



بخشی از ترجمه مقاله

عنوان فارسی مقاله :

ساخت سلسله مراتب و دسته بندی متن بر مبنای استراتژی
آزاد سازی و مدل حداقل اطلاعات

عنوان انگلیسی مقاله :

Hierarchy construction and text classification based on
the relaxation strategy and least information model



توجه !

این فایل تنها قسمتی از ترجمه میباشد. برای تهیه مقاله ترجمه شده کامل
با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، [اینجا](#) کلیک نمایید.



بخشی از ترجمه مقاله

6. Conclusions

We propose the hierarchical classification approach based on the relaxation strategy which alleviates the impact of the 'blocking' problem. It delays the uncertain category decision until it can be classified definitely, and so the error that has occurred in the upper level will not be transferred to the lower level. We also apply the Least Information Theory in term weighting and documentation representation and it offers a new basic information quantify model by different probability distributions.

The experiments on RCV1 data shows that the text hierarchical classification method based on the relaxation strategy has greater advantage on the time performance over SVM with an increasing data size. And it can also maintain both higher precision and recall simultaneously. Specially, by the use of LIT method for term weighting, the classification performance achieves significant improvement over classic TF*IDF on most classifiers. LIT measures the distance between two probability distributions and establishes a new basic information quantity approach. It performed better than TF*IDF not only in the field of text classification but also other natural language processing application, such as clustering, information retrieval and so on.

6. نتایج

ما رویکرد دسته بندی سلسله مراتب را بر مبنای استراتژی آزاد سازی ارائه می کنیم که تأثیر مسئله «انسداد» را کاهش می دهد. این رویکرد تصمیم دسته نامعین را تا زمانی که به صورت قطعی دسته بندی شود به تأخیر می اندازد و بنابراین خطایی که در سطح بالاتر روی داده است به سطح پایین تر انتقال داده نمی شود. همچنین نظریه حداقل اطلاعات را در وزن دهی واره و فایز سند سازی نیز اعمال کردیم که توسط توزیع های احتمال مختلف مدل تعیین کیفیت اطلاعات اساسی جدیدی را ارائه می کند. آزمایش های روی داده های RCV1 نشان می دهد که روش دسته بندی سلسله مراتبی متن مبتنی بر استراتژی آزاد سازی با افزایش اندازه داده در مقایسه با SVM کارایی بالاتری بر حسب عملکرد زمانی دارد. همچنین می تواند همزمان دقت بالاتر و فراخوانی را حفظ کند. به ویژه با استفاده از روش LIT برای وزن دهی واره عملکرد دسته بندی در بیشتر رده بندها در مقایسه با TF*IDF بهبود قابل توجهی می یابد. LIT فاصله بین دو توزیع احتمال را اندازه می گیرد و رویکرد کمیت اطلاعات اساسی جدیدی را وضع می کند. این رویکرد نه تنها در زمینه دسته بندی متن بلکه در کاربردهای پردازش زبان طبیعی، مانند خوشه بندی، بازیابی اطلاعات و غیره، نیز عملکرد بهتری در مقایسه با TF*IDF دارد.



توجه!

این فایل تنها قسمتی از ترجمه میباشد. برای تهیه مقاله ترجمه شده کامل با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، [اینجا](#) کلیک نمایید.

برای جستجوی جدیدترین مقالات ترجمه شده، [اینجا](#) کلیک نمایید.