



FINE-CHEM – AN INVENTORY ESTIMATION TOOL FOR FINE CHEMICALS

Jahresbericht 2006

Autor und Koautoren	Gregor Wernet, Ulrich Fischer, Konrad Hungerbühler
beauftragte Institution	ETH Zürich
Adresse	HCI G143, Wolfgang-Pauli-Strasse 10, 8093 Zürich
Telefon, E-mail, Internetadresse	044/632 34 71, gregor.wernet@chem.ethz.ch , http://www.sust-chem.ethz.ch
BFE Projekt-/Vertrag-Nummer	101711
BFE-Projektbegleiter	Martin Stettler
Dauer des Projekts (von – bis)	Juni 2006 – Mai 2009
Datum	30.11.06

ZUSAMMENFASSUNG

Ziele für das zweite Halbjahr 2006 waren die Datengewinnung und Verhandlungen mit den Kooperationspartnern aus der chemischen Industrie sowie die Entwicklung eines neuen Modells, um Parameter wie kumulierten Energiebedarf (CED) oder Umweltfolgen der Produktion einer Chemikalie zu bestimmen. Das Modell basiert auf der Analyse von Strukturparametern chemischer Substanzen.

Durch die Anwendung mehrerer Regressionsmodelle (multiple, lineare Regression sowie neuronale Netze) konnten Modelle erschaffen werden, durch die der CED und andere Parameter direkt aus der Molekülstruktur einer Chemikalie und ohne Informationen über den Produktionsablauf bestimmt werden können. Aus den Ergebnissen der linearen Modelle lassen sich direkt generelle, quantitative Aussagen über die Einflüsse einzelner funktioneller Gruppen machen. Die Modelle auf der Basis neuronaler Netze zeigen sich äusserst viel versprechend und erlauben eine sehr genaue Bestimmung von Parametern wie dem CED.

In mehreren Treffen mit den nationalen und internationalen Partnern aus der chemischen Industrie konnten die Partner überzeugt werden, das Projekt mit Daten zu unterstützen, auch wenn dies zusätzlichen Aufwand von Seiten der Partner bedeutet. Einige der Projektpartner haben bereits grosszügige Datenmengen zur Verfügung gestellt. Andere haben Mitarbeiter dazu abgestellt, innerhalb der Unternehmen Daten für die Unterstützung des Projektes zusammen zu tragen.

Projektziele

Projektziel ist ein Tool für die schnelle und unkomplizierte Erstellung von Inventardaten (Massen- und Energiebilanzen) der Produktion von Feinchemikalien. Da die herkömmliche Methode einer Inventarisierung bei Feinchemikalien zu einem extremen Arbeitsaufwand führen würde, werden Inventarisierungen nur selten und im Rahmen von Fallstudien durchgeführt. Das angestrebte Tool soll eine Abschätzung von Inventardaten bereits in der Phase der Produktplanung ermöglichen. Dadurch werden sich mehrere Alternativprodukte bezüglich des wahrscheinlichen Energieverbrauches bei der Produktion vergleichen lassen. Des Weiteren können verschiedene Produktionsmöglichkeiten für ein neues Produkt verglichen werden, so dass energieaufwändige Prozesse frühzeitig identifiziert und vermieden werden können. Zuletzt werden auch bestehende Prozesse ohne aufwändige Messungen auf übermässigen Energieverbrauch hin untersucht werden können, um eine Prozessoptimierung zu erleichtern.

Die Teilziele des Projektes für den Zeitraum Juni 2006 bis November 2006 waren:

Datengewinnung und Verhandlungen mit den Kooperationspartnern aus der chemischen Industrie. Für die Entwicklung von verlässlichen Modellen für die Produktion von Feinchemikalien sind Produktionsdaten aus der chemischen Industrie essentiell. In der zweiten Jahreshälfte 2006 sollten Einzelbesprechungen mit den Industriepartnern klarstellen, in welcher Weise die Partner das Projekt mit Daten unterstützen können.

