Automatic Non-Taxonomic Relation Extraction from Big Data in Smart City

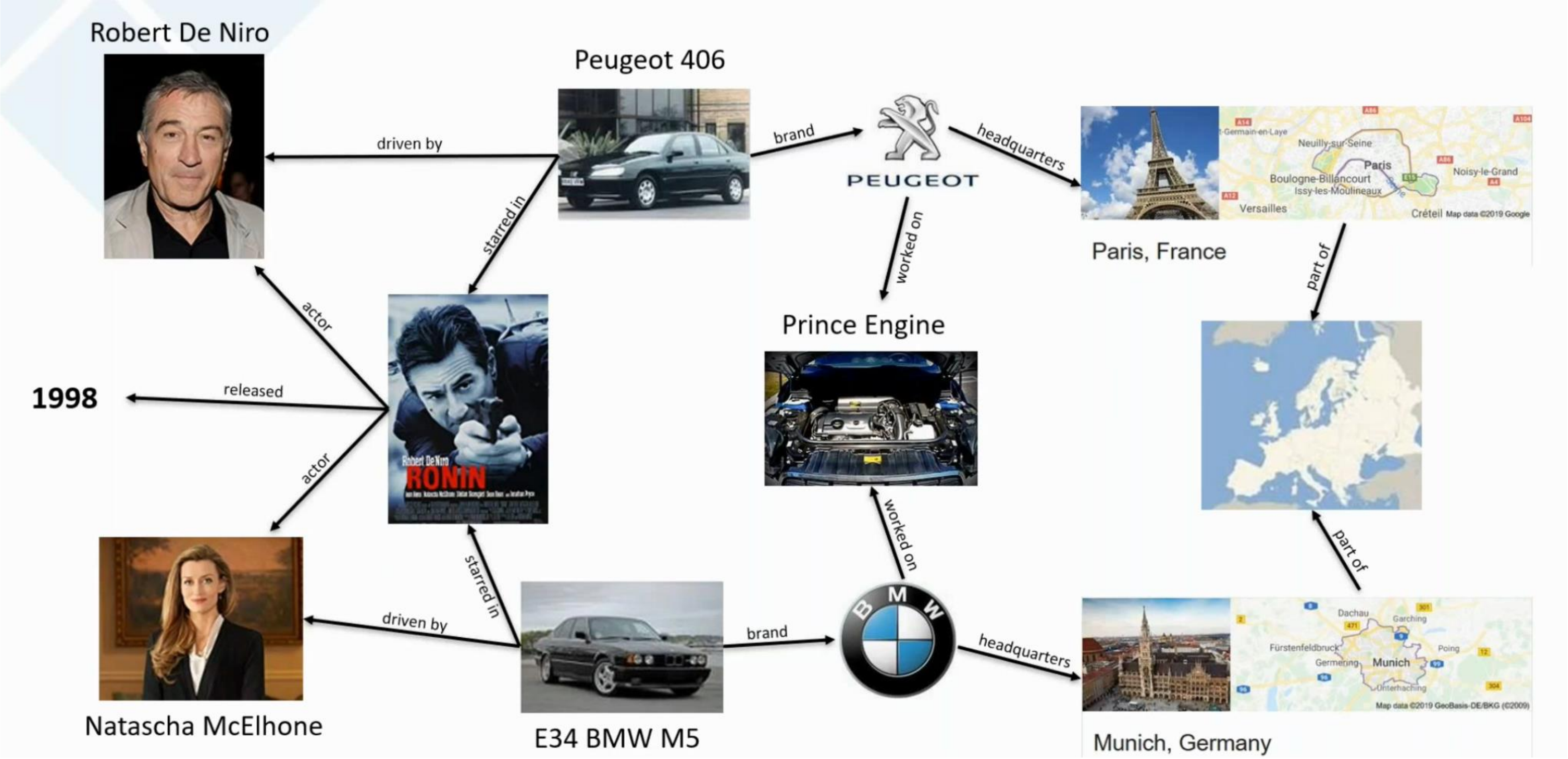
JING QIU, YUHAN CHAI, YAN LIU , ZHAOQUAN GU , SHUDONG LI , AND ZHIHONG TIAN

(Date of publication November 15, 2018, date of current version December 27, 2018, IEEE Access)

