

การ Clustering High-dimensional data เป็นปัญหาสำคัญใน Machine Learning โดยวิธี Subspace Clustering เป็นหนึ่งในแนวทางที่ได้รับความนิยม อย่างไรก็ตาม วิธีนี้มักไวต่อ Noise และ Outliers ซึ่งอาจทำให้ Affinity Matrix ผิดพลาด งานวิจัยนี้จึงนำเสนอ Game Model for Semi-supervised Subspace Clustering (GMSSC) ซึ่งใช้ Semi-supervised Learning ร่วมกับ Dynamic Affinity Learning โดยอาศัย Graph Convolutional Network (GCN) และ Self-expressive Learning เพื่อเพิ่มความแม่นยำและเสถียรภาพของการจัดกลุ่ม พร้อมทั้งประยุกต์ใช้ Game Theory เพื่อสร้างสมดุลระหว่างการเรียนรู้ของสองเมทริกซ์

แนวคิดหลักของ GMSSC คือ 1. **Data Self-expressiveness**: แนวคิดที่ว่าแต่ละตัวอย่างสามารถแทนตัวเองด้วยการรวมเชิงเส้นของตัวอย่างอื่นใน Subspace เดียวกัน วิธีแบบเดิมสร้าง Affinity Matrix จากข้อมูลดิบโดยตรง ซึ่งไวต่อ Noise และ Outliers ส่วนแนวคิด **Label Self-expressiveness**: นำมาใช้เพื่อเพิ่มความถูกต้องของ Assignment Matrix โดยให้มีโครงสร้างที่สอดคล้องกับข้อมูลต้นฉบับ ส่งผลให้การจัดกลุ่มมีประสิทธิภาพมากขึ้น 2. **Semi-supervised Subspace Clustering**: ถูกนำมาใช้เพื่อลดข้อจำกัดของ Unsupervised Clustering แบ่งเป็นสองประเภท ได้แก่ Pairwise Constraints (PC) และ Graph-based Label Propagation (GFHF) อย่างไรก็ตาม วิธีเหล่านี้ยังใช้ Static Affinity Matrix ซึ่งไม่สามารถปรับเปลี่ยนตามข้อมูลได้ 3. **Graph Convolutional Network (GCN)**: ใช้เพื่อเรียนรู้โครงสร้างข้อมูล ทำให้สามารถสร้าง Assignment Matrix ที่แม่นยำขึ้น และลดผลกระทบจาก Noise และ Outliers ได้ 4. **Game Theory และ Nash Equilibrium**: นำมาใช้ในการเรียนรู้ Assignment Matrix และ Affinity Matrix ไปพร้อมกัน โดยให้ทั้งสองโมดูลแข่งขันกันจนกระทั่งถึง Nash Equilibrium ซึ่งเป็นจุดที่โมเดลเสถียรที่สุด วิธีนี้ช่วยให้ Affinity Matrix อัปเดตได้แบบไดนามิก และปรับตัวตามข้อมูลอย่างเหมาะสม

งานวิจัยนี้ทดสอบโมเดลบน 4 ชุดข้อมูลมาตรฐาน ได้แก่ ORL, USPS, Yale และ COIL20 ซึ่งมีข้อมูลภาพใบหน้าตัวเลขดิจิทัล และวัตถุ โดยดำเนินการทดลอง คือ 1. **Baseline Comparison**: การทดลองเปรียบเทียบ GMSSC กับวิธีการ semi-supervised subspace clustering อื่น ๆ เช่น STSSL, NNLR, DCSSC, SSLRR, DPLRR และ SSC_TLRR โดยใช้เปอร์เซ็นต์ของ labels ที่ทราบล่วงหน้าตั้งแต่ 5% - 30% ผลลัพธ์ที่ออกมาแสดงให้เห็นว่า GMSSC มีความแม่นยำสูงกว่าในกรณีส่วนใหญ่ 2. **Clustering Results**: GMSSC ให้ผลลัพธ์ที่ดีกว่าวิธีอื่น ๆ โดยเฉพาะในชุดข้อมูล Yale ที่โมเดลสามารถเพิ่มความแม่นยำได้มากกว่าวิธีที่ติดอันดับสอง 3. **Ablation Study**: ได้มีการวิเคราะห์องค์ประกอบสำคัญของ Multi-scale Regularization (MSR) และ Label Self-expressiveness (LS) พบว่า การตัดองค์ประกอบใดจะออกส่งผลให้ค่าความถูกต้องลดลง แสดงให้เห็นถึงความสำคัญของการใช้ทั้งสององค์ประกอบร่วมกัน 4. **Parameter Analysis**: ทำการศึกษาผลกระทบของพารามิเตอร์ที่ใช้ในการเรียนรู้ Affinity Matrix และ Label Self-expressiveness พบว่า GMSSC สามารถรักษาประสิทธิภาพได้ดีแม้มีการเปลี่ยนแปลงค่าพารามิเตอร์ในช่วงกว้าง แสดงถึงความทนทานของโมเดล 5. **Convergence Analysis**: ทำการวิเคราะห์การลู่เข้าของ GMSSC แสดงให้เห็นว่า ค่าความถูกต้องของการจัดกลุ่มข้อมูลและฟังก์ชันเป้าหมายสามารถเข้าสู่ค่าที่เสถียรได้อย่างรวดเร็ว สะท้อนถึงประสิทธิภาพของกระบวนการฝึกโมเดล 6. **Robustness Analysis**: ทำการทดลองกับชุดข้อมูล USPS โดยเพิ่ม Gaussian Noise, Block-masking และ Salt & Pepper Noise ซึ่ง GMSSC สามารถรักษาค่าความแม่นยำของการจัดกลุ่มข้อมูลได้ดีกว่า

จากผลการทดลองแสดงให้เห็นว่าโมเดล GMSSC มีประสิทธิภาพสูงกว่าวิธีดั้งเดิม เช่น K-means และ Spectral Clustering โดยเฉพาะในกรณีที่มีข้อมูลปายากกักบัง นอกจากนั้น โมเดลสามารถเพิ่มความแม่นยำในการจัดกลุ่ม ทนทานต่อ noise และรองรับข้อมูลที่มีโครงสร้างซับซ้อน ซึ่งแนวทางในอนาคตอาจเน้นไปที่การเพิ่มประสิทธิภาพด้านการคำนวณ การประยุกต์ใช้กับชุดข้อมูลที่หลากหลายขึ้น และการพัฒนาโมเดลให้เหมาะสมกับงานที่ต้องจัดกลุ่มข้อมูลขนาดใหญ่ งานวิจัยนี้ถือเป็นแนวทางสำคัญในการพัฒนาเทคนิค Semi-supervised Subspace Clustering ให้มีประสิทธิภาพมากยิ่งขึ้น