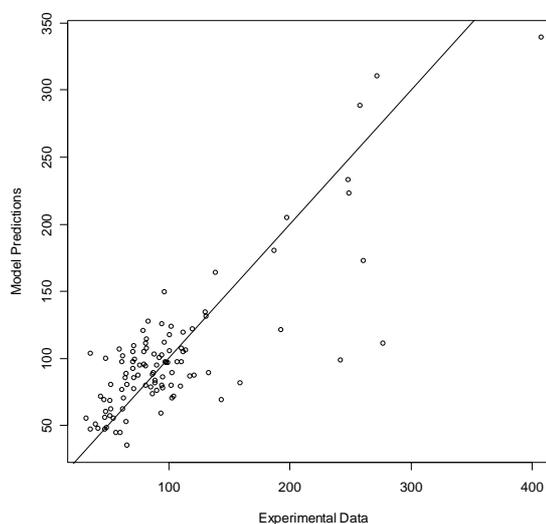
Analyse von Strukturparametern chemischer Substanzen. Eines der vorgeschlagenen Modelle basiert auf dem neuartigen Ansatz, Produktionsparameter wie den Energieverbrauch direkt auf der Basis der Molekülstruktur vorherzusagen. Die Validierung dieses Ansatzes sollte an Hand von Daten aus der *ecoinvent*-Datenbank stattfinden. Neue Erkenntnisse über Molekülstrukturen und deren Produktionsaufwand wären ein zusätzlicher, wissenschaftlicher Nutzen aus diesem Teil des Projektes.

Durchgeführte Arbeiten und erreichte Ergebnisse

Auf der Basis von 103 Inventardatensätzen von Chemikalien (46 Basischemikalien aus der *ecoinvent*-Datenbank, 12 Pestizide aus der *ecoinvent*-Datenbank und 45 Lösungsmittel, die in einem anderen Projekt in der Hungerbühler-Gruppe inventarisiert wurden) wurde nach Korrelationen zwischen der Struktur von Chemikalien und Parametern für die Produktion der Chemikalie gesucht. Die Parameter waren unter anderem auch der kumulierte Energiebedarf (CED) und das Global Warming Potential (GWP).

Mehrere Modelle wurden getestet und die Resultate verglichen. Zuerst wurde eine multiple, lineare Regression durchgeführt. Dabei wurden verschiedene Sätze von Deskriptoren (die alle direkt aus der Molekülstruktur gewonnen wurden) angewendet, um zu bestimmen, welche Moleküldeskriptoren die besten Korrelationen liefern. Als Resultat wurde ein Satz von Deskriptoren gewählt, welcher aus dem Molekulargewicht sowie der Anzahl 16 verschiedener funktioneller Gruppen im Molekül besteht. Fig. 1 zeigt ein Ergebnis dieser multiplen, linearen Regression des CED.

Fig. 1: True-Predicted Plot der vorhandenen Inventardaten gegen die berechneten Werte des Modells. Vorhersageparameter ist der CED. Die Linie ist das Idealziel, d.h. Punkte, die auf oder dicht an der Linie liegen, werden vom Modell gut vorhergesagt. Das Modell basiert auf 17 Deskriptoren, die sich aus dem Molekulargewicht und der Anzahl 16 verschiedener funktioneller Gruppen im Molekül zusammensetzen.



Das Quadrat des Korrelationskoeffizienten nach Pearson ist für dieses Modell 0.67. Dies zeigt eindeutig, dass gewisse Zusammenhänge zwischen der Molekülstruktur und dem kumulierten Energiebedarf bestehen. Eine quantitative Vorhersage wäre bei diesem Modell jedoch mit einer grossen Unsicherheit behaftet. Daher wurde in einem nächsten Schritt ein neues Modell auf der Basis von künstlichen neuronalen Netzen entwickelt.

Die auf neuronalen Netzen basierenden Modelle haben den Vorteil, dass sie komplexe, nichtlineare Zusammenhänge deutlich besser erfassen können als klassische Regressionsmodelle. Ein Nachteil dieser Netze ist allerdings, dass sie auf Grund ihrer hohen Komplexität eine konsequente Analyse des Modells nicht zulassen. Es ist also nur mit grossem Aufwand möglich zu bestimmen, welche Anteile die funktionellen Gruppen jeweils am Resultat des Modells haben.

Der oben beschriebene Deskriptorensatz wurde zur Erstellung eines Modells auf der Basis neuronaler Netze verwendet. Es konnte ein Modell erstellt werden, welches den CED sehr gut bestimmen kann (Fig. 2). Der quadratische Korrelationskoeffizient beträgt für dieses Modell 0.91.

