Feature Specific Sentiment Analysis of Reviews

Subhabrata Mukherjee and Pushpak Bhattacharyya
Dept. of Computer Science and Engineering,
IIT Bombay

13th International Conference on Intelligent Text Processing and Computational Intelligence - **CICLING 2012**,
New Delhi, India, March, 2012

MOTIVATION CONTD…

- Sentiment Analysis is *always* with respect to a particular entity or feature
- Feature may be *implicit or explicit*
- **This work concerns *explicit feature***

MOTIVATION

- *I have an* ***ipod*** *and it is a great buy but I'm probably the only person that dislikes the iTunes* ***software****.*
  - *Here the sentiment w.r.t ipod is positive whereas that respect to software is negative*

ENTITY AND FEATURES

- An entity may be analyzed from the point of view of multiple features

- Entity – Titanic
- Features – Music, Direction, Plot etc.

- Given a sentence how to identify the set of features ?

SCENARIO

- Each sentence can contain **multiple** features and **mixed** opinions (positive and negative)
- Reviews mixed from various domains
- No prior information about set of features except the ***target feature***

MAIN FEATURES OF THE ALGORITHM

- Does not require any prior information about any domain
- Unsupervised – But need a **small** untagged dataset to tune parameters
- Does not require any prior feature set
- Groups set of features into separate clusters which need to be pruned or labeled

Opinion Extraction Hypothesis

- *“More closely related words come together to express an opinion about a feature”*

Hypothesis Example

- *“I want to use Samsung which is a great product but am not so sure about using Nokia”.*
  - *Here “great” and “product” are related by an adjective modifier relation, “product” and “Samsung” are related by a relative clause modifier relation. Thus “great” and “Samsung” are transitively related.*
  - ***Here “great” and “product” are more related to Samsung than they are to Nokia***
  - *Hence “great” and “product” come together to express an opinion about the entity “Samsung” than about the entity “Nokia”*

Hypothesis Example

- *"I want to use Samsung which is a great product but am not so sure about using Nokia".*
  - *Here "great" and "product" are related by an adjective modifier relation, "product" and "Samsung" are related by a relative clause modifier relation. Thus "great" and "Samsung" are transitively related.*
  - ***Here "great" and "product" are more related to Samsung than they are to Nokia***
  - *Hence "great" and "product" come together to express an opinion about the entity "Samsung" than about the entity "Nokia"*

Hypothesis Example

- *"I want to use Samsung which is a great product but am not so sure about using Nokia".*
  - *Here "great" and "product" are related by an adjective modifier relation, "product" and "Samsung" are related by a relative clause modifier relation. Thus "great" and "Samsung" are transitively related.*
  - ***Here "great" and "product" are more related to Samsung than they are to Nokia***
  - *Hence "great" and "product" come together to express an opinion about the entity "Samsung" than about the entity "Nokia"*

Hypothesis Example

- *"I want to use Samsung which is a great product but am not so sure about using Nokia".*
  - *Here "great" and "product" are related by an adjective modifier relation, "product" and "Samsung" are related by a relative clause modifier relation. Thus "great" and "Samsung" are transitively related.*
  - ***Here "great" and "product" are more related to Samsung than they are to Nokia***
  - *Hence "great" and "product" come together to express an opinion about the entity "Samsung" than about the entity "Nokia"*

Hypothesis Example

- *"I want to use Samsung which is a great product but am not so sure about using Nokia".*

Adjective Modifier

  - *Here "great" and "product" are related by an adjective modifier relation, "product" and "Samsung" are related by a relative clause modifier relation. Thus "great" and "Samsung" are transitively related.*
  - ***Here "great" and "product" are more related to Samsung than they are to Nokia***
  - *Hence "great" and "product" come together to express an opinion about the entity "Samsung" than about the entity "Nokia"*

Hypothesis Example

- *"I want to use Samsung which is a great product but am not so sure about using Nokia".*

  Adjective Modifier

  - *Here "great" and "product" are related by an adjective modifier relation, "product" and "Samsung" are related by a relative clause modifier relation. Thus "great" and "Samsung" are transitively related.*
  - ***Here "great" and "product" are more related to Samsung than they are to Nokia***
  - *Hence "great" and "product" come together to express an opinion about the entity "Samsung" than about the entity "Nokia"*

Hypothesis Example

- *"I want to use Samsung which is a great product but am not so sure about using Nokia".*

Adjective Modifier

  - *Here "great" and "product" are related by an adjective modifier relation, "product" and "Samsung" are related by a relative clause modifier relation. Thus "great" and "Samsung" are transitively related.*
  - ***Here "great" and "product" are more related to Samsung than they are to Nokia***
  - *Hence "great" and "product" come together to express an opinion about the entity "Samsung" than about the entity "Nokia"*

Hypothesis Example

- *“I want to use Samsung which is a great product but am not so sure about using Nokia”.*

Relative Clause Modifier

Adjective Modifier

  - *Here “great” and “product” are related by an adjective modifier relation, “product” and “Samsung” are related by a relative clause modifier relation. Thus “great” and “Samsung” are transitively related.*
  - ***Here “great” and “product” are more related to Samsung than they are to Nokia***
  - *Hence “great” and “product” come together to express an opinion about the entity “Samsung” than about the entity “Nokia”*

Example of a Review

- *I have an **ipod** and it is a great buy but I'm probably the only person that dislikes the iTunes **software**.*

Example of a Review

- *I have an* ***ipod*** *and it is a great buy but I'm probably the only person that dislikes the iTunes* ***software****.*

Example of a Review

- *I have an* ***ipod*** *and it is a great buy but I'm probably the only person that dislikes the iTunes* ***software****.*

Example of a Review

- *I have an **ipod** and it is a great buy but I'm probably the only person that dislikes the iTunes **software**.*

Example of a Review

- *I have an **ipod** and it is a great buy but I'm probably the only person that dislikes the iTunes **software**.*

Feature Extraction : Domain Info Not Available

Feature Extraction : Domain Info Not Available

- Initially, all the Nouns are treated as features and added to the *feature list* ***F***.

Feature Extraction : Domain Info Not Available

- Initially, all the Nouns are treated as features and added to the *feature list* ***F***.

