A data-mining approach to the Parkour discipline

Un approccio data-mining alla disciplina del Parkour

Paola Pasca, Enrico Ciavolino and Ryan L. Boyd

Abstract Parkour is a relatively new discipline. As an uncommon nexus between risk, resistance, and the philosophy of overcoming obstacles, it is continuously gaining interest and popularity. However, the "voice" of parkour practitioners has been explored only qualitatively, on small samples and from a phenomenological point of view. In this work, raw data from the official American Parkour forum (from 2005 to 2013) have been web-scraped and treated with the Meaning Extraction Method (MEM), a simple and flexible technique providing optimal dimension reduction and the identification of broader themes related to the parkour discipline, thus providing a broader vision of a phenomena which is configuring itself as a true life-style.

Abstract *Il parkour è uno sport piuttosto recente: come nesso tra rischio, resistenza e filosofia degli "ostacoli da superare", è diventata una disciplina sempre più interessante e popolare. Tuttavia, la voce dei praticanti è stata esplorata solo in una ricerca qualitativa, effettuata su un piccolo campione, e solo da un punto di vista fenomenologico. In questo lavoro, per la prima volta, sono stati estratti dati provenienti dall'intero forum American Parkour (dal 2005 al 2013) e sono stati trattati con il Meaning Extraction Method (MEM), una tecnica semplice e flessibile che permette di ridurre la dimensionalità dei dati e di identificare i temi più ampi e caratteristici del parkour, una disciplina che si configura sempre più come uno stile di vita.*

Key words: parkour, online language use, Meaning Extraction Helper

Pasca Paola, e-mail: paola.pasca@unisalento.it
Enrico Ciavolino, e-mail: enrico.ciavolino@unisalento.it
Department of History, Society and Human Studies. University of Salento, via di Valesio, 73100, Lecce (Italy)
Ryan L. Boyd, e-mail: ryanboyd@utexas.edu
Department of Psychology, University of Texas at Austin, 108 E Dean Keeton St, Austin, TX 78712, Stati Uniti

1 An introduction to the Parkour discipline

Parkour is a relatively recent practice which is continuouly gaining interest and popularity, due to a significant presence on both tv and the *web*. Its first manifestations started as a game for its founders David Belle and Sébastian Foucault, in the context of the deprived parisian suburbs [3].

However, the roots of parkour can be dated back prior to the First World War: Georges Hébert, a military trainer of the French Navy, believing in intense physical training as a means for self-development, created a training method that combined physical obstacles with psycho-emotional barriers. The so-called *Natural Method* encouraged training in unconstrained, outdoor settings, where different terrain or obstacles could be faced and overcome. Grounded on ten main groups of exercises (walking, running, jumping, quadrupedal movement, climbing, balancing, throwing, lifting, defending and swimming), the *Natural Method* aims at making mind and body agile and adaptive in any situation: its fast, fluid and forward movements transposed into the urban environment are the essence of parkour [3, 17].

Despite its popularity, parkour is still permeated by prejudice. On a large scale, thoughtlessness, reckless behaviour and even suicide is sometimes attributed to it, even if there are no formal statistics related to deaths caused by parkour; moreover, training at one's own physical level is at the heart of the practice and practice itself starts from a ground level where falls are not life threatening. On a smaller scale, parkour practitioners, better known as *traceurs*, are often discouraged from practicing or banished by authorities, even if they almost always take care of their training spots by keeping them safe, clean and populated [17].

These controversial aspects make parkour an interesting topic for research and indepth exploration. Previous literature shed light on various aspects of the discipline [5, 19, 3, 13]. However, it is mainly focused on a small-group scale, based on qualitative data (interviews, narratives) on which no statistical processing has ever been performed.

This paper approaches the parkour discipline through simple data-mining techniques. It focuses on the entire American Parkour forum and lets the voice of *traceurs* and interested people emerge with the help of recent software developments for the treatment, processing and meaning extraction from linguistic data [10, 7].

A brief description of dataset, procedures and methods is reported in section 2. Some of the results of the natural language processing are reported in section 3.

2 Data and Methods

When it comes to the collection of unstructured data from the *web*, *web scraping* may help [16]. In the field of computer science, the term refers to several methods for data extraction and for their conversion into metadata which can further be processed and analyzed. In this case, one of the simplest, most intuitive software for efficient collection of *web* data has been used: *Helium Scraper* [2].

