تعزيز تنقيب البيانات المتدفقة في شبكات الاستشعار اللاسلكية باستخدام خوارزميات التجميع

ياسمين سند الغامدي

المستخلص

شهدت السنوات القليلة الماضية اهتماماً متزايداً في استخدام شبكات الاستشعار اللاسلكية المختصرة بـ (WSNs) في مجموعة واسعة من التطبيقات ، حيث أصبح هذا النوع من الشبكات مجالاً ساخناً للبحث ، وبسبب التقدم والنمو في مجال تكنولوجيا الاتصالات اللاسلكية ، فقد أصحبت شبكات الاستشعار اللاسلكية مرغوبة على نحو متزايد للعديد من التطبيقات في مجالات مختلفة ، كمجال المراقبة العسكرية ، والكشف عن الحرائق ، ومراقبة المساكن الطبيعية ، وفي الصناعة ، والمراقبة الصحية وغيرها الكثير .

تتكون شبكات الاستشعار اللاسلكية من عقد استشعار فردية صغيرة قادرة على التفاعل مع بيئاتها بواسطة ميزة الاستشعار عن بعد والسيطرة على البارامترات المادية ، لكن هذه العقد الصغيرة تعاني من قصور مصادر الطاقة الكائنة بداخلها مما قد يؤدي إلى تقصير عمر شبكة الاستشعار اللاسلكية بشكل عام .

تميل هذه الشبكات إلى توليد كمية كبيرة من البيانات الصغيرة المتتابعة والنابعة من عدة عقد استشعار تسمى بالبيانات المتدفقة ، تتدفق البيانات المتدفقة بسرعة عالية جداً عبر الإنترنت ، وتتميز الحزمة منها بأنها ضخمة وغير محدودة وقد يتم فقد السيطرة على معالجتها حسب ترتيب وصولها .

وبسبب القيود السابقة الذكر لشبكات الاستشعار اللاسلكية ، بالإضافة إلى سرعة وصول البيانات المتدفقة وضخامة حجمها ، توجد هناك حاجة ماسة لحل التحديات التي قد تعيق عمل شبكات الاستشعار اللاسلكية ، وتكمن هذه التحديات في محاولة إطالة عمر شبكة الاستشعار اللاسلكية ، كما تكمن في محاولة خفض استهلاك طاقة عقد الاستشعار ، وأخيراً تكمن في الحد من التأخير الذي يحدث نتيجة لبعض العمليات ولحركة البيانات داخل هذا النوع من الشبكات .

تقنية التنقيب عن البيانات هي تقنية يمكن أن تستخدم لمواجهة ومعالجة التحديات المذكورة بشبكات الاستشعار اللاسلكية ، وتشمل هذه التقنية عدة أنواع منها التجميع ، والتصنيف ، والعد المتكرر، وتحليل السلاسل الزمنية وغيرها .

كما أثبتت الأبحاث أن خوارزميات التجميع تلعب دوراً هاماً في تنظيم شبكات الاستشعار اللاسلكية وحل التحديات المذكورة سابقاً . بناء عليه سيقوم البحث بتطوير ومحاكاة خوارزمية باستخدام إحدى تقنيات التجميع وتسمى هذه الخوارزمية بخوارزمية التجميع الكثافي الشبكي المختصرة بـ (DeGiCA) . هذه الخوارزمية تعزز من آداء تقنية التجميع في شبكات الاستشعار اللاسلكية وذلك من خلال الجمع بين ثلاث تقنيات هي : تقنية التجميع وتقنية الكثافة وتقنية الشبكة .

تتميز تقنية الكثافة باستطاعتها استخلاص الأشكال التقديرية للتجميع ، أما تقنية الشبكة فبإمكانها إزالة العقد من جوانب الشبكة والتي وجودها قد يقلل من جودة التجميع ، وبذلك تكون خوارزمية الـ (DeGiCA) قادرة على مواجهة التحديات أثناء نقلها للبيانات المتدفقة .

وباستخدام برنامج النمذجة والمحاكاة المعروف بالـ (MATLAB) ، وبالمقارنة مع خوارزميات أخرى تعالج نفس التحديات المذكورة أعلاه بشبكات الاستشعار اللاسلكية ، تمت مقارنة نتائج خوارزمية (DeGiCA) بنتائج خوارزميتين أخريتين تعرفان باسم (K-means) و(FCM) ، كلتا الخوارزميتين تستخدم تقنية التجميع في شبكات الاستشعار اللاسلكية والتي تنقل البيانات المتدفقة .

على وجه الخصوص ، تتفوق خوارزمية الـ (DeGiCA) على خوارزمية الـ (K-Means) من حيث إطالة عمر الشبكة بنسبة ١٥% ، ونسبة ١٣% من حيث الحفاظ على الطاقة ، ونسبة ٤٠% من حيث وصول حزم البيانات المتدفقة. من جهة أخرى فإن خوارزمية الـ (DeGiCA) تتفوق على خوارزمية الـ (FCM) من حيث إطالة عمر الشبكة بنسبة ١٧% ، ونسبة ١١% من حيث الحفاظ على الطاقة ، ونسبة ٧٠% من حيث وصول حزم البيانات المتدفقة.

ENHANCING DATA STREAM MINING IN WIRELESS SENSOR NETWORIKS USING CLUSTERING ALGORITHMS

Yassmeen Sanad Alghamdi

ABSTRACT

The past few years have witnessed an increased interest in the potential use of Wireless Sensor Networks (WSNs) in a wide range of applications in the field of military surveillance, fire detection, habitat monitoring, industry, health monitoring and many more. WSNs consist of individual nodes that are able to interact with their environment by sensing and controlling physical parameters. Sensor nodes tend to generate a large amount of sequential small and tuple-oriented data that is considered as Data Streams. Data streams usually are huge data sets that arrive in an online fashion, flowing rapidly in a very high speed, where they are unlimited and there is no control on the arrival processing order. Due to sensor network limitations, some

challenges are faced and urgently need to be solved. Such challenges include long lasting the WSN lifetime and reducing nodes energy consumption. Data mining could deal with the WSN challenges. Clustering is one of mining techniques and plays an important role in organizing WSNs. It has proven its efficiency on network performance by extending network lifetime, saving energy of sensor nodes, reducing delay and delivering more data packets. This research develops an algorithm called the Density Grid-base Clustering algorithm (DeGiCA) that enhances the clustering mining technique in WSNs by combining density and grid techniques. The deployment density variation technique can find arbitrary shaped clusters while the grid technique is used to avoid clustering quality problems by discarding the boundary points of grids. DeGiCA helps to face the limitations found in WSNs that carry data streams. By using a MATLAB-based simulator, DeGiCA is compared with other clustering algorithms in WSNs that manipulate data streams. Fuzzy Clustering Means algorithm (FCM) and K-means algorithm are two selected algorithms used to be compared with DeGiCA performance metrics results. The simulation results indicate that the performance of DeGiCA outperforms K-Means in terms of network lifetime by 15%, energy consumption by 13% and packet delivery ratio by 40%. DeGiCA also outperforms FCM in terms of network lifetime by 17%, energy consumption by 11% and packet delivery ratio by 70%.