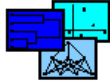


Gesellschaft für
Klassifikation e.V.



GfKI

<http://www.gfkl.de>

Arbeitsgruppe

Datenanalyse und Numerische Klassifikation (AG-DANK)

Arbeitsgebiete • Aktivitäten • Adressen
November 2012

Information/ Anschriften

Website der AG-DANK

<http://www.homepages.ucl.ac.uk/~ucakche/agdank/>

Leiter der AG-DANK:

Dr. Christian Hennig
University College London
Department of Statistical Science
Gover Street, London WC1E 6BT
United Kingdom
Tel: 0044-207-679-1862
E-mail: c.hennig@stats.ucl.ac.uk

Stellvertretender Leiter der AG-DANK:

Prof. Alfred Ultsch
Philipps Universität Marburg
Fb. 12 - Mathematik und Informatik
Hans-Meerwein-Straße
D-35032 Marburg
Deutschland

Das Arbeitsgebiet der AG-DANK

Daten und Informationen sind Grundlage jeder empirischen und angewandten *Forschung*, oft bilden sie den Ausgangspunkt wissenschaftlicher Erkenntnis und induktiver Theoriebildung. In der *Wirtschaft* und *Technik* sind Daten unverzichtbare Basis für wirtschaftliche und betriebliche Entscheidungen und die Planung technischer Anlagen und Prozesse.

Riesige *Datenbanken* sollen hierzu die jeweils verfügbare oder gesammelte Information bereitstellen.

Eine sinnvolle Nutzung von Daten und Informationen setzt indessen voraus, daß der (in der Regel umfassende) Datenbestand zuvor geeignet aufbereitet und nach den sachlichen und anwendungsorientierten Gesichtspunkten geordnet wurde. Dazu gehören etwa eine detaillierte *Analyse* der erfaßten Merkmale und ihrer Kenngrößen, die Entdeckung und Beschreibung von *daten-immanenten Beziehungen* zwischen verschiedenen Merkmalen, die Erkennung und Unterscheidung von unterschiedlich gearteten *Teilpopulationen* (Clustern, Objekttypen) und schließlich eine geeignete *Visualisierung* gefundener Beziehungen und Heterogenitäten in Form von übersichtlichen Grafiken und Diagrammen.

Erst eine solche Analyse ermöglicht es, den Informationsgehalt der Daten in Bezug auf die Anwendung zu beurteilen und die wesentlichen Charakteristika der Daten von Artefakten und Zufallsmustern zu unterscheiden. Es resultieren dann möglicherweise neue und interessante *Hypothesen* für den zukünftigen Forschungs- oder Arbeitsprozeß.

Ausgangspunkt solcher Untersuchungen und Analysen sind Datentabellen, Dokumentenbeschreibungen, relationale oder unscharfe Daten, aber auch sonstige quantitative wie qualitative Informationen. In neuerer Zeit werden insbesondere Textdaten, Signale, Szenen und Bilder, symbolische Objekte oder probabilistische Merkmale bearbeitet. Die verwendeten Methoden und Denksätze orientieren sich – je nach Problem- und Datentyp – an probabilistischen, algebraischen, geometrischen oder informatiknahen Konzepten. Oft ist das Ziel die Entwicklung *effizienter Algorithmen*.

Interessen- und Themenschwerpunkte der AG-DANK

- **Clusteranalyse, numerische Taxonomie** oder **unüberwachte Klassifikation**, d.h. die (automatische) Erstellung einer Klassifikation von Objekten mit “homogenen” Objektclustern oder Typen;
- **Diskriminanzanalyse** oder **überwachte Klassifikation**, d.h. die (automatische) Zuordnung von Objekten zu durch Regeln oder Funktionen beschriebenen Klassen aufgrund erhobener Merkmale und Eigenschaften der Objekte;
- **hierarchische Verfahren** über durch Ähnlichkeits- und Unähnlichkeitsmatrizen dargestellte Beziehungen zwischen den Objekten;
- multidimensionale, ordinale und algebraische **Skalierungsmethoden** für Merkmale;
- (statistische) **Modellbildung** und die Anpassung von theoretischen Modellen an empirische Daten, Parameterschätzung;
- **Vektorquantisierung**;
- **Lernverfahren**, z.B. für **neuronale Netze** oder **Support-Vector-Maschinen**, Vapnik-Chervonenkis-Theorie;
- selbstorganisierende Karten, **SOM**;
- **explorative Methoden** der Datenanalyse;
- Analyse **longitudinaler Daten** und **funktionale Datenanalyse**;
- Klassifikation von **Zeitreihen**;
- **Stabilität** und **Robustheit** datenanalytischer Verfahren
- **Kausal- und Zusammenhangsanalyse**;
- Auswertung **unscharfer** oder vager **Informationen**;
- Bestimmung zweckmäßiger **Systematiken**, etwa in Form von Hierarchien, Begriffsstrukturen oder Mengensystemen;
- Strategien zur **Evaluation** numerischer Ergebnisse (z.B. Tests, Simulation, Resampling-Methoden);

- **graphische Repräsentation** von Daten, Ähnlichkeitsbeziehungen und Relationen zwischen Objekten oder Personen in Form von Punkte- und Liniendiagrammen, Abhängigkeitsnetzen, Hierarchien etc.;
- **Optimierungsverfahren** zur Bestimmung von modellbasierten Klassifikationen;
- **Mustererkennung**, die Klassifikation von komplexen Objekten wie Bildern und Tonsignalen über die Extraktion und Reduktion von Merkmalen;
- automatische **Sprach- und Bildererkennung**;
- die in der Wirtschaftsinformatik als **Data Mining** bezeichnete Gewinnung von Information aus grossen und unübersichtlichen Datensätzen, Knowledge Discovery;
- quantitative Analyse zeitlicher und räumlicher Vorgänge, auch im Rahmen von **geographischen Informations-Systemen** (GIS);
- klassifikatorische und statistische Analyse von Zeichenketten, etwa aus **Genetik und Proteomik**;
- Analyse von **Microarray-Daten**.
- Realisierung von Klassifikations- und Datenanalyse-Methoden mittels geeigneter **Software**.