Through this software, linguistic data were exported from one of the biggest and most active online communities of parkour/freerunning practitioners: the American Parkour Forum [1]. There, *traceurs* and *freerunners* can share experiences, advices and ideas related to the discipline.

2.1 Dataset

The raw dataset covers a period of time ranging from 2005 to 2013. It consists of a .csv file containing *unique Id* of the post, *text in the post, number of responses, dates* and *times, author* of the thread, thread *title*.

On the raw dataset, a consistent amount of threads containing few hundreds of words could be noticed (precisely, 2692 threads contained an amount of words equals or inferior to 640 words).

In order to facilitate the information extraction in the text processing phase, authors decided to exclude from the analysis the threads with a word count lower than the minimum determined for segmentation, that is, 100 words. With respect to the whole corpus of data, this choice prevents from including threads likely to contain repetitions and comment citations, that is, less informative data. Building on these assumptions, we retained threads whose overall word count was above or equal to 5000 words.

Moreover, we excluded one thread which was in arabic language, and removed *links* and *urls* mentioned within the texts. Grouped by thread, the final dataset ($>$5000) included a total of 249 threads and 17275 comments.

2.2 Meaning Extraction Method

The Meaning Extraction Method (MEM), developed by Chung and Pennebaker [11] is based on the assumption that words related to a particular topic will tend to be used together. The mere observation of associated words allows researchers to draw inferences on how much people are talking, about what topics and in what way. This simple assumption makes application and interpretation procedures easier and more efficient.

MEM procedures start with a binary dataset of used-vs-non used words. This first step requires cleaning data from stop words (words that carry little to no meaning, such as ”the”, ”you”, ”did”), uncommon and non-informative words, in order to be sure that the most meaningful themes will be extracted. Then, a common factor analytic approach such as Principal Component Analysis can be applied to the word-by-observation table. Applied to natural language, results will provide word clusters reflecting broader themes emerged from the sample of text.

In order to speed up the front-end procedures used in this (and other) topic modeling approaches, Boyd released the Meaning Extraction Helper (MEH) [8], a free-

ware that automates many of the steps described above. Its simplest application requires the user to make few basic selections, to point the software to the location of the .txt files and run the analysis. MEH efficiently converts unstructured linguistic data into structured matrices ready to be analyzed, leaving researchers to apply statistical techniques for meaning extraction [7].

In literature, several applications confirmed MEM as a promising method for uncovering information regarding psychological dimensions [11], personal values [9] and cultural self-schemas [18]. In particular, a recent application [6] proved MEM efficacy in catching how people think and talk online.

Building on previous literature, the American Parkour dataset has been processed with MEH through the following settings:

- upload a list of stop words (both the default one and another stop word list determined by the first author who, as a parkour practitioner, is familiar to parkour language);
- upload conversions (e.g. words such as *compete, competitive, competing* to be coded as *competition*);
- split files into equally sized segments (38 per thread). The number of segments is the ratio between the *mean by thread* value and a pre-determined words-per-segment value, that is 250. As said above, the minimum size (in terms of word count) admitted for a segment to be included in the output is set to 100 words.
- setting the minimum word percentage of word appearance in order to be included in the outputs to 3%.

The rationale for text processing choices is to make the most out of the combination of data descriptive statistics and MEH characteristics. The average number of segments (38) will split small threads into very small segments (in terms of word count) and bigger threads into big ones. However, the 3% required to a word to be included for analysis ensures to take the very essential from the smallest threads and more contents from bigger threads.

MEH uses a dictionary to identify common content words in each segment and systematically assigns a binary score to each word according to presence ("1"), or absence ("0"). After processing each word in each segment, MEH generates an output file that identifies the words and shows in which segments they are used.

Moreover, frequency of each common word across all the text observation is computed. The output file can be uploaded into statistical software (e.g. R, SPSS). A Principal Component Analysis with varimax rotation allows to extract components reflecting broader themes across all texts.

3 Results

Top 50 most frequently used words are reported in table 1.

