ทิพยา ถินสูงเนิน : การแทนอนุกรมเวลาเพื่อการจัดกลุ่มที่มีประสิทธิภาพ (TIME SERIES REPRESENTATION FOR EFFICIENT CLUSTERING) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.นิตยา เกิดประสพ, 132 หน้า.

ข้อมูลคลื่นไฟฟ้าหัวใจและคลื่นไฟฟ้าสมองที่ได้จากการตรวจจับกระแสไฟฟ้าที่ออกมาจากหัวใจและสมองมีความรู้ซ่อนอยู่มากมายสามารถนำมาใช้เพื่อช่วยในการวินิจฉัยโรคหรือเตือนภัยก่อนโรคร้ายจะมาถึงได้ ดังนั้นหากสามารถจำแนกกลุ่มข้อมูลคลื่นไฟฟ้าหัวใจและคลื่นไฟฟ้าสมองเหล่านี้ได้ชัดเจนและแม่นยำขึ้นจะเป็นประโยชน์อย่างมากในทางการแพทย์ ข้อมูลคลื่นไฟฟ้าเหล่านี้จัดเป็นข้อมูลอนุกรมเวลาที่มักมีขนาดใหญ่ มิติสูงและซับซ้อน ซึ่งเป็นความท้าทายสำหรับการเรียนรู้ของเครื่องจักร หัวใจของความสำเร็จอย่างหนึ่งสำหรับการเรียนรู้ข้อมูลประเภทนี้คือการแทนข้อมูลที่ดี ดังนั้นงานวิจัยนี้จึงได้นำเสนอวิธีการแทนข้อมูลอนุกรมเวลาเพื่อให้เกิดประสิทธิภาพสูงขึ้นในการจัดกลุ่มสำหรับข้อมูลคลื่นไฟฟ้าหัวใจและคลื่นไฟฟ้าสมอง โดยอาศัยโครงข่ายการเข้ารหัสอัตโนมัติเชิงลึก (ดีเอเอ็น) ซึ่งเป็นวิธีที่นำเอาหลักการของ โบลทซ์มันน์แมชชีนเชิงจำกัดมาใช้ร่วมกับเทคนิคที่มีประสิทธิภาพสูงในด้านการแทนข้อมูลอย่างออโตเอ็นโคดเดอร์ งานวิจัยนี้ใช้อัลกอริทึมเชิงพันธุกรรมเพื่อหาโครงข่ายที่ดีที่สุดให้กับ ดีเอเอ็น เรียกว่า “ดีเอเอ็นจีเอ” งานวิจัยนี้จัดกลุ่มอนุกรมเวลาด้วยอัลกอริทึมพีดีซีและเค-มีนส์ โดยเปรียบเทียบประสิทธิภาพกับข้อมูลดั้งเดิมและเทคนิคการแทนข้อมูลด้วยวิธีอื่น ด้วยมาตรวัดค่าความถูกต้อง ค่าพิวริตี เวลาในการประมวลผลและประเมินค่าความเหมาะสมของกลุ่มด้วย ค่าซิลลูเอ็ต และค่าผลรวมความผิดพลาด ผลการวิจัยพบว่า ดีเอเอ็นจีเอสามารถผลิตตัวแทนอนุกรมเวลาที่ช่วยเพิ่มประสิทธิภาพการจัดกลุ่มให้ดีขึ้นได้ทั้งในอัลกอริทึมพีดีซีและเค-มีนส์ เป็นตัวแทนอนุกรมเวลามีประสิทธิภาพดีกว่าข้อมูลดั้งเดิม ดีกว่าตัวแทนข้อมูลจากเทคนิคอื่น แต่มีข้อจำกัดคือใช้เวลาในการประมวลผลสูงกว่าวิธีการอื่น เมื่อพิจารณาข้อมูลคลื่นไฟฟ้าหัวใจ ผลการวิจัยพบว่า ตัวแทนอนุกรมเวลาที่ได้ให้ประสิทธิภาพการจัดกลุ่มเพิ่มขึ้นทั้งอัลกอริทึมพีดีซีและเค-มีนส์ โดยเฉพาะในอัลกอริทึมพีดีซีให้ค่าความถูกต้อง และค่าพิวริตี เพิ่มขึ้นถึง 30% และ 23% ตามลำดับ นอกจากนี้ซิลลูเอ็ตและผลรวมความผิดพลาดยังบ่งชี้ให้เห็นว่าการจัดกลุ่มมีความเหมาะสมตามธรรมชาติของข้อมูลและกลุ่มที่แท้จริงอีกด้วย สำหรับข้อมูลคลื่นไฟฟ้าสมอง พบว่า ตัวแทนข้อมูลที่ได้ เหมาะกับการจัดกลุ่มด้วยอัลกอริทึมพีดีซีโดยสามารถเพิ่มประสิทธิภาพได้สูงมากคือให้ค่าความถูกต้อง และค่าพิวริตี เพิ่มขึ้นถึง 31% และ 61% ตามลำดับ

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนักศึกษา _____

ปีการศึกษา 2560 ลายมือชื่ออาจารย์ที่ปรึกษา _____

TIPPAYA THINSUNGNOEN : TIME SERIES REPRESENTATION FOR EFFICIENT CLUSTERING. THESIS ADVISOR : ASSOC. PROF. NITTAYA KERDPRASOP, Ph.D., 132 PP.

TIME SERIES ANALYSIS / DEEP LEARNING / RESTRICTED BOLTZMANN MACHINES / DEEP AUTOENCODER.

Electrocardiogram signals (ECGs) and electroencephalographic signals (EEGs) are time series detected from electrical flow of the heart and brain. Deep analysis of these data can reveal some hidden knowledge potentially useful for the accurate diagnosis or warning an early alarm for heart disease. Electrical signals are normally organized into time series data that are usually large, high dimensional, and complex in their components. Therefore, efficient analysis with the machine learning techniques is a challenging problem. One of the key successes for this kind of learning is to learn from the representative data that are carefully selected. In this research, we present a method for efficient casting of time series representatives that are to be used later for time series clustering for ECGs and EEGs. To find series representative, we propose to use Deep Autoencoder Networks (DANs), which is a technique based on Restricted Boltzmann Machines and Autoencoder. This research determines the appropriate network for DANs by using genetic algorithm called "DANGA". The signal representatives are then clustered using the PDC and k-Means algorithms. The clustering results obtained from our proposed method are compared against other time series representation techniques based on the cluster evaluation (accuracy), purity, processing time, silhouette, and the number of clusters considered from the sum of

square error (SSE). The experimental results show that our proposed method can cast for more appropriate time series representatives than others techniques with the longer processing time trade-off. The ECGs representatives yield the better performance on time series clustering with the 30% improvement in grouping accuracy and 23% increase in the purity metric. Furthermore, silhouette and SSE index indicate natural clusters. For EEGs clustering, the results yield the better performance with the 31% improvement in accuracy and 61% increase in purity.

มหาวิทยาลัยเทคโนโลยีสุรนารี

School of Computer Engineering Student's Signature

Academic Year 2017 Advisor's Signature