | : مشهد: مینوفر، ۱۳۹۵. |
| مشخصات ظاهری | : ۹۱ ص.: مصور (بخشی رنگی)، جدول، نمودار. |
| فروست | : فناوری؛ ۸۱۰۳. |
| شابک | : ۱۰۰۰۰۰ ریال: ۶-۷۹-۸۰۶۰-۶۰۰-۹۷۸ |
| وضعیت فهرست نویسی | : فیبا |
| یادداشت | : کتابنامه: ص. [۸۷]-۹۱. |
| موضوع | : داده کاوی |
| موضوع | : Data mining |
| شناسه افزوده | : چراغی فر، سعید، ۱۳۵۹ - |
| رده بندی کنگره | : ۹/۷۶۴QA ۱۳۹۵۲د ۴ |
| رده بندی دیویی | : ۰۰۶/۳۱۳ |
| شماره کتابشناسی ملی | : ۴۳۹۳۳۰۶ |



عنوان کتاب: بررسی روش‌های متن کاوی

نویسندگان: مصطفی برومندزاده - سعید چراغی‌فر

طراح جلد: مینوفر

صفحه‌آرا: اکرم ملک‌نژاد

شمارگان: ۱۰۰۰ نسخه

نوبت چاپ: اول - ۱۳۹۵

چاپ و صحافی: دقت

قطع: وزیری

شابک: ۹۷۸-۶۰۰-۸۰۶۰-۷۹-۶

قیمت: ۱۰,۰۰۰ تومان

تمام حقوق چاپ و نشر این اثر محفوظ است.

انتشارات مینوفر: مشهد مقدس، بلوار جلال‌آل احمد ۶۹ پ ۱۱۶، ص.پ: ۹۱۸۹۵-۱۷۵۵

همراه: ۰۹۳۵۲۱۶۲۷۵۵ - ۰۹۱۲۰۶۸۲۷۵۵ تلفن و دورنگار: ۰۵۱-۳۸۳۲۳۵۵۳

سایت: www.minufar.ir ایمیل: minufar@yahoo.com

فهرست مطالب

۱۱	پیشگفتار
۱۳	فصل اول: بررسی روش های انتخاب ویژگی
۱۳	مقدمه
۱۵	روش های مبتنی بر استخراج ویژگی
۱۶	روش های انتخاب ویژگی
۱۶	تعاریف
۲۰	بررسی توابع مختلف ریاضی و تولیدکننده
۲۰	توابع تولیدکننده
۲۱	جستجوی کامل
۲۱	جستجوی مکاشفه ای
۲۱	جستجوی تصادفی
۲۲	توابع ارزیابی
۲۵	دسته بندی و تشریح الگوریتم های مختلف انتخاب ویژگی
۲۶	تابع ارزیابی مبتنی بر فاصله - تابع تولیدکننده مکاشفه ای
۲۶	روش Relief
۲۸	روش Jakub
۲۸	تابع ارزیابی مبتنی بر فاصله - تابع تولیدکننده کامل
۳۰	تابع ارزیابی مبتنی بر اطلاعات - تابع تولیدکننده مکاشفه ای
۳۱	(۱) روش درخت تصمیم (DTM)
۳۱	الگوریتم C4.5
۳۴	(۲) روش استفاده شده توسط Koller و Sahami
۳۴	پوشش مارکوف
۳۷	تابع ارزیابی مبتنی بر وابستگی - تابع تولیدکننده مکاشفه ای
۳۷	Probability of Error & Average Correlation Coefficient (POE1ACC) (۱)
۳۸	روش PreSet (۲)

۳۹	 تابع ارزیابی مبتنی بر سازگاری - تابع تولیدکننده کامل
۳۹	 Focus روش (۱)
۴۵	 Schlimmer روش (۲)
۴۶	 MIFES1 روش (۳)
۴۶	 تابع ارزیابی مبتنی بر سازگاری - تابع تولیدکننده تصادفی
۴۸	 تابع ارزیابی مبتنی بر خطای طبقه‌بندی کننده - تابع تولیدکننده مکاشفه‌ای
۴۸	 SFS (Sequential Forward Selection) روش (۱)
۴۸	 SBS (Sequential Backward Selection) روش (۲)
۴۸	 SBS-Slash روش (۳)
۴۹	 PQSS ((p,q) Sequential Search) روش (۴)
۴۹	 BDS (Bi-Directional Search) روش (۵)
۴۹	 Schematic Search روش (۶)
۵۰	 RC (Relevance in Context) روش (۷)
۵۰	 Quinlan's and Gelsema روش (۸)
۵۰	 تابع ارزیابی مبتنی بر خطای طبقه‌بندی کننده - تابع تولیدکننده کامل
۵۰	 Michi84 Linear Classifier روش (۱)
۵۰	 Skla84 Box Classifier روش (۲)
۵۱	 AMB&B (Approximate monotonic branch and bound) روش (۳)
۵۱	 BS (Beam Search) روش (۴)
۵۱	 تابع ارزیابی مبتنی بر خطای طبقه‌بندی کننده - تابع تولیدکننده تصادفی
۵۱	 LVW روش (۱)
۵۲	 GA (Genetic Algorithm) روش الگوریتم ژنتیک (۲)
۵۲	 SA (Simulated Annealing) روش (۳)
۵۲	 GSS (Random Generation plus Sequential Selection) روش (۴)
۵۲	 RMHC-PF1 روش (۵)
۵۳	 جمع‌بندی روش‌های انتخاب ویژگی

