

Questions of Interest

1. Why are EV sales low?
2. What do owners think about their cars?
3. What are manufacturers doing wrong (and right)?

Challenging To Answer...

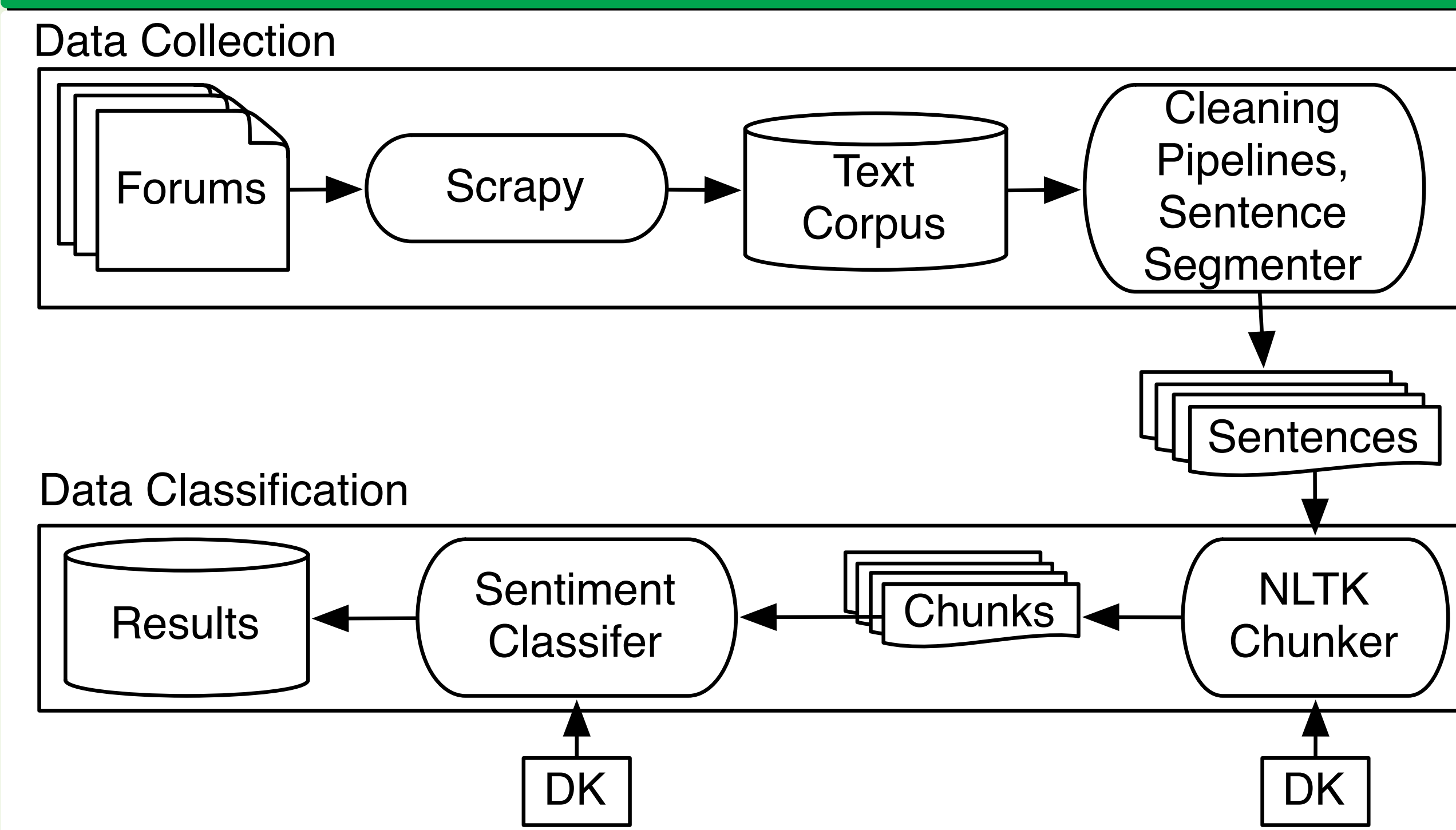
- ▶ Field trials are expensive
- ▶ Drivers to survey are hard to find
- ▶ No easily-consumable review database (e.g., Amazon)
- ▶ Reading text is time consuming/tedious (compared to ↑)

Our Solution

Automatically mine & summarize EV ownership forums based on the **features** the user cares about:

