

“Isn’t it obvious?” Rhetorical Question Identification

Lina Abu-Khader, Philipp Meier, Mehrdad Moslemi, Ines Pisetta
Universität Heidelberg

{labukhader, meier, moslemi, pisetta}@cl.uni-heidelberg.de

1. Einleitung

Das Ziel unseres Projektes ist es, ein möglichst universelles Modell zur Erkennung rhetorischer Fragen zu erstellen. Hierzu trainieren wir zunächst auf transkribierten Telefongesprächen aus dem Switchboard Dialogue Act Corpus und testen auf weiteren Datensätzen. Hierfür verwenden wir Daten aus dem Sarcasm V2 Corpus. Die Erkennung und Klassifizierung von Fragen setzt das Wissen über die verschiedenen Arten von rhetorischen Fragen voraus. Diese werden wir im weiteren Verlauf des Berichts erläutern.

2. Fachlicher Bezug

Um unser Ziel zu erreichen, werden wir uns stark an dem Werk von Bhattasali et al. (2015) [1] orientieren. In der Arbeit wird mit dem Switchboard Dialogue Act Corpus gearbeitet. Die rhetorischen Fragen wurden in zwei Arten unterteilt. Zum einen rhetorische Fragen, die anhand des Kontextes erkannt werden können, wie beispielsweise “Who likes winter? It is always cold and windy” und zum anderen die Fragen, die linguistische Stichwörter enthalten, welche typisch für rhetorische Fragen sind, zum Beispiel “Who ever lifted a finger to help George?”. Mit Hilfe von Bi-, Uni- und POS-n-grammen wurden Features extrahiert. Die Klassifikation erfolgte dann durch den Naive Bayes Algorithmus und durch eine Support Vector Machine. Ein weiteres hilfreiches Werk von Oraby et al. (2013) [2] arbeitet mit dem Sarcasm V2 Korpus, einem Subset des Inter Argument Korpus (IAC), welches sich aus politischen Diskussionen aus Internetforen zusammensetzt. Das Sarcasm v2 Korpus enthält generellen Sarkasmus, Hyperbeln und rhetorische Fragen. Jede Instanz ist entweder als sarkastisch oder nicht-sarkastisch klassifiziert.

3. Korpora

3.1. Switchboard Dialogue Act Corpus

Der Switchboard Dialogue Act Corpus beinhaltet Telefongespräche, welche anhand von Transkriptionsregeln niedergeschrieben wurden. Aus dem Korpus möchten wir

zunächst alle Fragen und auch den Kontext der Fragen extrahieren. Das Korpus enthält geschätzt 500 rhetorische Fragen. Der Aufbau des Korpus wird in Tabelle 1 durch ein Beispiel veranschaulicht. Die Art der Äußerung kann man in dem von Hand annotierten Korpus erkennen, indem man das Tag der Zeile betrachtet. Hierbei werden rhetorische Fragen als “qh” annotiert.

Tag	Sprecher	Aussage	Text
qy	B.48	utt3:	Can you force somebody to be a good productive citizen? /
b	B.50	utt1:	Yeah. /
sv	B.50	utt1:	I don’t think you can. /

Tabelle 1. Switchboard Dialogue Corpus: Beispiel

Die Tags, die bei der Annotation des Korpus verwendet wurden, sind Teil des Discourse Annotation and Markup System of Labeling (=DAMSL Tagset) aus dem Jahre 1997. Für das Projekt sind für uns vor allem die verschiedenen Fragen wichtig, denn wir möchten rhetorische Fragen von allen anderen Fragen unterscheiden können. Die für die Extraktion wichtigen Tags sind in Tabelle 2 zusammengefasst.

Tag	Bedeutung	Beispiel
qh	Rhetorische Frage	Ist das nicht offensichtlich?
qy	Ja-Nein-Frage	Hast du gerade Zeit?
qy^d	Deklarative Ja-Nein-Frage	Kannst du Englisch?
qw	W-Fragen	Wo bist du?
qw^h	Deklarative W-Fragen	You are what kind of buff?
bh	Nachfrage	Ist das richtig?
qo	Offene Fragen	Was ist mit dir?
^g	Tag-Frage	Richtig?