- ***F*** = { *ipod, buy, person, software* }

Feature Extraction : Domain Info Not Available

- Initially, all the Nouns are treated as features and added to the *feature list* ***F***.

- ***F*** = { *ipod, buy, person, software* }

- Pruning the feature set
  - Merge 2 features if they are **strongly related**

Feature Extraction : Domain Info Not Available

- Initially, all the Nouns are treated as features and added to the *feature list* ***F***.

- ***F*** = { *ipod, buy, person, software* }

- Pruning the feature set
  - Merge 2 features if they are **strongly related**

- *"buy"* merged with "*ipod*", when target feature = "*ipod*",
  - *"person, software"* will be ignored.

Feature Extraction : Domain Info Not Available

- Initially, all the Nouns are treated as features and added to the *feature list* ***F***.
- ***F*** = *{ ipod, buy, person, software }*
- Pruning the feature set
  - Merge 2 features if they are **strongly related**
- *"buy"* merged with "*ipod"*, when target feature = "*ipod"*,
  - *"person, software"* will be ignored.
- *"person"* merged with "*software"*, *when* target feature = "*software"*
  - *"ipod, buy"* will be ignored.

Relations

- Direct Neighbor Relation
  - Capture **short range dependencies**
  - Any 2 consecutive words (such that none of them is a StopWord) are directly related
  - Consider a sentence *S* and 2 consecutive words .
  - *If* $w_i, w_{i+1} \notin Stopwords$, then they are directly related. $w_i, w_{i+1} \in S$

- Dependency Relation
  - Capture **long range dependencies**
- Let *Dependency_Relation* be the list of **significant relations**.
- Any 2 words w_i and w_j in *S* are directly related, *if*

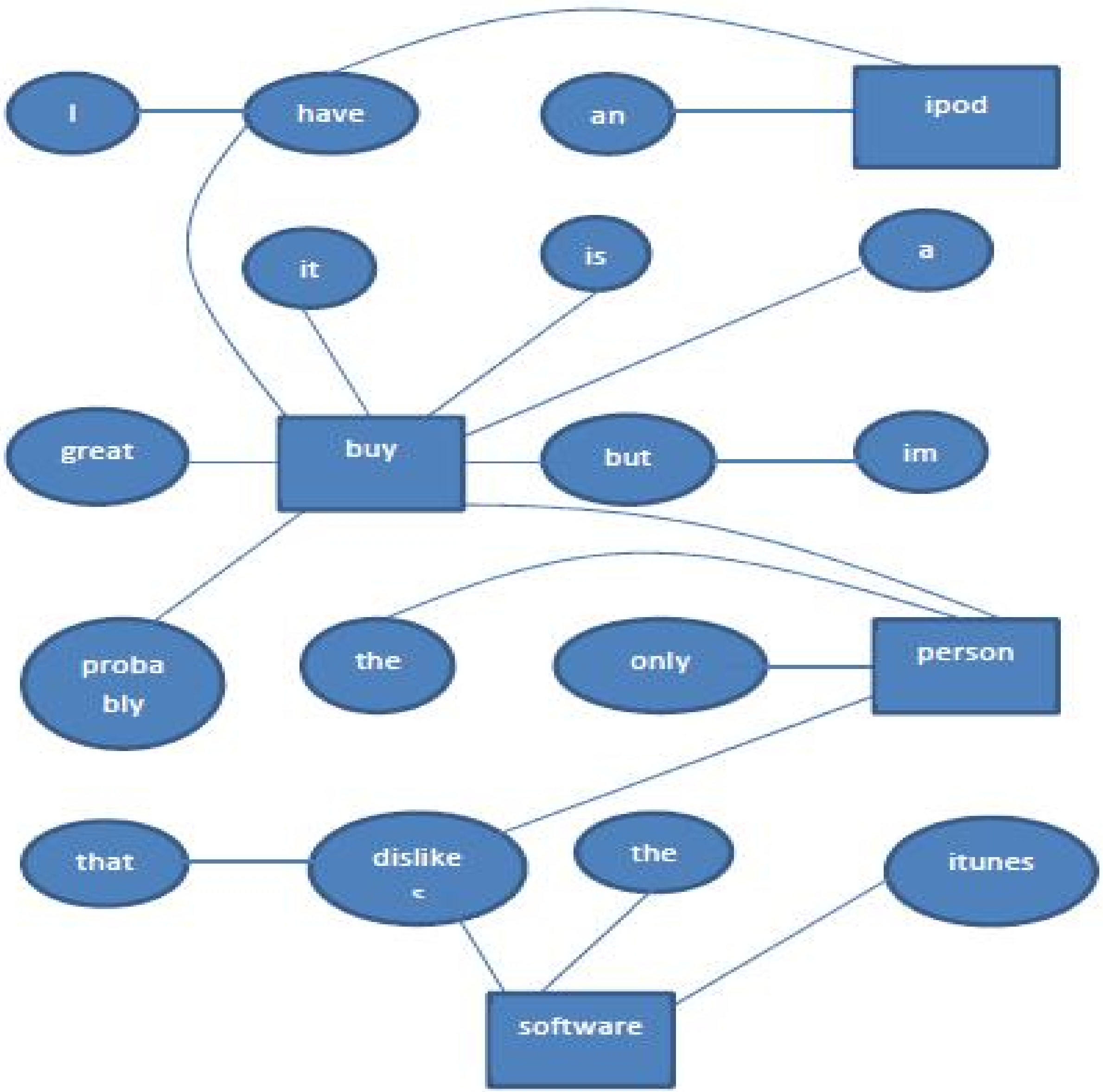
$\exists D_i$ s.t. $D_i(w_i, w_j) \in Dependency\_Relation$

Graph representation

Given a sentence S, let $\boldsymbol{W}$ be the set of all words in the sentence S.

A Graph $G(W, E)$ is constructed such that any $w_i, w_j \in W$ are directly connected by $e_k \in E$,*if* $\exists R_l \;\; s.t. \;\; R_l(w_i, w_j) \in R$.

Graph



I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
s
the
itunes
software

Algorithm

i. Initialize n clusters $C_i \ \forall i = 1..n$

ii. Make each $f_i \in F$ the clusterhead of C_i . The target feature f_t is the clusterhead of C_t. Initially, each cluster consists only of the clusterhead.