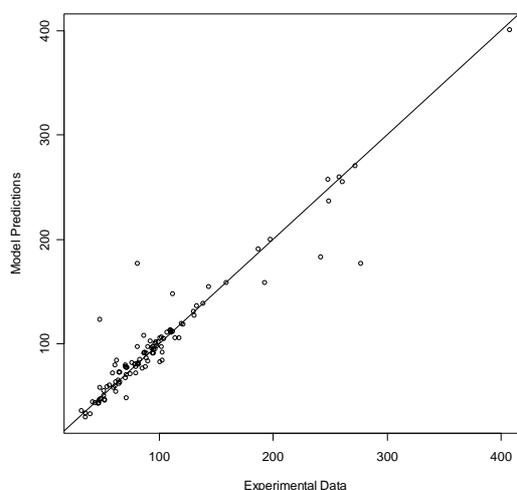


Fig. 2: True-Predicted-Plot des Modells auf der Basis neuronaler Netze. Es sind wieder die CED-Werte aus den vorhandenen Inventardaten gegen die vom Modell aus der Molekülstruktur bestimmten CED-Werte aufgetragen. Wie man erkennen kann, beschreibt das Modell den Datensatz deutlich besser als das lineare Regressionsmodell. Eine quantitative Vorhersage sollte auf der Basis dieses Modells möglich sein.

Nationale Zusammenarbeit

An dem Projekt FINE-CHEM beteiligen sich mehrere private Firmen aus der Schweizerischen chemischen Industrie. Namentlich sind dies Ciba SC, Hoffmann-La Roche, Novartis, Syngenta CP, Lonza. Alle Firmen haben sich bereit erklärt, Produktionsdaten zur Entwicklung des Tools beizutragen. Drei der Firmen beteiligen sich darüber hinaus auch finanziell an dem Projekt.

Im Zeitraum Juni 2006 bis Oktober 2006 fanden Einzeltreffen mit allen beteiligten Partnern aus der Schweizerische chemischen Industrie statt. In diesen wurde besprochen, welche Daten bei den Industriepartnern vorhanden sind, wie viel Zeit die Industriepartner investieren können, um Daten zu sammeln und zusammen zu stellen und wie mit Fragen der Geheimhaltung umgegangen werden sollte. Zwei der Industriepartner haben nach oder während der Treffen bereits Produktionsdaten zur Verfügung gestellt. Die drei anderen Partner haben einen Mitarbeiter abgestellt, der zur Unterstützung des Projektes geeignete Daten sammelt und zusammenstellt. Diese werden dem Projekt dann zur Verfügung gestellt werden.

Internationale Zusammenarbeit

Das Deutsche Unternehmen BASF AG hat sich nach einem Treffen im Oktober ebenfalls entschlossen, das Projekt mit Daten zu unterstützen. Herr Wernet wird zu Beginn des neuen Jahres mit Unterstützung eines BASF-Angestellten vorhandene Produktionsdaten auswerten können. Über eine zusätzliche, finanzielle Beteiligung der BASF wird noch verhandelt.

Bewertung 2006 und Ausblick 2007

Die Hauptziele für die zweite Jahreshälfte 2006 konnten beide erfüllt werden. Die Erfolge in der Analyse der Strukturparameter zeigen, dass der verfolgte Ansatz äusserst viel versprechend ist. Für das nächste Jahr sind Präsentationen sowie mehrere Publikationen mit den Resultaten geplant. Darüber

hinaus konnte durch die Korrelationen gezeigt werden, dass eine der Schlüsselannahmen des Projektes korrekt ist.

Die Datengewinnung und die Unterstützung der Industrie sind eines der kritischsten Elemente des Projektes. Ohne die Datenbasis aus der Industrie können die Modelle nicht geplant und entwickelt werden. Die besondere Zusammenarbeit mit der chemischen Industrie ist ein extrem wichtiger Teil des Projektes, und ohne die Unterstützung der Industrie wären weitere Arbeiten durch mangelnde Datenbasis extrem eingeschränkt. Daher ist die zugesicherte und zum Teil schon gelieferte Unterstützung mit Produktionsdaten ein Meilenstein des Projektes.

Für das Jahr 2007 sind folgende Ziele gesetzt worden:

- Beschaffung und Aufarbeitung von weiteren Produktionsdaten aus der chemischen Industrie. Da die Daten zur Erstellung der Modelle möglichst homogen sein sollten, ist hier mit einigem Aufwand zu rechnen.
- Weiterentwicklung der Modelle auf der Basis neuronaler Netze. Mit zusätzlichen Daten lassen sich die Resultate verbessern und die Vorhersagekapazitäten der Modelle erweitern. Das Modell wird an Vergleichsdaten oder in einer Fallstudie validiert.
- Erweiterung eines bestehenden Modells für die Inventaranalyse verschiedener chemischer Reaktionen. Durch die statistische Analyse von Produktionsdaten können Standardwerte für die Durchführung verschiedener Reaktionen bestimmt werden. So lässt sich zum Beispiel der Energiebedarf für die Produktion einer Chemikalie auf einzelne Prozesse aufschlüsseln.