Tabelle 2. Tags für verschiedene Fragetypen

3.2. Inter Annotator Agreement

Um zu überprüfen, wie qualitativ hochwertig die Daten sind, möchten wir auf einen kleinen Teil des Korpus Annotationen per Hand durchführen. Das sogenannte Inter-annotator Agreement dient lediglich zur Überprüfung der Annotationen. Dafür wurden aus dem Switchboard Korpus zufällig 25 rhetorische und 25 nicht rhetorische Fragen extrahiert. Daraus wurden zwei Datensets erstellt:

1. Datenset, welches nur die Fragen beinhaltet
2. Datenset mit Fragen und deren vorherstehendem und dahinterstehendem Kontext

Die Daten wurden dann als rhetorisch und nicht rhetorisch annotiert. Dazu wurden die Tags des Switchboard Korpus entfernt. Auf dem Datenset, das nur die Frage allein enthielt, erzielte die Gruppe einen Kappa-Score von 0.23, auf dem Datenset mit Kontext betrug der Kappa-Score 0.45. Aufgrund dessen liegt die Hypothese nahe, dass der Kontext eine wichtige Rolle bei der Identifikation rhetorischer Fragen spielt. Dies gilt es später beim Erstellen des Modells zu berücksichtigen. Beide Scores sind relativ niedrig. Dies ist sowohl auf die Schwierigkeit der Aufgabe zurückzuführen, als auch darauf, dass die Gespräche teilweise fragwürdig transkribiert und getaggt wurden. Oftmals endet eine Frage mit einem Punkt oder Sätze mit Fragezeichen wurden nicht als Frage getaggt. Ebenso scheinen die vorhandenen Tags (gerade für Fragen) nicht immer korrekt zu sein.

3.3. Sarcasm Corpus

Das Programm soll nicht nur bei der Erkennung von rhetorischen Fragen in Telefongesprächen, sondern auch auf übergreifenden Texten funktionieren. Dazu betrachten wir den Sarcasm V2 Corpus aus dem Werk von Oraby et al. (2013) [2], welcher Daten aus dem Internet Argument Corpus enthält. Auch hier werden rhetorische Fragen mit einer Markierung ("RQ") versehen. Die Daten setzen sich dabei aus einem Text und der zugehörigen Antwort zusammen (Die Antwort ist im Beispiel blau markiert). Ein Beispiel hierfür ist der folgende Satz, wobei die Frage in scharz und die Antwort in blau geschrieben ist:

RQ, sarc, RQ_sarc0001, "So Stalin didn't actually practice Communism (not that anyone has), but really just used bits and pieces to further his rule and such ideology us not communism but Stalinism. So on essence, Stalin wasn't a commie.", *You seem to require a slick answer to a complicated question - so how about this: Question: Was Stalin a communist ? Answer: That's what he said he was. Hope this is glib and vacuous enough to meet with your exacting standards.*

4. Herangehensweise

4.1. Projektablauf

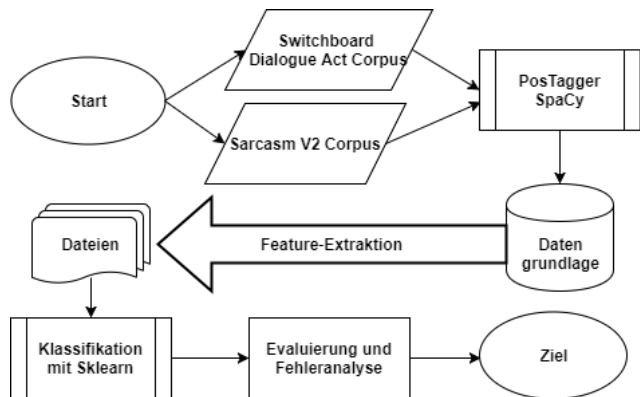


Abbildung 1. Projektablauf

4.2. Klassen rhetorischer Fragen

Rhetorische Fragen haben im Grunde folgende Eigenschaften: Sie sind eine Äußerung in interrogativer Form, dienen nicht dem Informationsfluss und somit nicht dem Informationsgewinn. Meist ist die Antwort auf eine rhetorische Frage offensichtlich und es wird keine ernsthafte Beantwortung benötigt. Es gibt verschiedene Arten von rhetorischen Fragen:

1. Implizit negativ formulierte Fragen, die das Gegenteil meinen: *"Für was soll das denn gut sein?"*
2. Fragen, die quasi schon eine Antwort auf eine andere Frage beantworten: *"Was ist dein Name?" "Seh ich so aus, als ob ich antworten will?"*
3. Fragen, für die es implizite Antworten gibt: *"Who is your daddy?"*
4. Blamierende oder kritisierende Fragen: *"Was denkst du eigentlich, was das ist?"*
5. Der Sprecher drückt seinen Unglauben in der Frage aus: *"Was macht die Fliege in der Suppe?"*

In unserem Projekt möchten wir zwar nicht die genannten Arten der rhetorischen Fragen voneinander unterscheiden können, jedoch lassen sich anhand der verschiedenen Arten möglichst viele rhetorische Fragen erkennen.