Algorithm Contd…

iii. Assign each word $w_j \in S$ to cluster C_k

$$s.t.\ \ k = \arg\min_{i \in n} dist(w_j, f_i),$$

Where $dist(w_j, f_i)$ *gives the number of edges, in the shortest path, connecting* w_j *and* f_i *in G.*

Algorithm Contd…

iv. Merge any cluster C_i with C_t *if*,

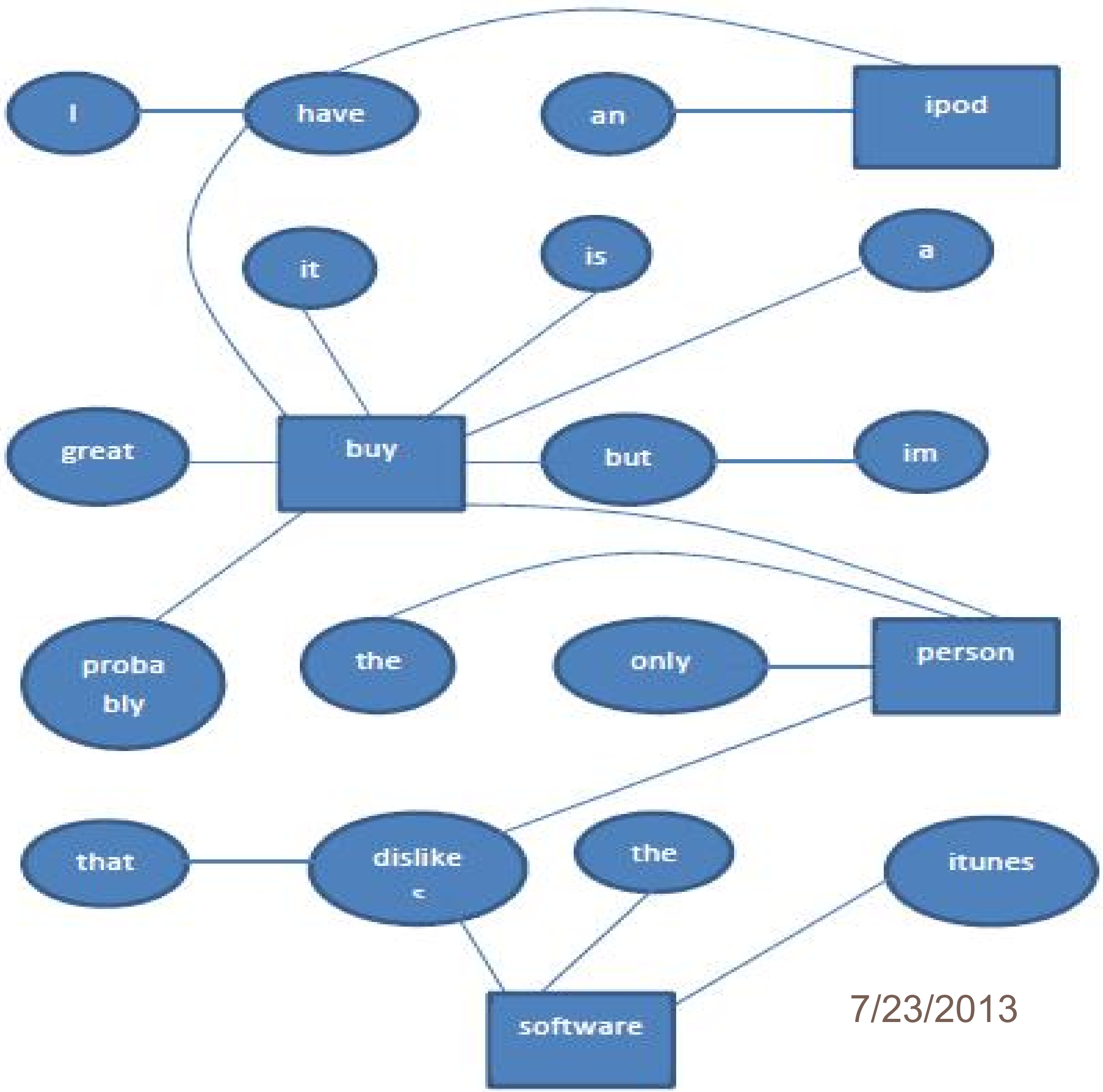
$$dist(f_i, f_t) < \theta ,$$

Where θ is some threshold distance.