- ▶ Get input features from user, e.g., range, battery-life,...
- ▶ Crawl forums to build a **review corpus**
- ▶ Mine **feature-opinion pairs**, or (f, o) s
- ▶ **Classify** (f, o) s as $+$, $-$, N
- ▶ Produce **data visualizations**

Architecture



Python Natural Language Toolkit Chunking

- ▶ **Tag** sentences for pos, producing tuples: $\dots, (w_n, pos_n), \dots$
- ▶ Define a NLTK [1] **chunking grammar**, a series of **regex rules** executed on tuple sequences to find **chunks** (combos)
- ▶ Rules executed in-order and non-overlappingly
- ▶ $*$, $.$, $+$, $?$ allow for arbitrarily long groups & optional pos:

$$\{<det>?<noun><verb>+<adverb>*<adjective>+\}$$
 chunks “the (product-name) is really superb” and “my (product-name) has been reliable”
- ▶ DK used to find chunks containing features

Handling Context-Dependent Opinions

1. For some **oriented features**, more or less is always better, e.g., “performance” and “price”
Intensity modifiers change sentiment w/ oriented features:

(Range, low) $\rightarrow -$
 (Maintenance, low) $\rightarrow +$
 (Gear, low/high) $\rightarrow N$

2. We build (laboriously using DK) a small **prod-feat sentiment dictionary**, $S_{f \in \mathcal{F}_p}^{p \in \mathcal{P}}$, for each {product, feature} pair:

$S[\text{Leaf}, \text{Quality}](\text{cheap}) \rightarrow -$
 $S[\text{Leaf}, \text{Price}](\text{cheap}) \rightarrow +$
 $S[\text{Leaf}, \text{Performance}](\text{awesome}) \rightarrow \text{unknown}$