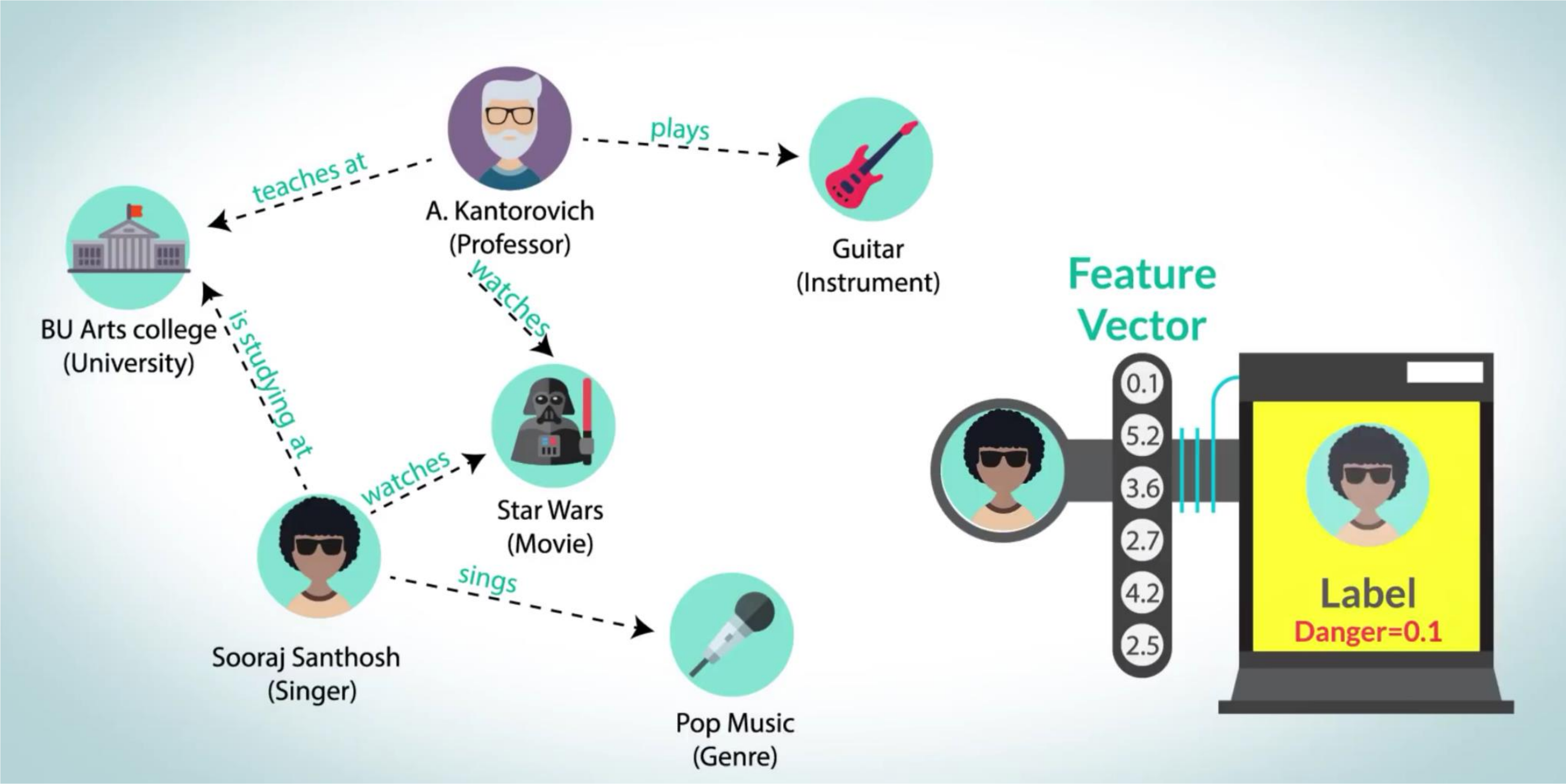
Research Area

- Knowledge Extraction
  - Non-taxonomic relations
  - Semantic graph
  - Dependency relation (Dependency syntactic identification)