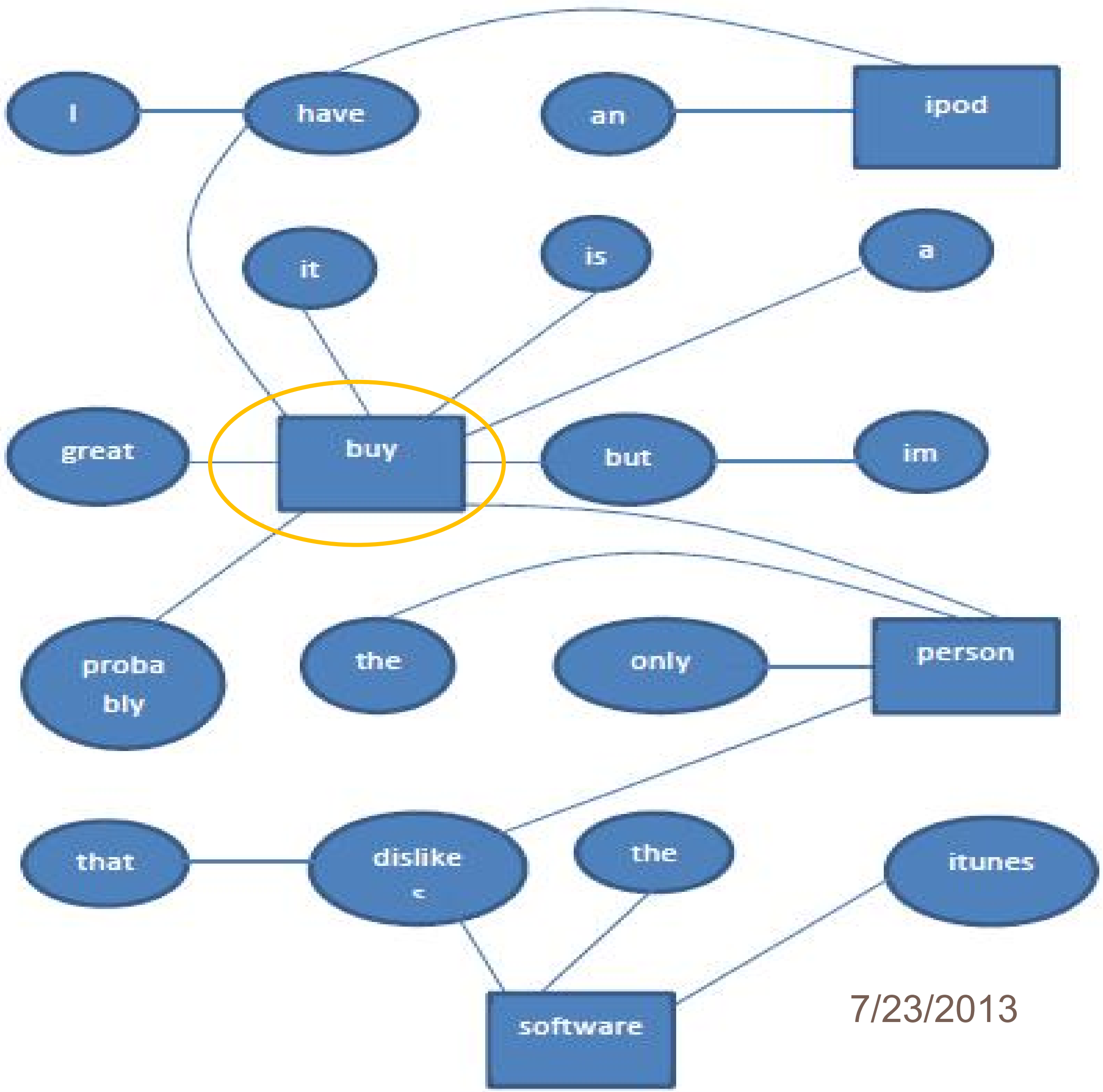
v. Finally the set of words $w_i \in C_t$ gives the opinion expression regarding the target feature f_t .

