



## **MASTERARBEIT**

### „Vorhersage von Gerüchen (PrOdour) - Datenexploration“

## **SLAVICA SUBIC**

Elitestudiengang

“Advanced Signal Processing and Communications Engineering”

Universität Erlangen-Nürnberg, Januar 2020

## Können Gerüche vorhergesagt werden?

Slavica Subic studiert im Elitestudiengang „Advanced Signal Processing and Communications Engineering“ (ASC) an der Friedrich-Alexander-Universität in Erlangen. Das Masterprogramm legt den Fokus auf aktuellsten Forschungsthemen der Kommunikationswissenschaft. Das hier beschriebene Projekt untersucht die Möglichkeiten, wie Gerüche mit Hilfe von Datenexploration vorhergesagt werden können.

## Was veranlasst eine Substanz, einen bestimmten Duft zu produzieren?

Man könnte diese Frage auf verschiedene Weise beantworten, und eine Möglichkeit ist, die Strukturen zu betrachten, mit der Moleküle dargestellt werden. Strukturformeln sind die Grundlage für eine Datenexploration. Sie sind wie eine ausgefeilte Sprache, um zu erklären, was unserer Meinung nach in der Materie während einer chemischen Reaktion geschieht. Wenn wir Strukturformeln der Chemie wie einzelne Wörter in einem Wörterbuch betrachten, muss nach denjenigen gesucht werden, die für bestimmte Gerüche am häufigsten verantwortlich sind.

Ausgehend von den Molekülformeln, deren Zerlegung in Millionen von Wörtern (Merkmale, submolekulare Strukturen) und der Suche nach den größten überlappenden Strukturen, erstellen wir ein „Wörterbuch der molekularen Merkmale“. Mit den Daten, die aus Experimenten gesammelt wurden, in denen Experten die sensorischen Eigenschaften von Substanzen bewerteten, konnte ein Datensatz erstellt werden, der die Anwendbarkeit einer Liste von Geruchsdeskriptoren auf die im Experiment verwendeten Substanzen enthält. Mit Hilfe der LASSO-Regressionsregulation (LASSO = Least Absolute Shrinkage and Selection Operator) für die Merkmalsextraktion sollen die Merkmale gefunden werden, die für die Anwendbarkeitswerte, nach denen die Deskriptoren im Experiment bewertet wurden, am meisten verantwortlich sind - mit anderen Worten, die Merkmale, die am meisten für die Ausprägung eines bestimmten Geruchs verantwortlich sind. Durch die Gruppierung von Deskriptoren mit Clustering-Techniken kann man Ähnlichkeiten von Merkmalen zwischen den Gruppen untersuchen und nach chemisch bedeutsamen Merkmalen suchen, die ein Deskriptor-Cluster definieren.

Mehr zum Elitestudiengang  
“Advanced Signal Processing and Communications Engineering”:

<https://www.asc.studium.fau.de/>