

Opinion Word Expansion and Target Extraction through Double Propagation

Guang Qiu, Bing Liu, Jiajun Bu, Chun Chen

Extraction



Opinion Lexicon Expansion

- Opinion Words :
 - adjectives (good, bad, ugly)
- Lexicon Expansion :
 - Adding more words to an existing dictionary of opinion words
- Semi-supervised

Opinion Target Extraction

- Targets
 - Nouns/Noun phrases on which opinions are expressed
 - car, lens, battery life

Double Propagation

- Using an increase in either the opinion lexicon or target lexicon to have a ripple effect on the other.

RuleID	Observations	output
$R1_1$	$O \rightarrow O-Dep \rightarrow T$ s.t. $O \in \{O\}, O-Dep \in \{MR\}, POS(T) \in \{NN\}$	$t = T$
$R1_2$	$O \rightarrow O-Dep \rightarrow H \leftarrow T-Dep \leftarrow T$ s.t. $O \in \{O\}, O/T-Dep \in \{MR\}, POS(T) \in \{NN\}$	$t = T$
$R2_1$	$O \rightarrow O-Dep \rightarrow T$ s.t. $T \in \{T\}, O-Dep \in \{MR\}, POS(O) \in \{JJ\}$	$o = O$
$R2_2$	$O \rightarrow O-Dep \rightarrow H \leftarrow T-Dep \leftarrow T$ s.t. $T \in \{T\}, O/T-Dep \in \{MR\}, POS(O) \in \{JJ\}$	$o = O$
$R3_1$	$T_{i(j)} \rightarrow T_{i(j)}-Dep \rightarrow T_{j(i)}$ s.t. $T_{j(i)} \in \{T\}, T_{i(j)}-Dep \in \{CONJ\}, POS(T_{i(j)}) \in \{NN\}$	$t = T_{i(j)}$
$R3_2$	$T_i \rightarrow T_i-Dep \rightarrow H \leftarrow T_j-Dep \leftarrow T_j$ s.t. $T_i \in \{T\}, T_i-Dep == T_j-Dep, POS(T_j) \in \{NN\}$	$t = T_j$
$R4_1$	$O_{i(j)} \rightarrow O_{i(j)}-Dep \rightarrow O_{j(i)}$ s.t. $O_{j(i)} \in \{O\}, O_{i(j)}-Dep \in \{CONJ\}, POS(O_{i(j)}) \in \{JJ\}$	$o = O_{i(j)}$
$R4_2$	$O_i \rightarrow O_i-Dep \rightarrow H \leftarrow O_j-Dep \leftarrow O_j$ s.t. $O_i \in \{O\}, O_i-Dep == O_j-Dep, POS(O_j) \in \{JJ\}$	$o = O_j$

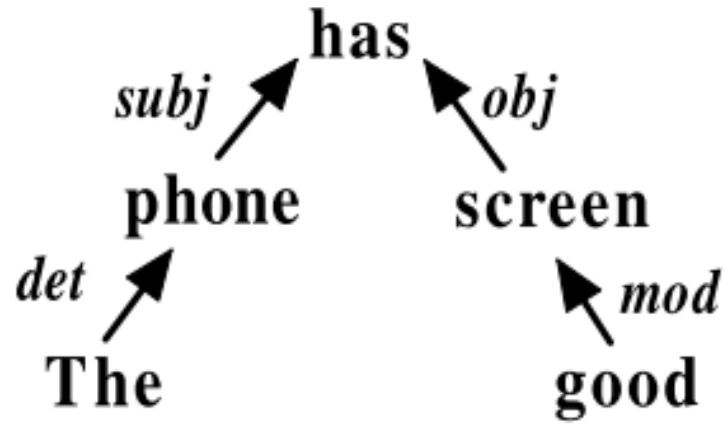
Rules of Extraction

Rule of Extraction (Cont.)