Clustering

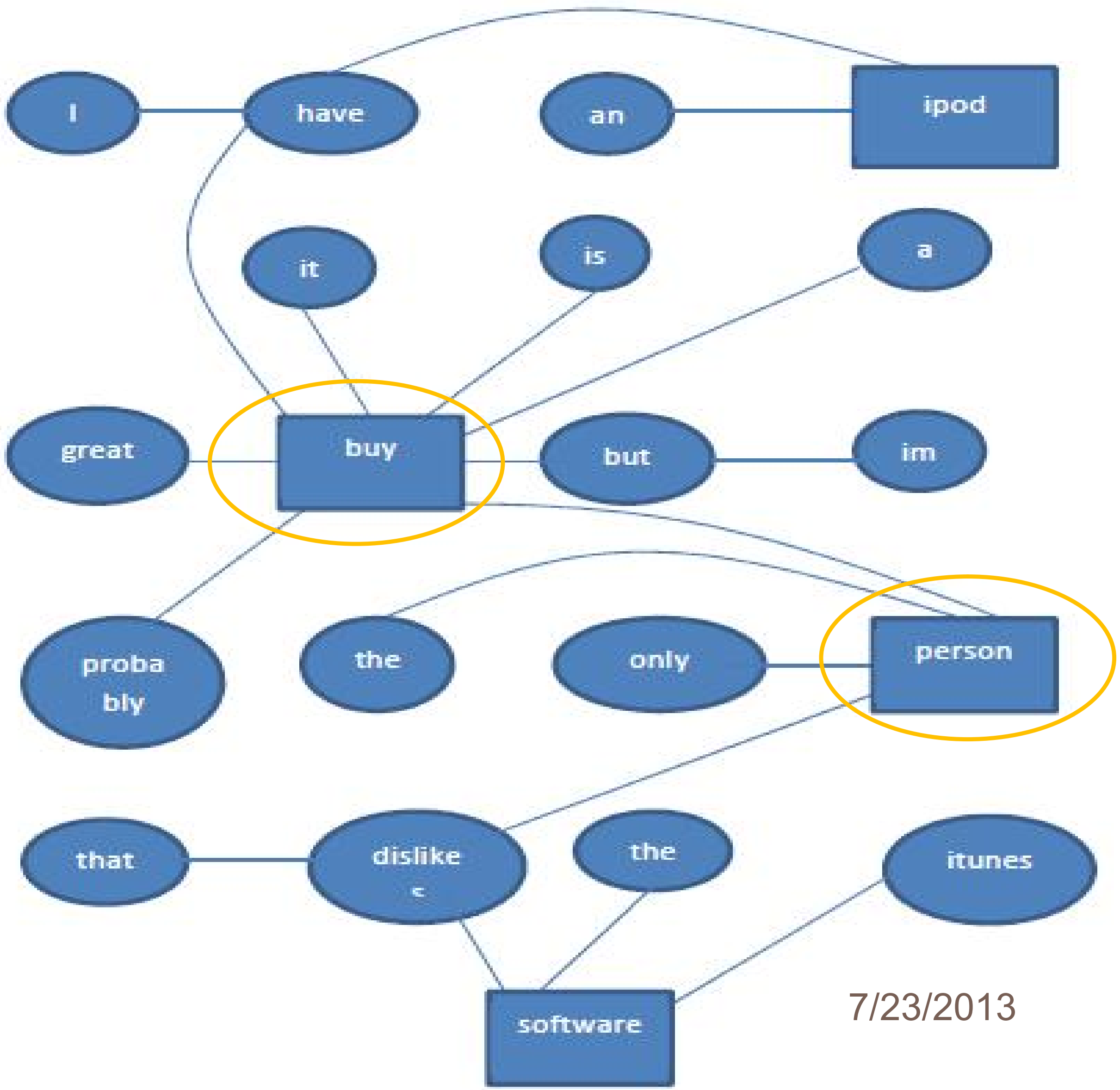
Clustering


I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
s
the
itunes
software

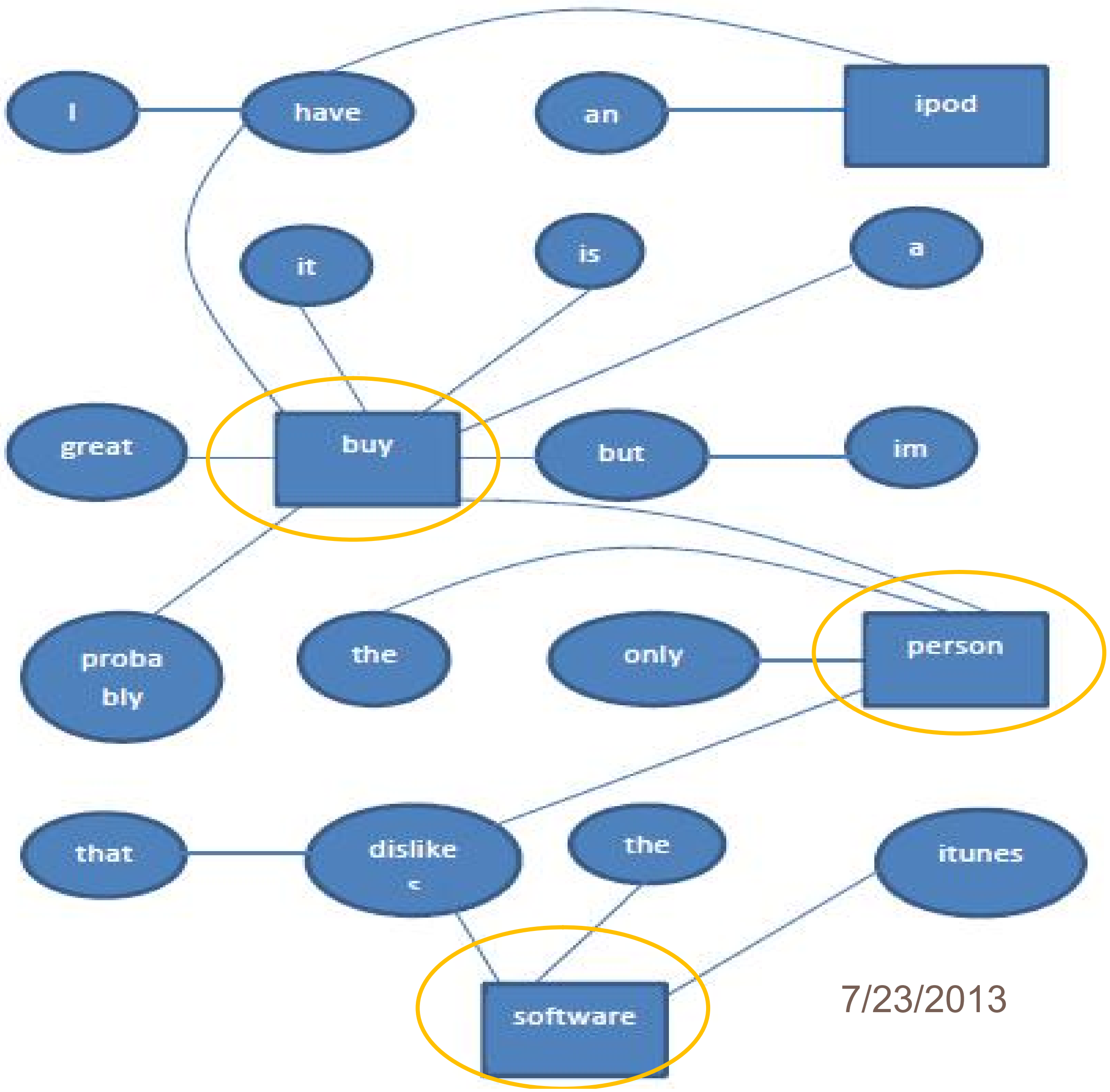
Clustering


I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
s
the
itunes
software

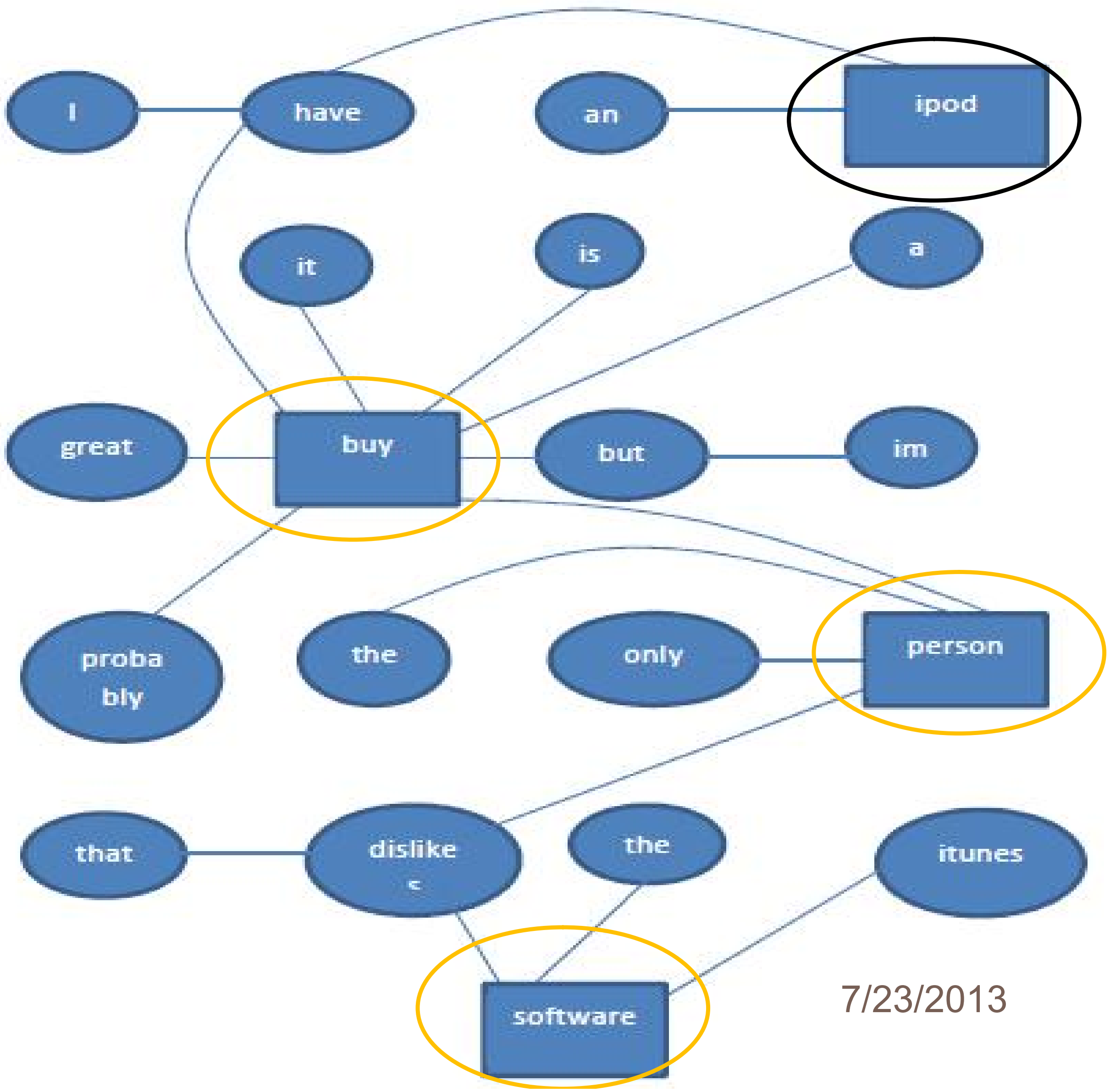
Clustering


