

معرفی شبکه عصبی مدل مخچه به عنوان روشی برای کاوش اطلاعات محتوایی صفحات وب

حسن طالبی

چکیده - یکی از روشهای اصلی در جستجوی محتوای وب، گروه بندی صفحات وب است، که اغلب با اطلاعات متنی، اطلاعات ساختاری و پیوندهای وب سروکار دارد و نقطه ثقل تحقیقات روی دسته بندی اتوماتیک اطلاعات متنی وب یعنی دسته بندی بر اساس محتوای متن است. در این مقاله سعی شده است مدل جدیدی از شبکه های عصبی پردازشگر محاسباتی مدل مخچه برای استفاده در دسته بندی صفحات وب پیشنهاد گردد، که نتایج نشانگر این بود که دسته بندی صفحات وب با این مدل کارایی بهتری نسبت به سایر الگوریتمها دارد.

کلید واژه- دسته بندی صفحات وب، جستجوی محتوا، شبکه عصبی، شبکه گسترده جهانی، مدل مخچه.

حقیقت وب، محیط پرباری برای تحقیقات داده کاوی روی حجم

مقدمه

عظیمی از اطلاعات آنلاین است.

داده کاوی فرایند شناسایی الگوهای معتبر، تازه، بالقوه، مفید و قابل درک موجود در داده است، وب کاوی نیز به کشف و تحلیل اطلاعات سودمند وب گفته می شود. در وب کاوی، داده می تواند در سمت سرویس دهنده، سمت مشتری و یا سرویس دهنده پروکسی باشد و یا از اطلاعات سازمان یافته در یک پایگاه داده بدست آید. همچنین بسته به محل منبع، نوع داده، می تواند متفاوت باشد که باعث می شود تنوع زیادی، هم در محتوای آنها (متن، تصویر، صوت، سمبل ها) و هم در فرا اطلاعات وجود داشته باشد، از این رو تکنیک های مورد استفاده در وب کاوی بسیار گسترده هستند.

وب کاوی بکارگیری تکنیک های داده کاوی به منظور کشف و استخراج خودکار اطلاعات از مستندات و محتویات و سرویسهای وب است [۱]. عمل دسته بندی، یکی از تکنیکهای اصلی در کاوش محتوای وب است. دسته بندی اتوماتیک صفحات وب با اطلاعات متنی، اطلاعات ساختاری و اطلاعات لینکهای صفحات وب سر و کار دارد که تمرکز تحقیقات روی دسته بندی اتوماتیک اطلاعات متنی وب، یعنی دسته بندی بر اساس محتوای متن است.

در نتیجه توصیف و سازماندهی این مقدار عظیم محتویات روی وب، برای فهم آن به عنوان یک منبع اطلاعات، ضروری است. در