Table 1 50 most frequent words across 9462 segments

Word	**Frequency**	**Word**	**Frequency**
competition	3357	friend	1311
work	3063	strong	1294
jump	2427	complete	1197
community	2338	technique	1181
body	1980	level	1124
life	1979	ability	1120
flip	1807	fast	1096
freerun	1763	martial	1078
foot	1726	goal	1072
discipline	1673	walk	993
sport	1668	important	975
hard	1651	interest	956
vault	1606	difference	953
change	1516	focus	925
fun	1457	roll	925
mind	1439	hand	914
strength	1434	group	911
kid	1421	climb	908
physical	1417	challenge	905
experience	1400	fight	903
philosophy	1398	improve	901
wall	1379	practitioner	893
skill	1378	environment	890
movement	1350	human	879
love	1344	style	876

The top 50 most frequent words themselves reveal relevant aspects of parkour: words related to it as a concept (*competition, discipline, philosophy, sport*), as well as words reminding to social aspects (*community, friend, group, practitioner*) and more technical words (*jump, flip, vault, wall, roll*).

PCA results along with descriptives are reported in table 2. The parkour corpora can be summarized by 11 themes that globally accounted for the 26% of variance.

Table 2 Corpora descriptives

Component	1	2	3	4	5	6
Eigenvalue	2.95	1.81	1.69	1.62	1.59	1.47
Variance %	4.27	2.62	2.45	2.36	2.31	2.13
M (*SD*)	.08 (.27)	.08 (.27)	.08 (.26)	.08 (.26)	.08 (.26)	.09 (.28)
Component	7	8	9	10	11	
Eigenvalue	1.45	1.43	1.42	1.36	1.32	
Variance %	2.1	2.07	2.05	1.97	1.91	
M (*SD*)	.08 (.26)	.09 (.28)	.08 (.27)	.07 (.25)	.08 (.26)	

Notes. The Kaiser-Meyer-Olkin measure was .79, above the recommended value of .6 [14, 15], and Bartlett's Test of Sphericity [4] reached statistical significance, thus confirming the correlation matrix of components.

Table 3 illustrates the 11 themes emerged from PCA analysis, their meaning and the loadings.

Table 3 Component information and loadings

Component	1 - Being Practitioner	2 - Values	3 - Outdoor	4 - Technique
	ability .559	discipline .582	climb .679	land .68
	strength .546	practitioner .482	wall .613	roll .65
	physical .56	philosophy .361	vault .461	technique .353
	improve .465	sport .329	jump .432	jump .35
	method .451	community .323	building .358	body .32
	skill .398	freerun .315	small .178	hurt .282
	challenge .395	martial .29	environment .176	vault .268
	strong .375	experience .278	efficient .176	basic .235
	body .318	purpose .274		ground .213
	life .314	competition .27		strength .187
	goal .296	respect .247		strong .17
	movement .277	physical .233		
	environment .271	involve .218		
	important .238	important .18		
	limit .236	environment .179		
	face .229			
	focus .221			
	work .217			
	hard .213			
	technique .212			
	basic .21			
	experience .2			
	practitioner .178			
	mind .169			

Component	5 - Freerunning	6 - Meetings	7 - Mindset	8 - Interactions
	flip .616	jam .454	mind .53	kid .464
	trick .589	group .438	open .453	fun .418
	freerun .536	community .45	human .362	friend .372
	movement .278	together .43	body .328	life .357
	efficient .231	work .33	change .33	love .318
	style .24	basic .254	community .245	sport .252
	fun .189	small .229	close .24	honest .24
	competition .167	experience .229	competition .229	money .29
	experience -.206	month .24	turn .214	hard .186
		human -.252	grow .199	hurt .179
		body -.244	involve .185	philosophy .177
			goal .179	
			purpose .17	

Component	9 - Competition	10 - Martial Art	11 - Injury
	push .588	martial .593	face .55
	limit .43	fight .534	head .379
	competition .418	style .477	ground .374
	sport .31	technique .291	fall .371
	hard .262	focus .245	fight .351
	focus .255	movement .212	hand .288
	money .26	efficient .212	turn .248
	challenge .176	hand .183	hurt .17
	efficient -.217		sport -.207

Values with an absolute value potentially rounded to .2 were used in order to identify PCA components [11, 9]. The themes emerged have been evaluated (in terms of meaning) by both the first author and an italian parkour instructor (see

the *acknowlegdement* section for more info). They cover many facets of the parkour discipline, related both to the global status of being a practitioner (*component 1*), but also to more specific dimensions: the idea of parkour as discipine and philosophy (*component 2*) is largely reported by the most conservative practitioners on a smaller scale [12, 19].