I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
s
the
itunes
software

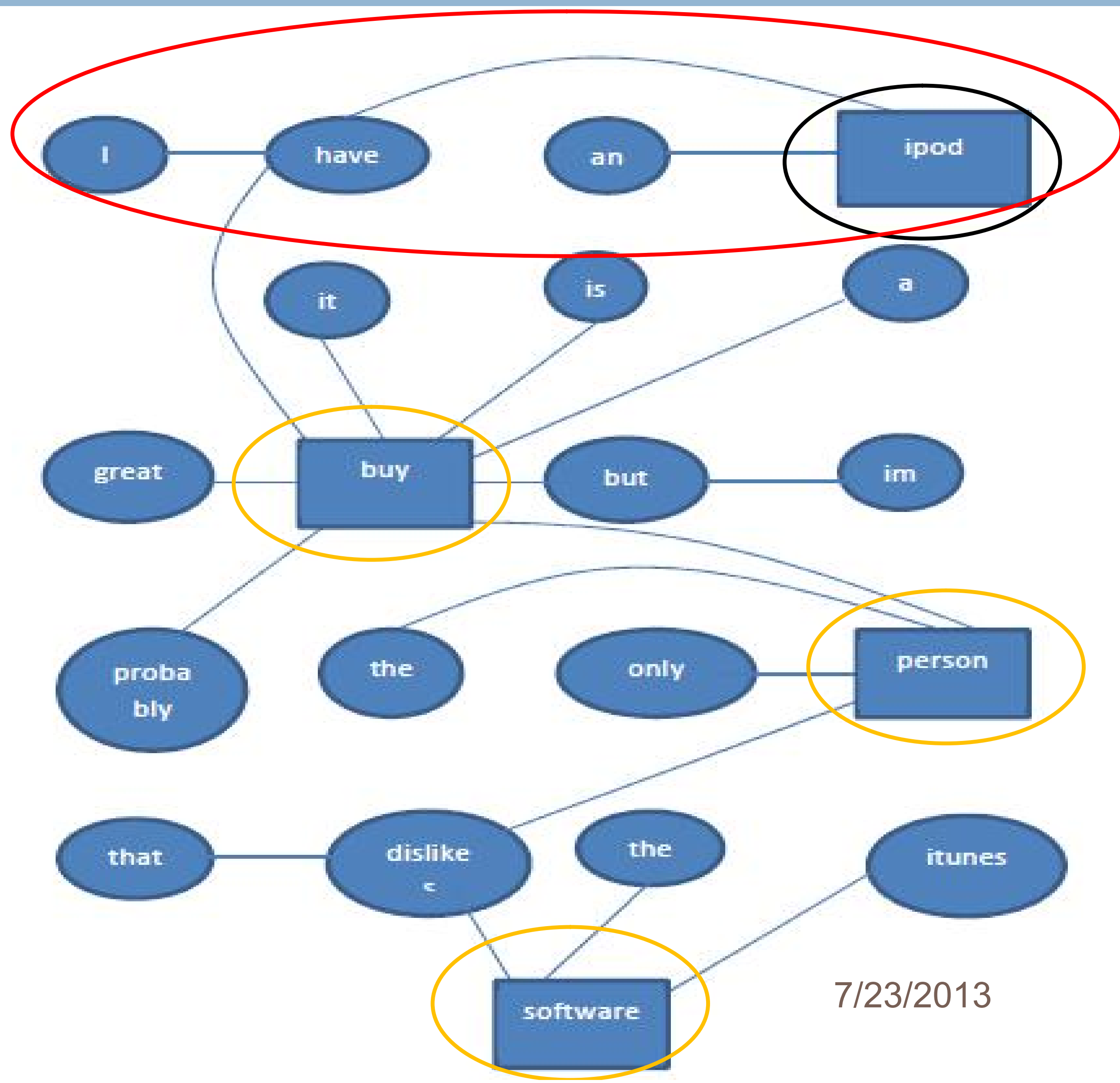
Clustering


I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
s
the
itunes
software

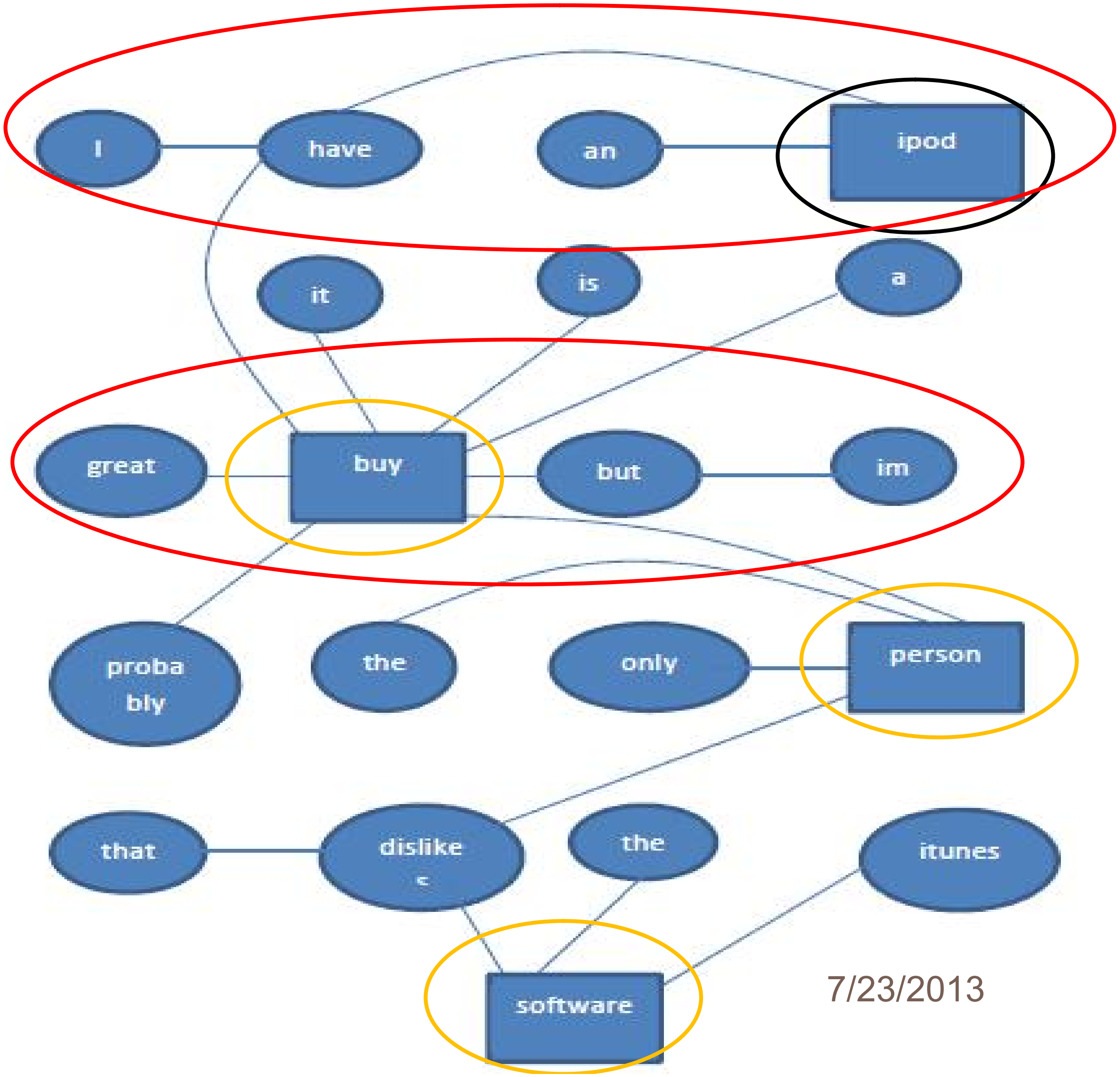
Clustering


I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
c
the
itunes
software

