Abstract

Recent advances in self-supervised representation learning have introduced VICReg (Variance-Invariance-Covariance Regularization), a method designed to maximize agreement between embedding vectors produced by different encoders to generate meaningful representations. VICReg outperforms other self-supervised methods and does not rely on collapse-prevention techniques (e.g., stop-gradient operations, memory banks, or output quantization) addressing the common representation learning problem where models produce nearly identical embeddings for all inputs, yielding meaningless representations. Initially proposed for computer vision, VICReg uses differently augmented views of the same image, training their representations to align. In this thesis, VICReg is adapted to computational chemistry, where different molecular modalities serve as distinct views in a multi-modal training approach. Specifically: Graph Isomorphism Networks (GIN) capture graph-structured data, Equivariant Graph Neural Networks (EGNN) represent conformational data, and Long Short-Term Memory (LSTM) networks encode chemical language representations. The resulting modality-specific encoders are evaluated after VICReg pre-training through both linear and transfer learning evaluation protocols on an independent Absorption, Distribution, Metabolism, and Excretion (ADME) dataset. Additional experiments compare against model-agnostic supervised pre-training and modality-specific self-supervised pre-training using identical evaluation protocols. This work addresses several research questions: (i) How do supervised model-agnostic pre-training strategies compare to self-supervised, modality-specific approaches? (ii) Can VICReg be successfully adapted from computer vision to computational chemistry using multi-modal molecular representations without collapse? (iii) Which experimental configurations optimize downstream performance with VICReg? (iv) Does VICReg outperform both modality-specific and model-agnostic pre-training strategies in downstream tasks?

Zusammenfassung

VICReg (Variance-Invariance-Covariance Regularization) beschreibt eine Technik des selbstüberwachten Repräsentationslernens, um die Übereinstimmung zwischen von unterschiedlichen Encodern erzeugten Embedding-Vektoren zu maximieren. VICReg übertrifft andere selbstüberwachte Methoden und ist nicht auf Techniken angewiesen, um den sogenannten „Collapse"-Effekt zu vermeiden (z.B. Stop-Gradient-Operationen, Memory Banks oder Output-Quantisierung), ein Phänomen, bei dem das Modell nahezu identische Embeddings für alle Eingaben erzeugt. VICReg kommt ursprünglich aus der Computer Vision Domäne und nutzt unterschiedlich augmentierte Ansichten desselben Bildes, um die encodeten Repräsentationen unter einander anzugleichen. In dieser Arbeit wird VICReg auf den Bereich der Chemoinformatik übertragen, indem verschiedene Modalitäten eines Moleküls als unterschiedliche Ansichten in einem multimodalen Trainingsansatz dienen. Konkret erfassen Graph Isomorphism Networks (GIN) die Graph-Modalität, Equivariant Graph Neural Networks (EGNN) repräsentieren die Konformer-Modalität, und Long Short-Term Memory (LSTM)-Netzwerke kodieren die chemische Sprachmodalität. Die daraus resultierenden, modalitätsspezifischen Encoder werden nach dem VICReg Pre-training mithilfe einer linearen Evaluierung und einer Transfer-Learning-Evaluierung auf einem unabhängigen Datensatz für Absorption, Distribution, Metabolism und Excretion (ADME) getestet. Zudem wird in weiteren Experimenten jeder Encoder in einem modellunabhängigen, überwachten Pre-training und in einem modalitätsspezifischen, selbstüberwachten Pre-training trainiert und mit derselben Evaluierungsstrategie evaluiert. Abschließend erörtert diese Thesis verschiedene Forschungsfragen: (i) Wie vergleichbar sind überwachte, modellunabhängige Pre-training-Strategien gegenüber selbstüberwachten, modellspezifischen Pre-training-Strategien? (ii) Lässt sich VICReg erfolgreich von der Computer Vision auf die Chemoinformatik übertragen, indem Graph-, Konformer- und Textmodalitäten eines Moleküls genutzt werden? (iii) Welche VICReg Einstellungen führen zu besseren Downstream-Performances? (iv) Übertrifft VICReg sowohl modalitätsspezifisches Pre-training als auch modellunabhängige Pre-training-Strategien in Bezug auf die Downstream-Performance?