

Tossaporn Sirampuj 2013: Integrating Constraints in Semi-supervised Stream Clustering. Master of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Associate Professor Kitsana Waiyamai, D.U. 57 pages.

Large number of stream clustering techniques has been proposed in recent years. However, these techniques still lack of using background knowledge which is available from domain expert. The advantages of using knowledge improve accuracy and performance of final clusters. In this research work, CE-Stream, an incremental method for stream clustering by using background knowledge as constraints is proposed.

Instance-level constraints have been used to guide better clustering behaviors i.e. using Must-Link constraints on Cluster Splitting and using Cannot-Link constraints on Cluster Assignment. Also, constraint operators are introduced to support evolving characteristics of dynamic constraints i.e. constraint activation, fading and outdating. Constraints operators seamlessly integrate into E-Stream to check active and update constraints and prioritize constraints. Likewise, CE-Stream reduces an excessive splitting during clustering process and misguide of cluster assignment.

Experimental results, using Coverttype and KDDCup'99 datasets which are Rain forest typology by geographical and Network Intrusion Detection respectively, show that both F-measure and Purity increased with respect to an original technique E-Stream.

---

Student's signature

---

Thesis Advisor's signature