Polyhedral clustering

Summary

The title of this thesis, Polyhedral clustering, is twofold. On the one hand it refers to the polyhedral approach to clustering problems, and on the other hand it refers to a cluster of related polyhedra. Both aspects are treated in this thesis.

The polyhedral approach is applied to two different clustering problems. First, the clique partitioning problem is considered, which can be formulated as follows. Given is a graph $G = (V, E)$ with weights, both positive and negative, for all edges of the graph. Find a partition of V, such that each element of the partition induces a clique in G and the sum of the weights on the edges within the elements of the partition is maximized. The clique partitioning polytope $P(G)$ is the convex hull of all edge sets of clique partitions of G. This gives an implicit description of this polytope. Polyhedral combinatorics is the art of finding an *explicit description*. Several techniques from this branch of mathematics are used to reveal the structure of the clique partitioning polytope, such as establishing ties with other polyhedra whose structure is already (partially) known, identifying new classes of facets, generalizing individual facets to classes, and lifting facets of the clique partitioning polytope of subgraphs to facets of the polytope corresponding to the original graph. Next to these techniques patching is used, a new technique for constructing facets from facets. Patching combines the structure of a number of facets to obtain a new facet. One could say that the known structure of a polyhedron is used to reveal part of the unknown structure. Another aspect of the clique partitioning problem that is considered is separation, that is: given a graph $G = (V, E)$ and a point $x \in [0,1]^{|E|}$, find a hyperplane in $\mathbb{R}^{|E|}$ separating $P(G)$ and x, or prove that no such hyperplane exists. For some classes of facets the separation problem is NP-hard, whereas for other classes the separation problem is polynomially solvable.

The second clustering problem considered in this thesis is the graph decomposition problem, which can be formulated as follows. Given is a graph $G = (V, E)$ and an integer $c \geq 1$. Find a subset $S \subseteq V$ of minimal size, such that the graph obtained from G by deleting all vertices in S and all edges incident to S has no component containing more than c vertices. For $c = 1$ this problem is equivalent to finding a maximal set of independent vertices in G, and therefore the graph decomposition problem is NP-hard in general. The structure of the corresponding polytope $P(G, c)$ is revealed with help of the same techniques as used for the clique partitioning polytope, in particular lifting and establishing ties with other polyhedra. Furthermore, it is examined how useful the known structure of this polytope is to solve instances of this clustering problem.

The second aspect of polyhedral clustering, namely clusters of related polyhedra, is the subject of research in the first and the last chapter of this thesis. The clique partitioning polytope belongs to a family of polyhedra, which also contains the clique polytope, the (multi)cut polytope, and the bipartite subgraph polytope. A more extended overview of this family of polyhedra is given in Figure 1.3. Another cluster is formed by the set of $\{0, 1\}$-polyhedra for which patching is a useful technique to investigate the structure. Different variations of patching for these kinds of polytopes are described in the last chapter.

Samenvatting

De titel van dit proefschrift, Polyhedrale clustering, is tweeledig. Enerzijds verwijst zij naar de polyhedrale benadering van clusteringsproblemen, anderzijds verwijst zij naar een cluster van gerelateerde polyhedra. Beide aspecten komen in dit proefschrift aan bod.

De polyhedrale benadering wordt toegepast op twee verschillende clusteringsproblemen. Als eerste wordt het kliek-partitie probleem beschouwd, dat als volgt geformuleerd kan worden. Gegeven is een graaf $G = (V, E)$ met gewichten, zowel positief als negatief, voor alle kanten. Gevraagd wordt een partitie van V zodanig dat elk element van de partitie een kliek in G induceert, en de som van de gewichten op de kanten binnen de elementen van de partitie maximaal is. Het kliek-partitie polytoop $P(G)$ is de convex omhullende van alle kanten-verzamelingen van kliek-partities van G. Dit geeft een impliciete beschrijving van dit polytoop. Polyhedrale combinatoriek is de kunst van het vinden van een expliciete beschrijving. Er worden diverse technieken uit deze tak van de wiskunde gebruikt om de structuur van het kliek-partitie polytoop bloot te leggen, zoals het leggen van verbanden met andere polyhedra waarvan de structuur al (gedeeltelijk) bekend is, het identificeren van nieuwe klassen van facetten, het generaliseren van individuele facetten tot uitgebreide klassen, en het liften van facetten van het kliek-partitie polytoop van deelgrafen tot facetten van het polytoop dat correspondeert met de oorspronkelijke graaf. Daarnaast wordt er gebruikt gemaakt van patching, een nieuwe techniek voor het construeren van facetten. Bij patching wordt de structuur van meerdere facetten gecombineerd tot een nieuw facet. Men zou kunnen zeggen dat de bekende structuur van een polyhedron gebruikt wordt om een deel van de onbekende structuur bloot te leggen. Een ander aspect van het kliek-partitie probleem dat belicht wordt is separatie, dat wil zeggen: gegeven een graaf $G = (V, E)$ en een punt $x \in [0,1]^{|E|}$, vind een hypervlak in $\mathbb{R}^{|E|}$ die $P(G)$ en

x separeert, of bewijs dat zo'n hypervlak niet bestaat. Voor sommige klassen van facetten is het separatieprobleem NP-lastig, terwijl het voor andere klassen polynomiaal oplosbaar is.

Het tweede clusteringsprobleem dat in dit proefschrift beschouwd wordt is het graaf-decompositie probleem, dat als volgt geformuleerd kan worden. Gegeven is een graaf $G = (V, E)$ en een geheel getal $c \geq 1$, gevraagd wordt een deelverzameling $S \subseteq V$ van minimale grootte, zodanig dat de graaf die uit G ontstaat door alle punten uit S en alle kanten incident met S te verwijderen geen componenten heeft die meer dan c punten bevatten. Voor $c = 1$ is dit probleem equivalent met het vinden van een maximale onafhankelijke verzameling punten in G, en bijgevolg is het graaf-decompositie probleem in het algemeen NP-lastig. De structuur van het bijbehorende polytoop $P(G, c)$ wordt bloot gelegd met behulp van dezelfde technieken die zijn gebruikt voor het kliek-partitie polytoop, met name lifting en het leggen van verbanden met andere polyhedra. Daarnaast wordt nagegaan hoe bruikbaar de bekende structuur van dit polytoop is om instanties van dit clusteringsprobleem op te lossen.

Het tweede aspect van polyhedrale clustering, namelijk verzamelingen van gerelateerde polyhedra, komt aan bod in het eerste en het laatste hoofdstuk van dit proefschrift. Het kliek-partitie polytoop behoort tot een familie van polyhedra, waartoe ook het kliekpolytoop, het (meervoudige) snedenpolytoop, en het bipartiete deelgraaf-polytoop behoren. Een uitgebreider overzicht van deze familie van polyhedra is gegeven in Figuur 1.3. Een ander cluster wordt gevormd door de verzameling van $\{0, 1\}$-polytopen waarvoor patching een geschikte techniek is om de stuctuur te onderzoeken. Verschillende varianten van patching voor deze polytopen worden beschreven in het laatste hoofdstuk.