جامعة الفيوم
كلية الحاسبات والمعلومات
قسم نظم المعلومات

البحث رقم (٦)

A Proposed Method for Minimizing Mining Tasks' Data Dimensionality

اسم البحث	**طريقة مقترحة لتقليل أبعاد بيانات مهام التعدين**
ملخص المشكلة	ساهمت تقنيات اكتشاف المعرفة بشكل كبير في العديد من المجالات بنجاح كبير. ومع ذلك، مع النمو المستمر للبيانات ، تعاني هذه التقنيات من اختناقات في معالجة هذه البيانات
سياق البحث	من الاتجاهات التي تقلل من تأثير حجم البيانات هو تقليل أبعاد البيانات والتي تركز على التخلص من السمات التي ليس لها تأثير كبير على دقة تقنية الاكتشاف. يقترح هذا البحث طريقة جديدة لتقليل أبعاد البيانات. تعتمد الطريقة المقترحة على ركيزتين رئيسيتين، الأول هو تطبيق طريقة Saaty المعدلة لتحديد اتساق السمات مع اقتراح مزيد من التكيف يستهدف تحديد دقة أكثر دقة. الركيزة الثانية هي تطبيق تقنيات التجميع على السمات المتسقة لإزالة السمات الأقل ترجيحًا في كل مجموعة والتي لديها أيضًا أقل المقاييس اتساقًا. نتيجة تطبيق الخطوتين هي إبراز أهم سمات مجموعة البيانات.
أسلوب البحث	يمكن تلخيص أهمية البحث الرئيسية في النقاط التالية: ١. تتكيف الطريقة المقترحة مع تطبيق تقنيات ترجيح مختلفة من خلال اقتراح نهج تعاوني لهذه التقنيات في تحديد أهمية السمات. ٢. يتم إجراء تقييم تقنيات الترجيح من خلال تطبيق تعديل مقترح على طريقة Saaty لاكتشاف اتساق تقنية الترجيح. ٣. يتم بعد ذلك الكشف عن اتساق السمات باستخدام نفس الأسلوب لإزالة السمات غير المتسقة كمرحلة الحذف الأولى. ٤. بالنظر إلى تقنيات الترجيح، يتم إجراء الترجيح الهام للسمات المتسقة فيما يتعلق بالنسب المتسقة لتقنية الترجيح. ٥. نظرت مرحلة الحذف الثانية في تطبيق تقنيات التجميع لتحديد العلاقات بين السمات واكتشاف أعلى السمات المرجحة في كل مجموعة على أنها سمات مهمة. ٦. يتم تمثيل مخرجات الطريقة المطلوبة على أنها المجموعة النهائية من السمات التي تتضمن السمات المكتشفة من مرحلة التجميع. تم تطبيق الطريقة المقترحة بنجاح على مجموعة بيانات أمراض الجهاز الهضمي والتي تم تقليل السمات من ٦٢ سمة إلى ٣١ سمة. تم تطبيق مجموعة من تقنيات التصنيف على مجموعة البيانات لإثبات أن تقليل الأبعاد قد احتفظ بدقة مهمة التصنيف، وقد أظهرت النتائج أن خوارزمية Random Forest تتمتع بدقة تصل إلى ٩٥,٣٦٪ عند تطبيقها على مجموعة البيانات المعتمدة.
النتائج المستخلصة	تؤدي القدرة على استخدام تقنيات الترجيح إلى تحديد اتساق السمات الذي يسلط الضوء على أن التقنيات الإحصائية وتقنيات التعلم الآلي يمكن اعتبارها عملة ذات وجهين. نجح نهج البحث المقترح في استخدام هذه التقنيات في بيئة تعاونية للوصول إلى هدف تقليل الأبعاد المطلوب.

جامعة الفيوم
كلية الحاسبات والمعلومات
قسم نظم المعلومات