Current paradigms in intelligent transportation systems

S.L. Toral[1] M.R. Martínez Torres[2] F.J. Barrero[1] M.R. Arahal[1]

[1]E.S. Ingenieros, University of Seville, Avda. Camino de los Descubrimientos s/n, Seville 41092, Spain
[2]E.U.E. Empresariales, University of Seville, Avda. San Francisco Javier s/n, Seville 41018, Spain
E-mail: toral@esi.us.es

Abstract: Intelligent transportation systems (ITS) constitute today a multidisciplinary field of study involving a large number of different research areas. As a consequence, it is difficult to achieve a structured view of ITS, which is necessary to unify efforts and as guidance for future developments. This study aims to identify the main paradigms in the field of ITS by semantically analysing studies related to this general topic. An understanding about which research is considered valuable by the research community to build upon may provide valuable insights in this field. As a result of the statistical treatment of data, up to 13 paradigms are obtained. The scope of these paradigms and the relationships between them have also been detailed, providing a structured vision of ITS synthesised in a map form.

1 Introduction

Intelligent transportation systems (ITS) have been investigated for many years in Europe, North America and Japan, with the aim of improving the safety and efficiency of road transport and environmental conservation. To this end, new technologies and computer power have been applied to freeway, traffic and transit systems [1, 2]. ITS can be considered a global phenomenon, attracting worldwide interest from transportation professionals, the automotive industry and political decision makers [3].

ITS involves a large number of research areas spread over many different technological sectors such as electronics, control, communications, sensing, robotics, signal processing and information systems [4, 5]. This multidisciplinary nature increases the problem's complexity because it requires knowledge transfer and cooperation among different research areas [6]. One of the main problems of being a complex and multidisciplinary field is the difficulty of dealing with ITS as a development and research area [7, 8]. Traditional topics of interest are changing and new ones are emerging due to the continuous advance of emergent technologies and the economical, social and environmental implications of ITS.

The complexity of the topic suggests that it will be of benefit to define a global, structured view. This view will be useful to support close integration of ITS with conventional transportation initiatives as well as to provide guidance for future ITS deployments. Sharing a common structured view will also help the promotion of ITS standards development, the identification and confirmation of needs, problems, objectives and issues, and the alignment of researchers, companies and users for synergy [3].

The idea of obtaining this structured view of a particular research and development area is not new, but nowadays is taking on major importance. The most recent and ambitious attempt has been made by the European commission by launching the technological platform called ARTEMIS [9]. The aim of ARTEMIS is to develop and drive a joint European vision on embedded systems connecting research and development with innovation to align fragmented R&D efforts along common strategic agenda and looking for improvements in European companies' efficiency and competitiveness. ARTEMIS is following a bottom-up scheme in which the most prominent European companies are getting involved in the definition of this unique European vision through the development of a strategic research agenda (SRA). In particular, ITS can be located in one of the four applications contexts identified by ARTEMIS SRA devoted to public infrastructure [9].

Although some structured view of ITS has been proposed based on market areas [3] or on databases and cumulative experience [10], this study proposes a quantitative and systematic methodology to identify the main paradigms within the broader field of ITS. The starting point is the information provided by the Institute for Scientific Information's (ISI) massive datasets. The abstracts of the published papers included in datasets will have been analysed using text categorisations tools to obtain the final paradigms. To meet these objectives, the paper has been structured as follows. In Section 2, ITS will be analysed as a field study, describing previous attempts of structuring this research topic. In Section 3, the proposed methodology will be described in detail, including a discussion about some other approaches. The obtained paradigms in ITS will be presented in Section 4, through the application of the proposed methodology. Finally, the main conclusions of the work are included in Section 5.

2 Analysis of ITS as a field study

There are different ways to classify and segment the ITS field. Six major categories were reviewed in [6] from a technological perspective:

- Advanced traffic management systems (ATMS), used to improve traffic service quality and to reduce traffic delays.

- Advanced travellers information systems (ATIS), used to supply real-time traffic information to travellers.

- Commercial vehicles operation (CVO), systems that use different ITS technologies to increase the safety and efficiency of commercial vehicles and fleets.

- Advanced public transportations systems (APTS), which make use of electronic technologies to improve the operation and efficiency of high-occupation transports, such as buses and trains.

- Advanced vehicles control systems (AVCS), which joint sensors, computers and control systems in driving assistance solutions.

- Advanced rural transports systems (ARTS), used to solve problems arising in rural zones (steep grades, blind corners, curves, scarce navigational signs, mix of users, lack of alternative routes).