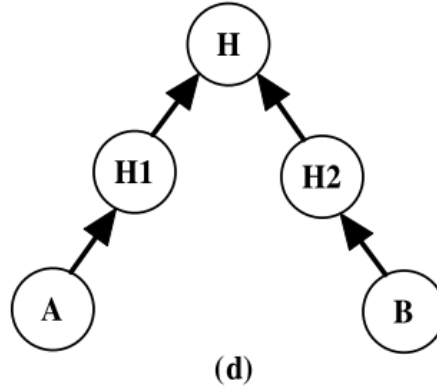
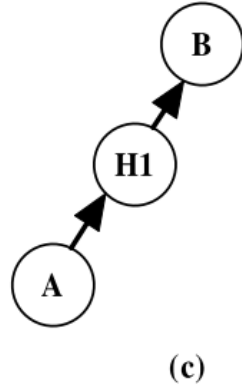
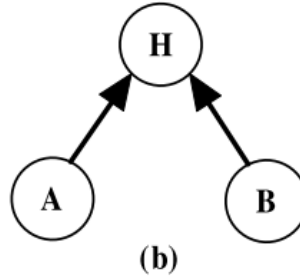
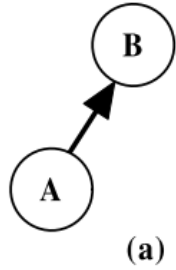
- o (or t) stands for the output (extracted) opinion word (or target).
- {O} (or {T}) is the set of known opinion words (or the set of targets) either given or extracted.
- H means any word.
- POS(O(or T)) and O(or T)-Dep stand for the POS information and dependency relation of the word O (or T) respectively.
- {JJ} and {NN} are sets of POS tags of potential opinion words and targets, respectively.
- {JJ} contains JJ, JJR, and JJS;
- {NN} contains NN and NNS.
- {MR} consists of dependency mod, pnm, subj, s, obj, obj2 and desc
- {CONJ} contains conj only.

Rules of Extraction (Cont.)

- $O \rightarrow O\text{-Dep} \rightarrow T$ means O depends on T through a syntactic relation $O\text{-Dep}$.
- “==” represents the same or equivalent (Here equivalent specifically means mod is the same as pnmod , and s or subj is the same as obj).
- For example, $T_i\text{-Dep} == T_j\text{-Dep}$ means $T_i\text{-Dep}$ being the same as $T_j\text{-Dep}$ or equivalent (e.g., subj and obj in R32).



Word Dependency Graph with POS tagging



Direct and Indirect Dependencies

Polarity Assignment



Polarity Assignment

- Dictionary Based
 - Domain Independent
 - Create a network of words on the basis of word, definition pair
 - consider the word 'unpredictable'
- Corpora Based

Rules

- **Heterogeneous Rule:** Opinions extracted from the same target in the same review have the same polarity
- **Homogeneous Rule :** Opinions extracted from the same opinions have the same polarity in the same review
- **Intra-review Rule :** Across reviews there are no relations between opinions and targets
- Rules are applied barring words like but, however

Rules (Cont.)

- In case of conflict, sum up the total positive and negative polarities
- Every review has an overall polarity

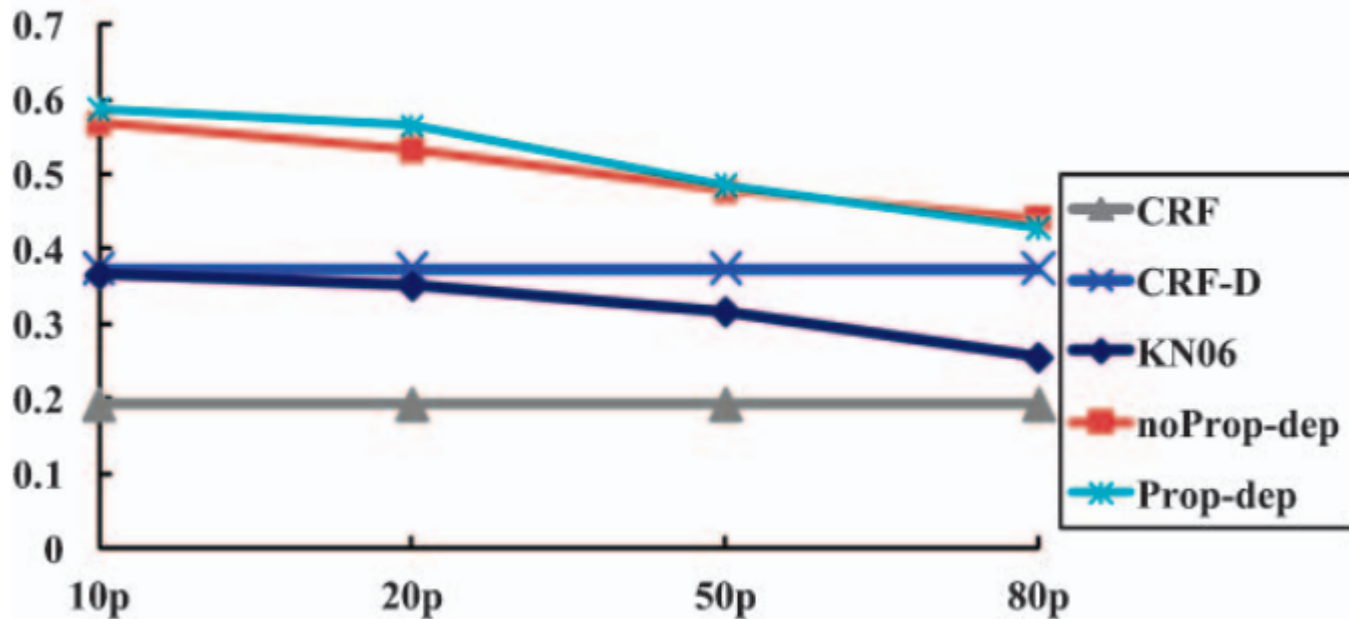
Target Pruning

- Clause Pruning
 - Each clause has only one target, in case of multiple remove one by frequency analysis
- Global Pruning
 - Group words like battery and life together and apply frequency analysis on them together