Das resultierende *Wissen* unterstützt den Anwender und Forscher bei seinen Entscheidungen, Planungen und Diskussionen, und die graphische Visualisierung erlaubt es, die in den Daten enthaltene, oft sehr komplexe Information auch einem weniger sachkundigen Publikum zu vermitteln.

Wesentliche Motivation der AG-DANK ist es, zu einem möglichst umfassenden Verständnis der datenanalytischen Methoden zu gelangen. Mathematische Theorie und Simulationen sind hierfür ebenso nützlich wie die Arbeit an Anwendungsproblemen mit ihren unterschiedlichen Randbedingungen auf den verschiedensten Gebieten. Die Arbeitsgruppe fördert die Anwendung solcher Methoden z.B. auf den folgenden Bereichen: Archäologie, Biologie, Chemie, Informationswissenschaften, Kriminalistik, Linguistik, Medizin, Psychologie, Sozialwissenschaften, Technik, Verwaltung, Marketing, Wirtschaft.

Aktivitäten der AG-DANK

Die Arbeit der AG-DANK erfolgt einerseits im Rahmen der persönlichen Kontakte ihrer Mitglieder, andererseits im Rahmen der *Jahrestagungen* der GfKl sowie der jährlichen *Herbsttagungen* der AG-DANK.

Die AG-DANK organisiert regelmäßig Vortragssektionen im Rahmen der *Jahrestagungen* der GfKl, wo Spezialergebnisse präsentiert werden. Die Jahrestagungen dienen dazu, die in der GfKl vorhandenen Interessenschwerpunkte und Arbeitskreise im Hinblick auf eine interdisziplinäre Diskussion und Kooperation zusammenzuführen, insbesondere zwischen

- Anwendern und Entwicklern von Methoden,
- Anwendern verschiedener Fachrichtungen,
- Datenanalytikern, Mathematikern, Statistikern und Informatikern,
- Spezialisten verschiedener Länder.

Demgegenüber sind die jährlichen *Herbsttagungen* der AG-DANK eher speziellen Einzelthemen gewidmet. Der Schwerpunkt liegt dabei auf Übersichts- und Spezialvorträgen, der Erörterung von Einzelproblemen im Expertenkreis sowie auf der Diskussion einzelner Datensätze.

Frühere Veranstaltungen (Auszug):

- Herbsttagung 1995 (Heidelberg), gemeinsam mit der AG 'Statistische Methoden in der Medizin' der Biometrischen Gesellschaft und der AG 'Statistische Methoden' der GMDS.

- Tutorial 'Neuronale Netze zur Klassifikation' bei der 20. Jahrestagung 1996 (Universität Freiburg)
- Herbsttagung 1996 (Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin)
- Herbsttagung 1997 (Universität Passau)
- Tutorial 'Anwendungen der Datenanalyse in der Mobilfunk-Telekommunikation' bei der 22. Jahrestagung 1998 (Dresden)
- Herbsttagung 1998 (Universität Regensburg)
- Herbsttagung 1999 (Universität Hamburg)
- Herbsttagung 2000 (Universität Aachen)
- Herbsttagung 2001 (Technische Universität Berlin)
- Herbsttagung 2002 (Universität Bonn)
- Herbsttagung 2003 (Universität Düsseldorf)
- Herbsttagung 2004 (Universität Marburg)
- Herbsttagung 2005 (Infratest München)
- Herbsttagung 2006 (Universität Dortmund)
- Herbsttagung 2007 (Universität Dresden)
- Herbsttagung 2008 (Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin)
- Herbsttagung 2009 (RWTH Aachen, zusammen mit dem First German-Polish Symposium on Data Analysis and its Applications)
- Herbsttagung 2010 (Florenz, gemeinsam mit der Cladag)
- Herbsttagung 2011 (Universität Düsseldorf; Schwerpunkt medizinische Daten)
- Herbsttagung 2012 (LVR-Amt für Bodendenkmalpflege im Rheinland, Bonn; Schwerpunkt archäologische Daten)
- Herbsttagung 2013 (geplant: University College London, mit der British Classification Society)

Ergebnisse und Vorträge der Tagungen der AG-DANK werden in Form von *Publikationen* oder auf der *Website* der AG-DANK veröffentlicht.

Geschichte und Organisation

Die Arbeitsgruppe '*Datenanalyse und Numerische Klassifikation*' ist als formelle Arbeitsgruppe der GfKl konzipiert. Sie besteht seit 1979 und vereint zur Zeit ca. 200 Personen inner- und außerhalb der GfKl, die sich in Forschung und Praxis mit mathematisch-statistischen Problemen der Datenanalyse und Klassifikation sowie deren Anwendungen beschäftigen.

Die Mitarbeit in der AG kann durch ein formloses Schreiben an den Leiter der AG erklärt werden. Mitgliedschaft in der GfKl ist zwar erwünscht, aber keine Bedingung. Die AG erhebt keinen gesonderten Mitgliedsbeitrag.

Bei jeder Jahrestagung der GfKl findet eine *Mitgliederbesprechung* statt, auf der alle anstehenden Fragen diskutiert werden und das Arbeitsprogramm festgelegt wird.

Der Leiter der AG und sein Stellvertreter werden alle drei Jahre bei einer Mitgliederbesprechung der AG-DANK gewählt.