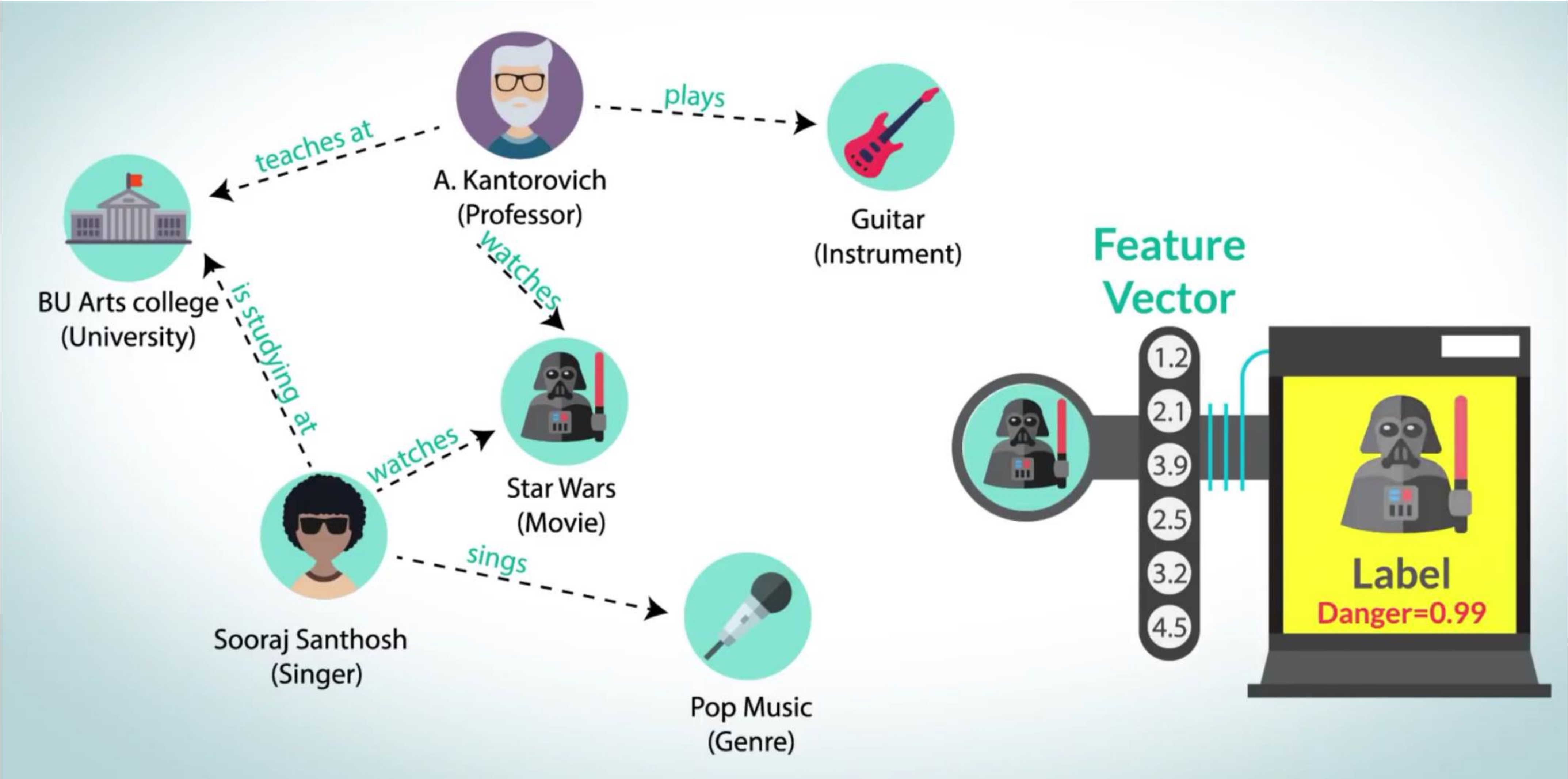
Knowledge Graph



Decision from Knowledge Graph (1/2)



Decision from Knowledge Graph (2/2)



The Problem

- Huge amount of data in smart city makes challenges for extracting useful information.
- Major research works and applications are in *Deep Learning* and *Knowledge Graph*, involving subtask (most difficult!) in extracting non-taxonomic relationship.
- Most research work have:
  - Limited practical applications
  - Neglected in syntactic and semantic information in relations that are extracted

Research Contribution

- This research contributes to:
  A. Extracting non-taxonomic relation at higher precision, based on combination of (i) semantic graph and (ii) context information.
  B. Providing ‘better’ labels for relations that have been extracted from A, thanks to combination of (i) dependency syntactic information and (ii) statistical information

Approach

Step 1:
Identify non-taxonomic relations

- SGNRI (semantic graph based non-taxonomic relationships identification)
- LDA
- Word2Vec

- Context 1: domain relative text
- Context 2: encyclopedia

Step 2: Labelling non-taxonomic relations

- **Verb** extraction
- Syntactic dependency parsing
- Statistical selection

Result – Step 1

TABLE 2. Results of different methods for identify non-taxonomic relations.

Method	P	R	F1
Apriori-based model	37.5%	8.1%	13.3%
Word2vec-based model	53.9%	66.2%	59.4%
LDA-based model	79.7%	68.9%	73.9%
SGNRI[1](LDA)	76.7%	75.7%	**76.2%**
SGNRI[2](LDA)	77.3%	78.4%	**77.9%**
SGNRI[1](Word2Vec)	77.9%	81.1%	**79.5%**
SGNRI[2](Word2Vec)	86.4%	77.0%	**81.4%**

Result – Step 2

- Each relation has top-5 verb candidate labels (total 370 = 74*5)
- Domain experts rating ‘Good’/’Bad’ verb per relations

TABLE 4. **Extraction results for the concept pair <Team, Player>.**

Concept Pair	Label Set	Common Sense	Score	Label Rating
“Qiudui, Qiuyuan” <Team, Player>	“chengwei” (“become”) “chengwei” (“become”)	“chengwei” (“become”)	3.569 3.569	Bad
	“qianding” (“signed”) “qianding” (“signed”) “qianyue” (“signed”) “qian” (“signed”) “qianding” (“signed”)	“yueidng” (“signed”)	2.99 1.21 0.91 0.53 0.34	Good
	“yinjin” (“introduce”) “yinjin” (“introduce”)	“tichu” (“propose”)	2.68 2.68	Good
	“tixing”(“remind”) “tixing” (“remind”) “zhonggao” (“advise”)	“quanshuo” (“persuade”)	2.65 1.36 1.29	Bad
	“baokuo” (“contain”) “baokuo” (“contain”)	“baokuo” (“contain”)	2.26 2.26	Good

Future Works

- Extracting good formal verb
- Clustering verbs with same meaning
- Considering information from multiple sources (e.g., sensing devices, multimedia data,…) – not only web sources as in this paper - can help to improve performance.

thank you