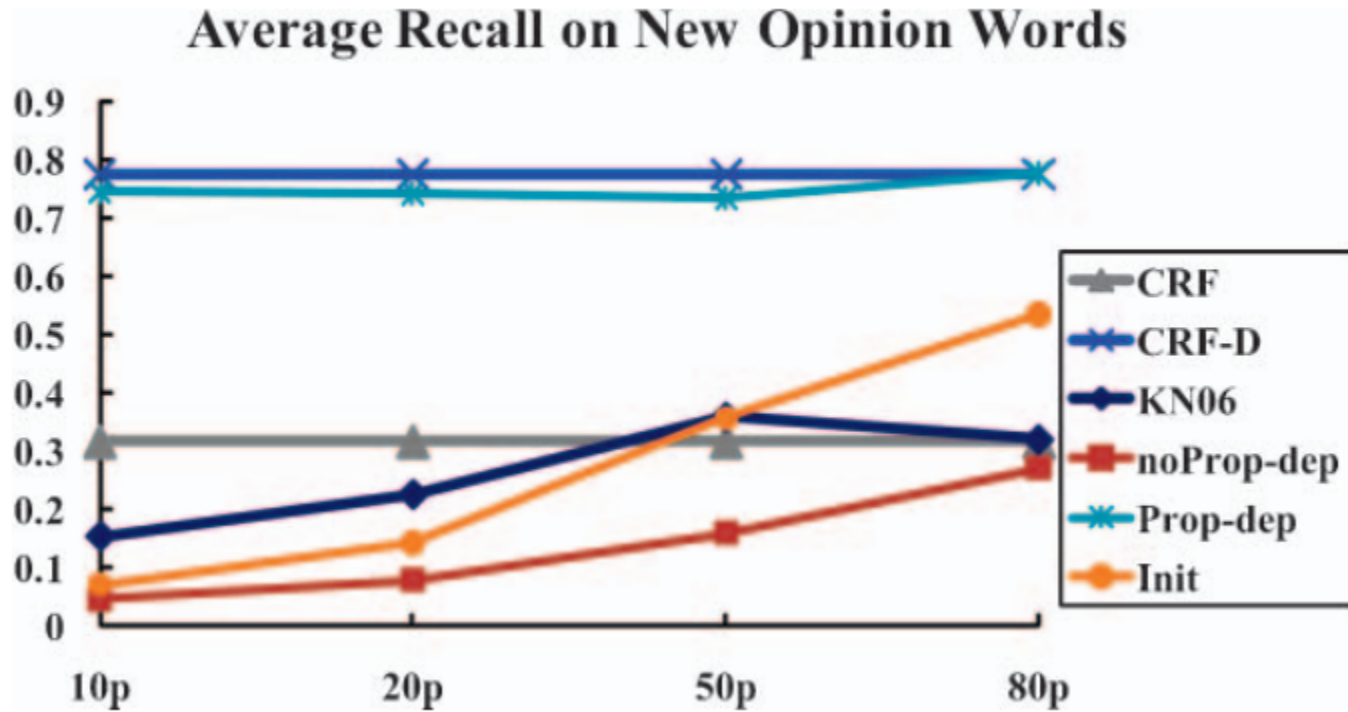
Results



Average Precision on New Opinion Words

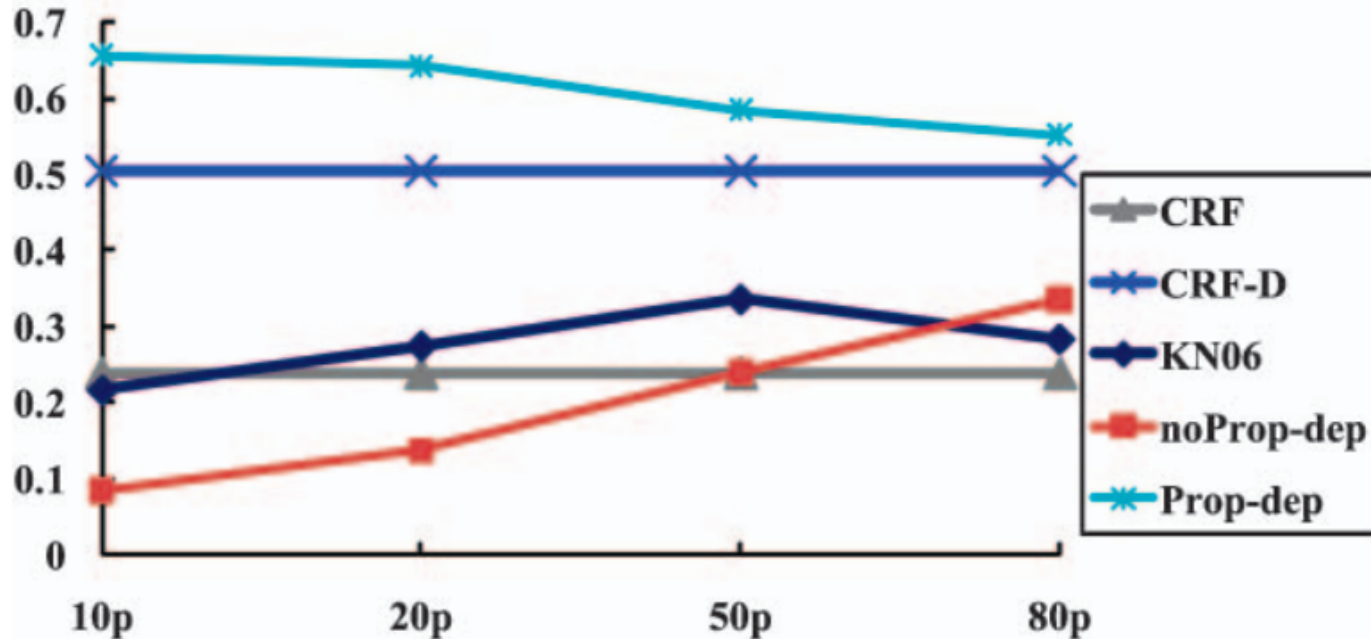


Precision is high, rules are good



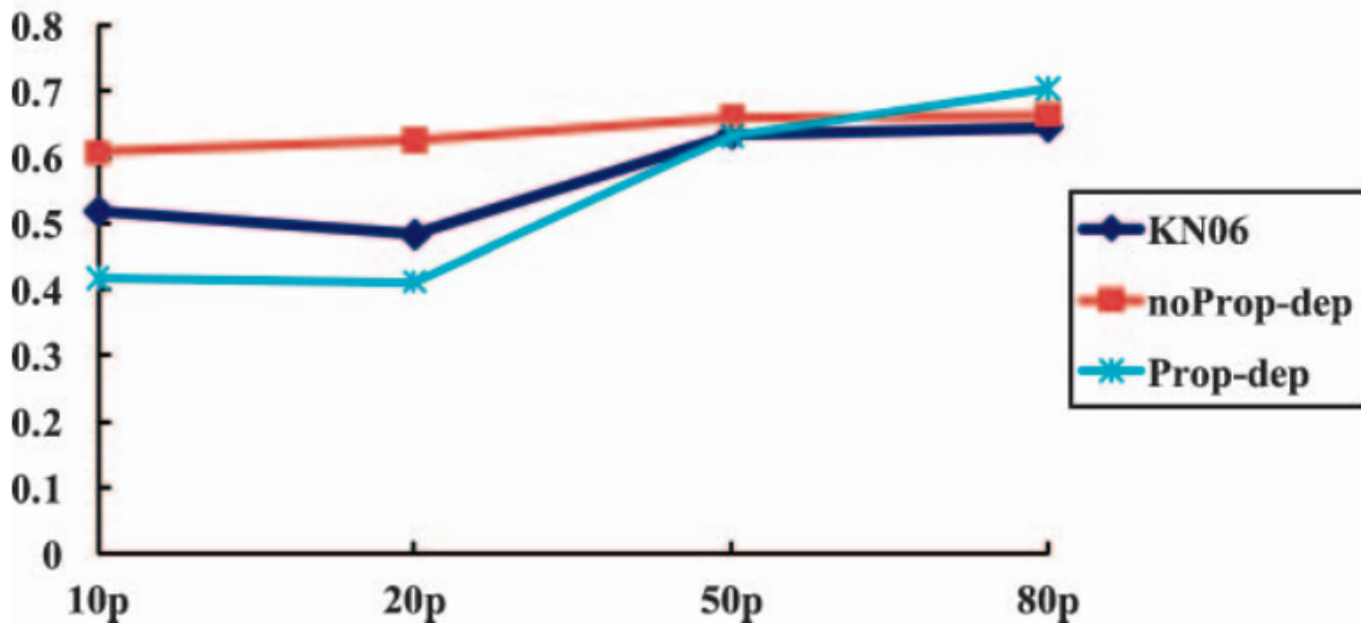
Recall is constant, harder to extract more

Average F-Score on New Opinion Words



Propagation is Effective and method works on small seed

Average Polarity Assignment Accuracy on Correct New Opinion Words



Propagation introduces noise at low seed

Thanks