Although this classification is shaped by emerging technologies and is useful from the viewpoint of system designers, some interesting topics related to ITS are excluded. For instance, transport policy and planning, traffic modelling and forecasting, and the sociological and behavioural influences of ITS are not included in the previous classification. The requirements and preferences of ITS users may also be given inadequate attention.

Another way of looking at ITS is to consider market areas as a representation of ITS users and operators with similar needs. Nine major market areas, detailed in Table 1, are defined in [3].

Similar classifications can also be found in [10], although they consider up to 13 areas by sub-dividing several of the nine market areas defined in Table 1.

As a difference to the previous classifications, mainly focused on one aspect of ITS, an integral analysis of the ITS field is proposed in this paper. The starting point will be the abstracts and keywords of published papers related to ITS included in the ISI Web of Science database [11, 12]. ISI Web of Science includes journals of almost

Table 1 Market areas in ITS

	Market area	Goal
area 1	traffic management	manage the entire road network on behalf of the general public
area 2	emergency management	respond to incidences and emergencies (fire, police, ambulance)
area 3	transportation planning	match transportation supply with demand both now and in the future
area 4	traveller information	supply information to traveller and subscribers
area 5	commercial vehicles	provide travel information and fleet management services
area 6	transit management	plan and operate transit systems in both urban and rural areas
area 7	intelligent vehicles	enhance the capabilities of road vehicles through the use of electronics, sensors, communications and control actuator technologies
area 8	incident management	concerned with efficiency and safety of the roadway network
area 9	payment systems	encompass all the people that take money in return for providing a service (toll road operators, transit agencies, car parking operators, etc.)

each scientific field. Consequently, not only technological and market approaches to ITS will be gathered, but any scientific discipline strongly or weakly related to ITS, such as planning and logistic, psychology or social sciences. The selected papers will be processed using a statistical text categorisation tool and, as a result, major paradigms in the field of ITS will be identified.

3 Methodology

The starting point of the proposed methodology consists of paper extraction from ISI databases. Particularly, papers related to the topic 'intelligent transportation systems'. Instead of analysing the full text, which would be an enormous task, a representative piece of text summarising the whole paper has been selected, i.e. the abstracts and keywords or index terms. An abstract is a condensed version of a paper that highlights the major points covered and concisely describes the content and scope of the writing, while keywords review the writing's contents in abbreviated form. Keywords are included because they emphasise the content of the paper. Although keywords themselves can constitute an adequate description of the paper content [13], they can also restrict the number of different topics to be obtained. Notice that sometimes keywords must be chosen among a closed list provided by the journal publisher. Consequently, the proposed methodology will consider both abstracts and keywords with the aim of leaving open the number of topics to be obtained. It is implicitly assumed that authors write good abstracts (representative of the content of the paper) and that keywords are carefully chosen.

In general, bibliometric research is devoted to quantitative studies of literature. Several empirical methods can be found in the literature. Co-citation methods are perhaps the most employed [14], and they have been frequently used to analyse the intellectual structure of many disciplines [15]. The basis of co-citation methods consists of counting the number of times certain markers occur or co-occur, giving rise to information on such author co-citation [16], journal co-citation, keyword co-citation, etc. [13].

The main drawback of traditional bibliometric techniques, such as author or journal co-citation methods, is that they are not concerned about the content of considered papers but on references usually delayed between 2 and 5 years after a paper is first drafted. Although they lead to interesting results, they do not provide an immediate picture of the actual content of the research topic dealt with in the literature. As a difference, semantic analysis based on co-words analysis (co-occurrences of words in the publications on a given subject) has the potential of solving this kind of problem [13–17].

Semantic analysis usually employs a vector space model [18], in which documents are summarised and represented by vectors of words (term vectors). However, a central problem in this kind of statistical analysis is the high dimensionality of the feature space (one dimension for each unique word). Therefore, it is desirable to first project the documents into a lower-dimensional subspace in which the semantic structure of the document space becomes clear [19]. In the low-dimensional semantic space, the traditional clustering algorithms can then be applied. To this end, spectral clustering [20, 21], clustering using latent semantic indexing (LSI) [22] and clustering based on non-negative matrix factorisation [23, 24] are the most well-known techniques. Particularly, LSI decomposes a term document matrix using a technique called singular value decomposition to construct new features as combinations of the original features, significantly reducing the high-dimensionality problem of the feature space [25]. Moreover, LSI considers documents that have many words in common to be 'semantically close', and ones with few words in common to be 'semantically distant'. The LSI approach makes three basic claims: that semantic information can be derived from a word-document co-occurrence matrix; that dimensionality reduction is an essential part of this derivation; and that words and documents can be represented as points in a Euclidean metric space.

