

Gewinnung höherwertiger Information aus Low-Level Datenströmen (DataStreamMiner)

Beteiligte an der Hochschule

- Prof. Dr. Reinhold Kröger
- Dipl.-Inform. (FH) Jan Schäfer, M.Sc.
- Dipl.-Inform. (Univ. Brasov) Anda Ignat
- Dipl.-Inform. (Univ. Brasov) George Moldovan
- Dipl.-Inform. (FH) Markus Fischer
- Dipl.-Inform. (FH) Marcus Thoss M.Sc.
- Fabian Meyer B.Sc.
- Stefan Berntheisel B.Sc.
- Dipl.-Inform. (FH) Patrick Deuster
- Andreas Textor B.Sc.
- Dipl.-Inform. (FH) Michael Frey
- Christoph Anders B.Sc.

Kooperationspartner

- DFS Deutsche Flugsicherung GmbH
- blue elephant systems GmbH

Laufzeit

Beginn: Juli 2009

Ende: Februar 2012

Finanzierung

- 100 % BMBF im Rahmen der Initiative FHprofUnd (Förderkennzeichen 1727X09)

Veröffentlichungen

- Ignat, Anda: "Ein Data-Stream-Mining-Ansatz zum proaktiven Management von kritischen IT-Anwendungen", M.Sc. Thesis, FH Wiesbaden, FB Design Informatik Medien, November 2009
- Moldovan, George: "Proaktives Management einer kritischen IT-Anwendung unter Nutzung von Semantic Web-Ansätzen", M.Sc. Thesis, FH Wiesbaden, FB Design Informatik Medien, November 2009
- Großmann, Peter: "Klassifizierung von Laufzeit-Zuständen kritischer Anwendungen mit Hilfe von Bayes-Netzen", interne Diplomarbeit, FH

Wiesbaden, FB Design Informatik Medien, November 2009

- Fischer, Markus: "Ein Framework für dynamische, hochgradig parallel arbeitende Data Stream Mining Strukturen", M.Sc. Thesis, Hochschule RheinMain, FB Design Informatik Medien, Mai 2010
- Meyer, Fabian: "Analyse von Event Streams im Kontext semantischer Systemmodelle", M.Sc. Thesis, Hochschule RheinMain, FB Design Informatik Medien, September 2010
- Frey, Michael: "Ein Ansatz für eine Spezifikationssprache zur Generierung von Verarbeitungsvorschriften in heterogenen Complex-Event-Processing-Systemen", M.Sc. Thesis, Hochschule RheinMain, FB Design Informatik Medien, Oktober 2010
- Berntheisel, Stefan: "Vergleich von CEP-Systemen im Hinblick auf den Einsatz im IT-Management", M.Sc. Thesis, Hochschule RheinMain, FB Design Informatik Medien, Mai 2011

Kurzbeschreibung

Ziel des Vorhabens ist die Entwicklung von Methoden und entsprechenden Werkzeugen zur Online-Analyse von Datenströmen sowie deren Anwendung in den beiden Bereichen IT-Management und Ambient Assisted Living (AAL).

Das klassische Data Mining basiert darauf, einen persistent, i.d.R. in einer Datenbank gehaltenen, oft sehr großen Datenbestand mit meist mathematischen und statistischen Methoden in Hinblick auf die Erkennung von Mustern/Struktur und Besonderheiten zu untersuchen. Oft handelt es sich dabei um Geschäftsdaten, die Erkennung von Trends, von Korrelationen zwischen verschiedenen Größen, Gruppenbildung (Clustering), usw.. Charakteristisch ist, dass die Daten vollständig in der Datenbasis vorliegen und i.d.R. kein Modell für die Datenentstehung vorhanden ist.

Seit einigen Jahren sind jedoch immer häufiger Anwendungen im Einsatz, die kontinuierlich Daten, oft mit einer hohen Datenrate, in Form von transienten Datenströmen, sogenannten Data Streams, liefern. Der Auswertung dieser Datenströme widmet sich das sich erst in den letzten Jahren gebildete Forschungsgebiet "Data Stream Mining". Beispiele solcher Datenströme sind sehr vielfältig wie etwa Paketströme in Rechnernetzen, Messdaten aus Sensornetzen einschl. Systemen der Telemedizin zur Erfassung von lebenswichtigen Körperfunktionen oder Smart Metering im häuslichen Bereich, Folgen von Log/Trace-Records von IT-Anwendungen oder Betriebssystemen, fortlaufende Koordinaten oder Vektoren sich bewogender Objekte, Klick-Folgen auf Web-Seiten, usw. Ziel der Analyse dieser Low-Level-Datenströme sind höherwertige Informationen, z.B. Intrusion Detection (Einbruchsversuche von Hackern in

Datennetze), Erkennen von lebensbedrohenden Gesundheitszuständen, Ermittlung der Dienstgüte von IT-Anwendungen, Accounting in Anspruch genommener Leistungen, Klassifizierung von Fehlersituationen, Spam-Erkennung, Kollisionsgefahr sich bewegender Objekte, Erkennung und Abrechnung der Nutzung von Werbeelementen auf Web-Seiten usw.. Charakteristisch ist, dass häufig Modelle der Anwendungsdomäne vorhanden sind, dass zu jedem Zeitpunkt nur ein begrenztes Fenster von Datensätzen im kontinuierlichen Datenstrom zur Verfügung steht und dass die Verarbeitung der Daten online, d.h. unmittelbar nach ihrer Entstehung und unter Umständen sogar in Echtzeit vorgenommen werden muss.

Das Vorhaben wird durch einen Verbund erbracht, bestehend aus der Fachhochschule Wiesbaden (Labor für Verteilte Systeme), einem KMU (blue elephant systems GmbH (BES), Stuttgart) sowie dem Systemhaus der bundeseigenen DFS Deutsche Flugsicherung GmbH in Langen im Rhein/Main-Gebiet. Mit beiden Partnern werden anwendungsbezogene Problemstellungen aus dem Bereich des IT-Managements bearbeitet, an denen die zu entwickelnden Methoden und Werkzeuge validiert werden.