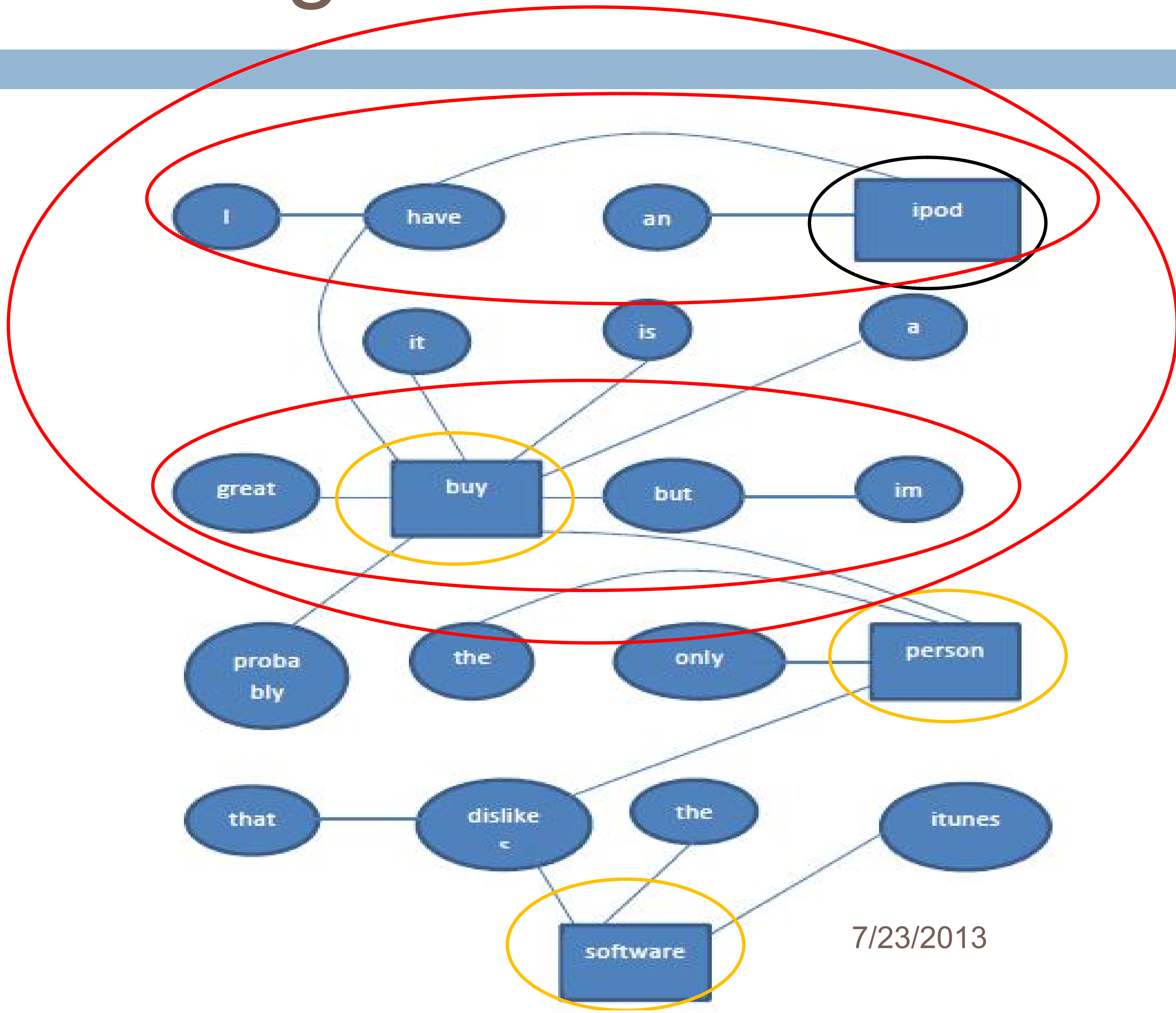
Clustering


I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
s
the
itunes
software

Clustering


I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
s
the
itunes
software

Clustering


I
have
an
ipod
it
is
a
great
buy
but
im
proba
bly
the
only
person
that
dislike
s
the
itunes
software

