

Table of contents

Abstract	3
1 Introduction	8
2 Literature review	9
2.1 Data mapping	9
2.2 Recent research papers	10
2.2.1 Electronic payments data	10
2.2.2 Mobile phone usage data	11
2.2.3 Sensor data	12
2.2.4 Satellite images data	13
2.2.5 Price data	14
2.2.6 Textual data	15
3 A general data conversion framework for unstructured numerical Big Data	16
3.1 Conceptual setting	16
3.2 Aggregation	17
3.3 Features extraction.....	18
3.4 Data mining	18
3.4.1 Random subsampling	21
4 Empirical examples	26
4.1 Financial markets data	26
4.2 Mobile phone usage data.....	28
4.3 Sensor data	32
4.4 Online prices data	36
4.5 Online search data	38
4.6 Social media.....	41
5 From Reuters News data to uncertainty indexes	44
5.1 Obtaining the data	44
5.2 Obtaining a list of links.....	45
5.3 Removing duplicates.....	46
5.3.1 Removing broken links	47
5.3.2 Removing duplicate URLs.....	48
5.3.3 Removing duplicate headlines	48

5.4	Scraping new articles.....	49
5.5	Article volume over time.....	50
5.6	Constructing the uncertainty index.....	51
	5.6.1 Empirical estimator.....	51
	5.6.2 From word mentions to indexes.....	52
5.7	Indices.....	52
6	Conclusions	53
	References	54
	Appendix	57