Man muss bei der Betrachtung von rhetorischen Fragen zwischen Suggestivfragen differenzieren. Diese können ähnlich formuliert sein wie eine rhetorische Frage, "verlangen" durch ihre Formulierung aber ein bestimmtes Antwortverhalten.

4.3. Datenvorverarbeitung

Im ersten Schritt der Datenvorverarbeitung wird der Switchboard Korpus nach Fragen durchsucht. Im Hinblick auf die rhetorischen Fragen ist es außerdem von Interesse den Kontext, also die vorherstehenden und nachstehenden Sätze, zu extrahieren. In diesem Schritt entstehen unterschiedliche Ausgabedateien, welche entweder die Annotation von “discfluencies“ [6] oder nur reinen Text beinhalten. Die genannten Ausgabedateien sollen zunächst mit POS-Tags versehen werden. Hierzu verwenden wir für die Korpora das Python-Tool “SpaCy“. Die Pos-Tags werden für die Feature-Extraktion benötigt.

4.4. Features

Mögliche Feature-Ideen nach Bhattasali et al. (2015) [1] sind Uni-, Bi-, sowie POS-n-Gramme der Fragen und deren umliegenden Sätzen. Diese sind hilfreich bei der Einbeziehung des Kontextes. Bei der Betrachtung der Fragen planen wir zwischen folgenden Fällen unterscheiden:

1. Die Frage ohne Kontext
2. Die Frage und die davorstehende Äußerung
3. Die Frage und die dahinterstehende Äußerung
4. Die Frage und beide Kontexte

Zum einen werden wir fragenbasierte und zum anderen kontextbasierte Features extrahieren. Um rhetorische Fragen erkennen zu können ist wie das Inter-annotator Agreement beweist, der Kontext einer Frage von großer Bedeutung. Viele Indizien im Kontext lassen auf eine rhetorische Frage schließen.

Der jeweilige Sprecher ist ein wichtiges Feature, da rhetorische Fragen häufig auftreten, wenn der Sprecher A, der die rhetorische Frage stellt, nach der gestellten Frage weiter spricht. Im Gegenzug erwartet man bei einer klassischen Frage eine Antwort von einem anderen Sprecher.

Auch können “strong negative polarity items“ (NPIs), wie “give a damn“ oder “lift a finger“, ein deutlicher Hinweis auf eine rhetorische Frage sein. Fragen, in denen Modalverben zusammen mit einer Negation auftreten tendieren ebenfalls dazu, rhetorisch zu sein.

Es gibt einige Schlüsselwörter, welche bei rhetorischen Fragen häufig verwendet werden. Ein Beispiel für diese Wörter sind “yet“ und “after all“.

Diese Besonderheiten möchten wir bei der Feature-Extraktion beachten.

4.5. Zeitplan

Der momentane Zeitplan der Gruppe ist in Abbildung 2 veranschaulicht. Im Dezember sollen die vorgestellten Features extrahiert werden und anschließend werden Klassifizierungen in Sklearn durchgeführt werden, so dass die Evaluierung rechtzeitig stattfinden kann.