Evaluation – Dataset 1

- 2500 sentences
- Varied domains like *antivirus, camera, dvd, ipod, music player, router, mobile*
- Each sentence tagged with a *feature* and polarity w.r.t the *feature*
- *Acid Test*
  - *Each Review has a mix of positive and negative comments*

Parameter Learning

- Dependency Parsing uses approx. 40 relations.
- Relation Space – $(2^{40} - 1)$
- Infeasible to probe the entire relation space.
- *Fix relations certain to be significant*
  - *nsubj, nsubjpass, dobj, amod, advmod, nn, neg*
- *Reject relations certain to be non-significant*

Parameter Learning Contd…

- This leaves around 21 relations some of which may not be signficant.
- Compute Leave-One-Relation out accuracy over a training set.
- Find the relations for which there is ***significant accuracy change.***

Ablation test

Relations	Accuracy (%)
All	63.5
Dep	67.3
Rcmod	65.4
xcomp, conj_and ccomp, iobj	61.5
advcl, appos, csubj, abbrev, infmod, npavmod, rel, acomp, agent, csubjpass, partmod, pobj, purpcl, xsubj	63.5

Ablation test

Relations	Accuracy (%)
All	63.5
Dep	67.3
Rcmod	65.4
xcomp, conj_and ccomp, iobj	61.5
advcl, appos, csubj, abbrev, infmod, npavmod, rel, acomp, agent, csubjpass, partmod, pobj, purpcl, xsubj	63.5

Ablation test

Relations	Accuracy (%)
All	63.5
Dep	67.3
Rcmod	65.4
xcomp, conj_and ccomp, iobj	61.5
advcl, appos, csubj, abbrev, infmod, npavmod, rel, acomp, agent, csubjpass, partmod, pobj, purpcl, xsubj	63.5

Significant Relations Contd…

Relation Set	Accuracy
With Dep+Rcmod	66
Without Dep	69
Without Rcmod	67
Without Dep+Rcmod	68

- Leaving out *dep* improves accuracy most

Significant Relations Contd…

Relation Set	Accuracy
With Dep+Rcmod	66
Without Dep	69
Without Rcmod	67
Without Dep+Rcmod	68

- Leaving out *dep* improves accuracy most

Inter cluster distance

□	**Accuracy (%)**
2	67.85
3	**69.28**
4	68.21
5	67.4

Inter cluster distance

	Accuracy (%)
2	67.85
3	**69.28**
4	68.21
5	67.4

Lexicon based classification

Domain	Baseline 1 (%)	Baseline 2 (%)	Proposed System (%)
Antivirus	50	56.82	**63.63**
Camera 1	50	61.67	**78.33**
Camera 2	50	61.76	**70.58**
Camera 3	51.67	53.33	**60.00**
Camera 4(Nikon)	52.38	57.14	**78.57**
DVD	52.21	63.23	**66.18**
IPOD	50	57.69	**67.30**
Mobile 1	51.16	61.63	**66.28**
Mobile 2	50.81	65.32	**70.96**
Music Player 1	50.30	57.62	**64.37**
Music Player 2	50	60.60	**67.02**
Router 1	50	58.33	**61.67**
Router 2	50	59.72	**70.83**

Lexicon based classification

Domain	Baseline 1 (%)	Baseline 2 (%)	Proposed System (%)
Antivirus	50	56.82	**63.63**
Camera 1	50	61.67	**78.33**
Camera 2	50	61.76	**70.58**
Camera 3	51.67	53.33	**60.00**
Camera 4(Nikon)	52.38	57.14	**78.57**
DVD	52.21	63.23	**66.18**
IPOD	50	57.69	**67.30**
Mobile 1	51.16	61.63	**66.28**
Mobile 2	50.81	65.32	**70.96**
Music Player 1	50.30	57.62	**64.37**
Music Player 2	50	60.60	**67.02**
Router 1	50	58.33	**61.67**
Router 2	50	59.72	**70.83**

Overall accuracy

Method	Average Accuracy(%)
Baseline 1	50.35
Baseline 2	58.93
Proposed System	**70.00**

Overall accuracy

Method	Average Accuracy(%)
Baseline 1	50.35
Baseline 2	58.93
Proposed System	**70.00**

Evaluation – Dataset 2

- Extracted 500 sentences
- Varied domains like *camera, laptop, mobile*
- Each sentence tagged with a *feature* and polarity w.r.t the *feauture*

Results

Method	Accuracy (%)
Baseline 1	68.75
Baseline 2	61.10
CFACTS-R	80.54
CFACTS	**81.28**
FACTS-R	72.25
FACTS	75.72
JST	76.18
Proposed System	**80.98**

Results

Method	Accuracy (%)
Baseline 1	68.75
Baseline 2	61.10
CFACTS-R	80.54
CFACTS	**81.28**
FACTS-R	72.25
FACTS	75.72
JST	76.18
Proposed System	**80.98**

Results

Method	Accuracy (%)
Baseline 1	68.75
Baseline 2	61.10
CFACTS-R	80.54
CFACTS	**81.28**
FACTS-R	72.25
FACTS	75.72
JST	76.18
Proposed System	**80.98**

CONCLUSIONS

- Incorporating feature specificity improves sentiment accuracy.

- Dependency Relations capture long range dependencies as is evident from accuracy improvement.

- Work to be extended for ***implicit features and domain dependent sentiment.***