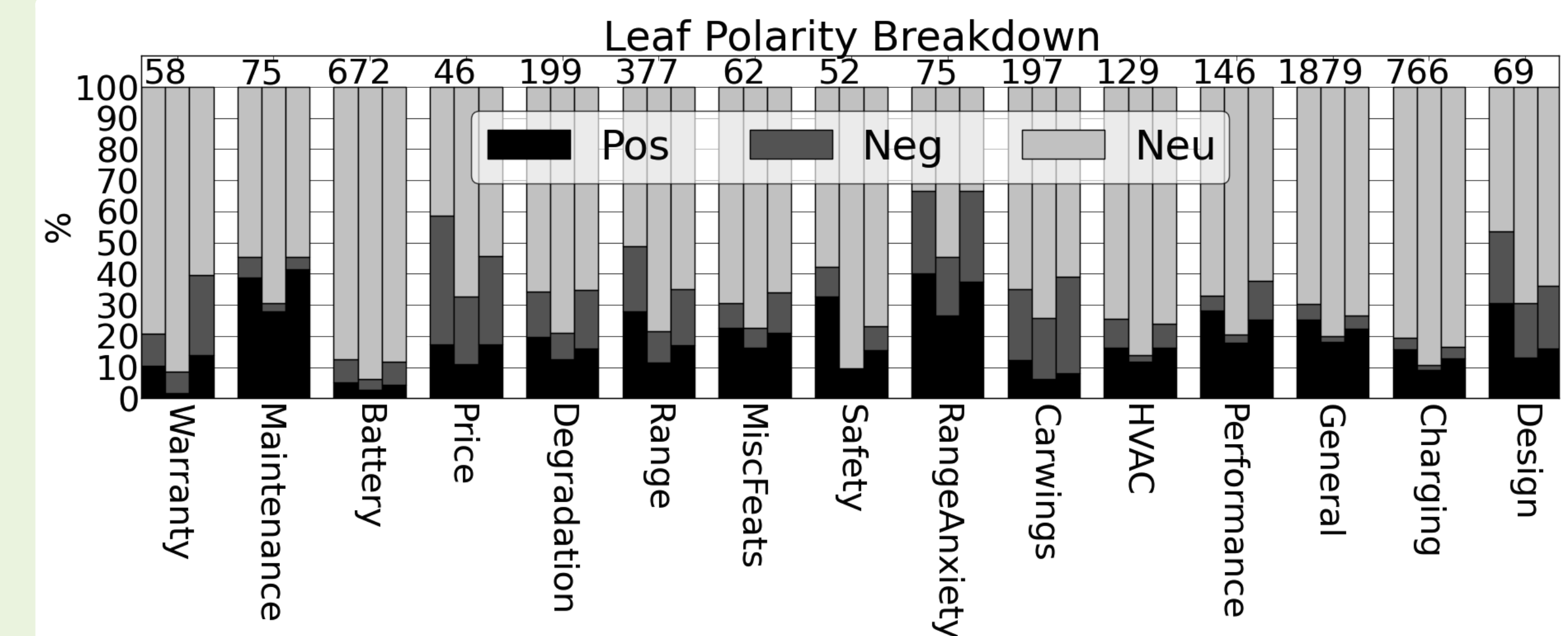
Sentiment Querying (f, o) s

1. If o is an intensity modifier & f “oriented”:
 $(o : +, f : +) \rightarrow \text{return } +$, $(o : +, f : -) \rightarrow \text{return } -$
 $(o : -, f : +) \rightarrow \text{return } -$, $(o : -, f : -) \rightarrow \text{return } +$
2. Querying the product-feature sentiment dictionary S_f^p
3. Query the **default sentiment dictionary** [2]
4. Return N

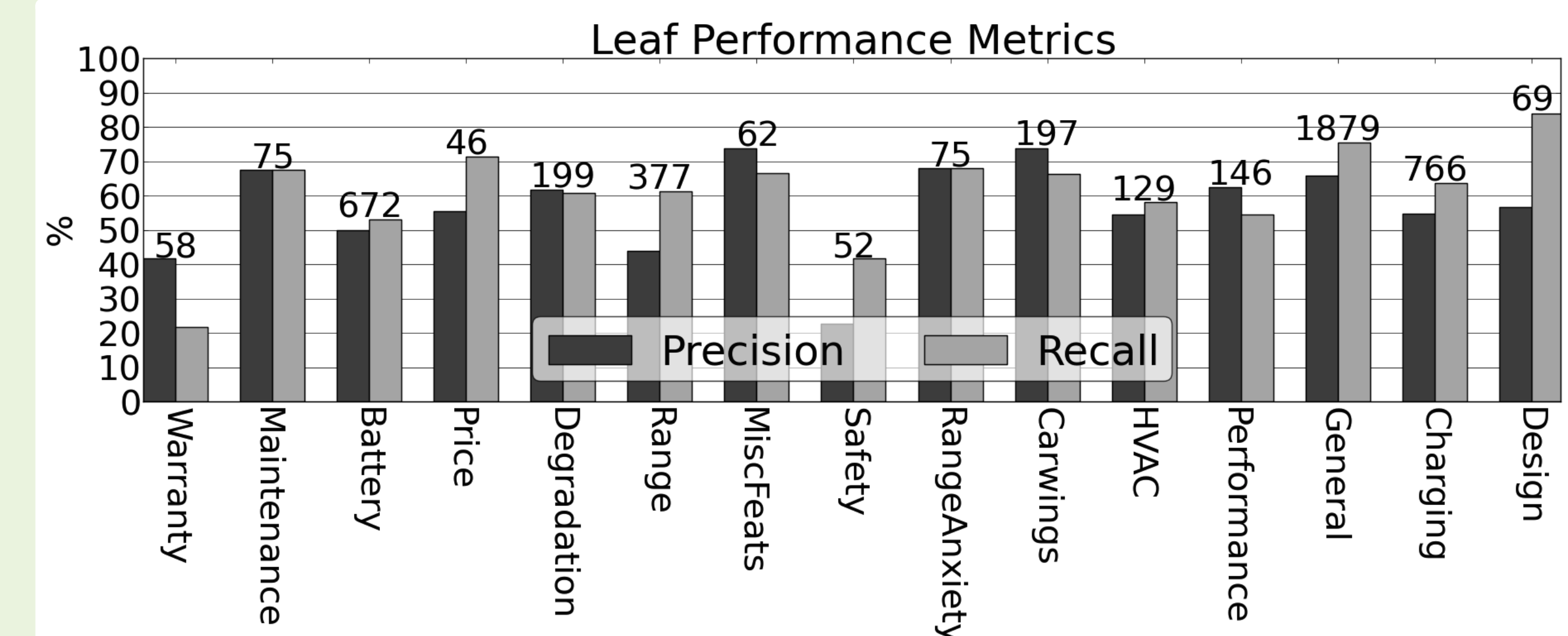
Eval Corpus

8,000 manually labelled sentences w/ product features

Data Visualization



Current Performance



References

1. <http://nltk.org/>
2. Theresa Ann Wilson, Fine-grained Subjectivity and Sentiment Analysis, PhD Thesis, 2008, University of Pittsburgh
3. Kunpeng Zhang et al., Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking. WOSN, 2010.