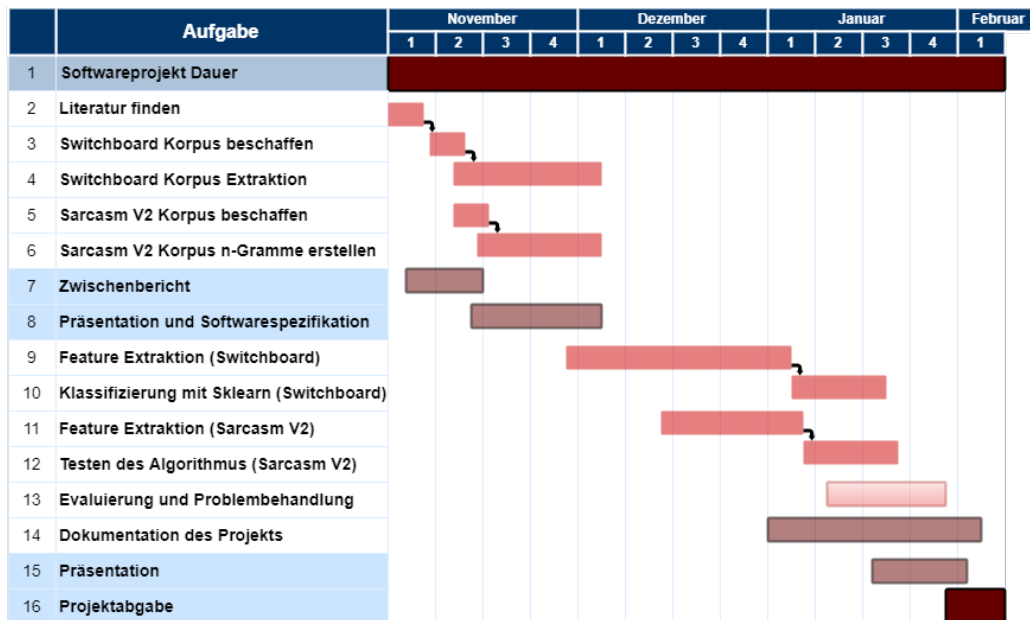


Abbildung 2. Zeitplan

4.6. Klassifikation

Für die Klassifikation wenden wir die Implementierung aus dem Tool Sklearn auf unsere gewonnenen Daten an. Als Klassifikationsalgorithmen bevorzugen wir Naive Bayes und eine Support Vector Machine, da diese besonders für kleine Datensätze geeignet sind. Die Klassifizierer werden wir mithilfe von Cross-Validation trainieren. Bei all dem halten wir uns wieder nah am Werk von Bhattasali et al. (2015) [1].

4.7. Evaluationsmethoden

Als Evaluationsmaße dienen uns die klassischen Maße Accuracy, Precision, Recall und F1-Measure. Für die Baseline orientieren wir uns am Modell von Bhattasali et al. (2015) [1]. Da wir unterschiedliche Featuresets erstellen, dient uns das Set bestehend aus den Fragen allein als Baseline.

Zu beachten gilt bei der Evaluation das Auftreten von möglichen "false positives" bei unterschiedlichen Datensets. Diese werden wir in einer ausführlichen Fehleranalyse betrachten, um das Auftreten weitestgehend zu verringern.

4.8. Optionale Ziele

Falls im Projekt noch genügend Zeit verbleibt, können Strong Negative Polarity Items explizit als eigenes Feature implementiert werden. Dazu stehen zwei Dateien mit NPIs zur Verfügung. Interessant wäre außerdem eine Experimentreihe um festzustellen, welche NPIs besonders gut auf den Datensets funktionieren.

Der zweite optionale Punkt ist die Anwendung des Algorithmus auf weiteren Korpora, beispielsweise auf Twitter-Daten.

Literatur

- [1] BHATTASALI, Shohini; CYTRYN, Jeremy; FELDMAN, Elana & PARK, Joonsuk, "Automatic Identification of Rhetorical Questions", Cornell University, 743-749, 2015.
- [2] ORABY, Shereen; HARRISON, Vrindavan; REED, Lena & HERNANDEZ, Ernesto; RILOFF, Ellen; WALKER, Marilyn; "Creating and Characterizing a Diverse Corpus of Sarcasm in Dialogue", University of California, Santa Cruz, 31-41, 2013.
- [3] RANGANATH, Suhas; HU, Xia; TANG, Juliang; WANG, Suhang; LIU, Huan; "Identifying Rhetorical Questions in Social Media", Arizona State University, 667-670, 2017.
- [4] BAMMAN, David; SMITH, Noah A.; "Contextualized Sarcasm detection on Twitter", Carnegie Mellon University, 574-577, 2015.
- [5] LU, Yujie; SAKAMOTO, Kotaro; SHIBUKI, Hideyuki; MORI, Tatsunori; "Construction of a Multilingual Annotated Corpus for Deeper Sentiment Understanding in Social Media", Carnegie Mellon University, 205-265, 2017.
- [6] METEER, Marie; TAYLOR, Ann A.; "Dysuency Annotation Stylebook for the Switchboard Corpus", University of Pennsylvania, 1995.