فصل دوم: بررسی روش‌ها و الگوریتم‌های فرا اکتشافی

۵۵	 روش‌های فرا اکتشافی
۵۶	 روش‌های مکاشفه‌ای
۵۷	 انواع الگوریتم‌های مکاشفه‌ای
۵۹	 پیاده‌سازی الگوریتم‌های فرا اکتشافی

۶۰	ویژگی‌های مشترک روش‌های فرااکتشافی
۶۰	دسته‌بندی الگوریتم‌های فرااکتشافی
۶۱	الگوریتم ژنتیک (Genetic Algorithm)
۶۵	مراحل الگوریتم ژنتیک
۶۵	انواع کدینگ
۶۶	روشهای کدینگ
۶۷	روش‌های پیاده‌سازی عملگر ترکیب
۷۰	انواع روش‌های جهش
۷۱	الگوریتم ژنتیک برای انتخاب ویژگی
۷۲	الگوریتم بهینه‌سازی جمعیت مورچگان (ACO)
۷۶	الگوریتم ACO برای انتخاب ویژگی
۷۸	الگوریتم بهینه‌سازی انبوه ذرات (PSO)
۷۹	الگوریتم PSO برای انتخاب ویژگی
۸۳	الگوریتم جستجوی ممنوعه (TABU SEARCH)
۸۵	استراتژی‌های پیشرفته جستجوی ممنوعه
۸۶	حافظه‌ها در جستجوی ممنوعه
۸۷	مفهوم همسایگی
۸۸	الگوریتم جستجوی ممنوعه برای انتخاب ویژگی
۹۱	فهرست منابع و مراجع

پیشگفتار

امروزه با گسترش سیستم‌های اطلاعاتی و حجم بالای داده‌های ذخیره‌شده در این سیستم‌ها، نیاز به ابزاری است تا بتوان داده‌های ذخیره‌شده را پردازش نمود و اطلاعات حاصل از این پردازش را در اختیار کاربران قرارداد. با استفاده از پرسش‌های ساده در SQL و ابزارهای نوگون گزارش‌گیری، می‌توان اطلاعاتی را در اختیار کاربران قرارداد تا بتوانند به نتیجه‌گیری در مورد داده‌ها و روابط منطقی میان آن‌ها بپردازند اما وقتی که حجم داده‌ها بالا باشد، کاربران هرچند برده‌ها و با تجربه باشند نمی‌توانند الگوهای مفید را در میان حجم انبوه داده‌ها تشخیص دهند و یا اگر قادر به این کار هم باشند، هزینه عملیات از نظر نیروی انسانی و مادی بسیار بالا است. از سویی دیگر کاربران معمولاً فرضیه‌ای را مطرح می‌کنند و سپس براساس گزارشات مشاهده‌شده به دنبال اثبات یا رد فرضیه می‌پردازند، درحالی‌که امروزه نیاز به روش‌های دقیق و منطقی‌تری است که به‌واسطه آن‌ها اصطلاحاً به کشف دانش بپردازند و با کمترین دخالت کاربر و به صورتی خودکار. تمام این روابط منطقی میان آن‌ها را بیان نمایند. داده‌کاوی و یکی از زیرشاخه‌های آن یعنی کاشی‌کاوی، از مهم‌ترین این روش‌ها هستند که به‌وسیله آن‌ها الگوهای مفید در داده‌ها با حداقل دخالت کاربران شناخته‌شده و اطلاعاتی را در اختیار کاربران و تحلیلگران قرار می‌دهند تا براساس آن‌ها تصمیمات مهم و حیاتی در سازمان‌ها اتخاذ شود. در این علوم از تحلیل اکتشافی داده‌ها که بخشی از علم آمار است، استفاده می‌شود که در آن بر کشف اطلاعات نهفته و ناشناخته از درون حجم انبوه داده‌ها تأکید می‌شود. علاوه بر این با آمیزش تئوری‌های پایگاه داده‌ها، هوش مصنوعی، یادگیری ماشین و علم آمار زمینه کاربردی‌تری در پردازش داده‌های بزرگ فراهم می‌گردد.