The *outdoor* and *technique* components mention specific parkour moves (*vault, jump, wall, roll*) and environment words (*building, ground*). It is interesting to note that the word *efficient* loads positively within the *outdoor* component, whereas it loads negatively in the *competition* component suggesting, in agreement with literature [3, 12], that when it comes to moving in real environments, efficiency matters more than spectacularity or aesthetic . This is also coherent with Georges Hébert's principle that inspired parkour: *be strong to be useful* [19].

Also, references to parkour *jams*, that is, intensive training meetings, as well as a strong social component appeared: in the context of an open-minded attitude towards discipline and life, a precious contribution to personal and physical development comes from confrontation and advise of other *traceurs*, and from sharing both sweat and goals.

4 Conclusions

In this paper, Meaning Extraction Method was used to web-scrape and analyse data from the official American Parkour forum. The method provided an optimal dimensionality reduction and identify some latent themes that underlay the vision of a Parkour discipline.

The obtained results showed that the Parkour discipline confirmed and extended the vision of a phenomenon involving global existence and perceived as a way of life.

Acknowledgements A special thanks to Ryan L. Boyd for providing the first author with the largest parkour dataset so far and introducing her to MEM.

A special thanks to Alessandro Pennella, one of the best certified parkour trainers in Italy, for examining and commenting the themes emerged from the analysis.

References

[1] APK, American Parkour Forum. http://americanparkour.com/smf/. Accessed: 2013-09-10.

[2] Helium Scraper. Extract data from any website. http://www.heliumscraper.com/en/index.php?p=home. Accessed: 2013-09-10.

[3] M. Atkinson. Parkour, anarcho-environmentalism, and poiesis. *Journal of sport and social issues*, 33(2):169–194, 2009.

[4] M. S. Bartlett. A note on the multiplying factors for various χ 2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 296–298, 1954.

[5] N. Bavinton. From obstacle to opportunity: Parkour, leisure, and the reinterpretation of constraints. *Annals of leisure research*, 10(3-4):391–412, 2007.

[6] K. G. Blackburn, G. Yilmaz, and R. L. Boyd. Food for thought: Exploring how people think and talk about food online. *Appetite*, 123:390–401, 2018.

[7] R. Boyd and J. Pennebaker. A way with words: Using language for psychological science in the modern era, 10 2015.

[8] R. L. Boyd. Meaning Extraction Helper (Version 1.2.65)[Software], 2015. URL http://meh.ryanb.cc.

[9] R. L. Boyd, S. R. Wilson, J. W. Pennebaker, M. Kosinski, D. J. Stillwell, and R. Mihalcea. Values in words: Using language to evaluate and understand personal values. 2015.

[10] G. G. Chowdhury. Natural language processing. *Annual review of information science and technology*, 37(1):51–89, 2003.

[11] C. K. Chung and J. W. Pennebaker. Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of research in personality*, 42(1):96–132, 2008.

[12] J. L. Clegg and T. M. Butryn. An existential phenomenological examination of parkour and freerunning. *Qualitative research in sport, exercise and health*, 4(3):320–340, 2012.

[13] P. Gilchrist and B. Wheaton. Lifestyle sport, public policy and youth engagement: examining the emergence of parkour. *International journal of sport policy and politics*, 3(1):109–131, 2011.

[14] H. F. Kaiser. A second generation little jiffy. *Psychometrika*, 35(4):401–415, 1970.

[15] H. F. Kaiser. An index of factorial simplicity. *Psychometrika*, 39(1):31–36, 1974.

[16] D. Murthy, A. Gross, A. Takata, and S. Bond. Evaluation and development of data mining tools for social network analysis. In *Mining Social Networks and Security Informatics*, pages 183–202. Springer, 2013.

[17] A. O'grady. Tracing the city–parkour training, play and the practice of collaborative learning. *Theatre, dance and performance training*, 3(2):145–162, 2012.

[18] N. Ramírez-Esparza, C. K. Chung, G. Sierra-Otero, and J. W. Pennebaker. Cross-cultural constructions of self-schemas: Americans and mexicans. *Journal of Cross-Cultural Psychology*, 43(2):233–250, 2012.

[19] S. J. Saville. Playing with fear: Parkour and the mobility of emotion. *Social & cultural geography*, 9(8):891–914, 2008.