A different approach has been applied in this paper. This approach is consistent with the first two of these claims, but it differs in the third, describing a class of statistical models in which the semantic properties of words and documents are expressed in terms of probabilistic topics [26]. The topic model is a statistical language model that relates words and documents through topics. It is based upon the idea that documents are mixtures of topics, where a topic is a probability distribution over words [26–28].

In this paper, the methodology proposed for ITS paradigms identification consists of a four-step procedure based on the latent Dirichlet allocation (LDA) method of [26] (see Appendix for more details) and illustrated in Fig. 1.

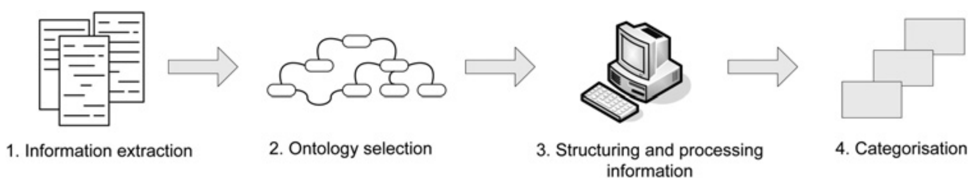

Figure 1 *Dataflow of the proposed methodology*

1. *Information extraction:* the extraction process involves access to ISI databases to annotate the abstract and keywords of a paper dealing with the topic 'intelligent transportation systems'.

2. *Ontology:* an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations between terms [29]. An ontological model of the domain is used as a facilitator throughout all the processes. This step provides a common vocabulary and specifies the semantics of key relationships within the domain. The selection is based on the obtained frequency-of-occurrence-rates of words over the total corpus of extracted texts.

3. *Structuring and processing information:* in statistical natural language processing, one common way of modelling the contributions of different topics to a document is to treat each topic as a probability distribution over words, viewing a document as a probabilistic mixture of these topics [27].

4. *Categorisation:* The algorithm outlined above can be used to find the topics that account for the words used in a set of documents. In this study, every abstract and its associated keywords are considered a document.

4 Paradigms in ITS

A total of 1147 papers related to the topic 'intelligent transportation systems' has been obtained from ISI databases, covering subject areas like engineering, transportation, computer science, telecommunications, automation and control systems, and operations research and management science. The majority of them belong to the general category of Science & Technology, but a small percentage is associated to Social Sciences and Arts & Humanities.

According to the explained procedure, abstracts and keywords have been used as documents in the terminology of semantic analysis. Up to 64 132 words have been extracted from these documents, leading to a vocabulary (non-repeated words) of 5485 words. Working with such an amount of words would be prohibitive in terms of computing time, so ontology is selected to facilitate the implementation of the procedure. The criterion for the ontology selection was based on the frequency of terms in the document collection. Particularly, words that occurred in more than 15 documents were considered [30]. As a result, an ontology of 267 words were obtained. This is the most manual-step of the procedure, because the final result must be supervised to remove words that are not directly related to ITS issues. Table 2 shows the key dimensions used in the topic model.

Table 2 Dimensions of the topic model

Parameter	Description	Value
D	number of documents in corpus	449
N	total number of words in corpus	34 132
L	average length of document in words ($L = N/D$)	76
V	number of words in vocabulary	5485
W	ontology	267
T	number of topics	–
ITER	number of iterations	200

T and ITER represent topic model run parameters. The number of Gibbs sampler iterations was chosen to be ITER = 200. This is a large enough value to guarantee the convergence of the algorithm [21]. The number of topics was selected using the perplexity value. Perplexity is a standard measure of performance for statistical models of natural language [31, 32] defined by (1).

$$pplex = \exp\left(-\frac{1}{W}\sum_{n=1}^{W}\log P(w_n|d_n)\right) \tag{1}$$

The role of perplexity has mostly been discussed on an intuitive level as average uncertainty when predicting the next word given its history. Perplexity indicates the uncertainty in predicting a single word. A lower perplexity score indicates better generalisation performance. Perplexity varies from 1 to W; lower perplexity is better, and the maximum perplexity of W is reached when all words in the vocabulary are equally likely. In our case study, the LDA algorithm was run for a number of topics varying between 1 and 30. The results are illustrated in Fig. 2.

The minimum perplexity value is reached for a number of topics equal to 13. Consequently, 13 was chosen as the number of selected topics. Table 3 shows the 13 topics obtained, with the most likely words in each topic, and their probabilities $P(w|t)$.

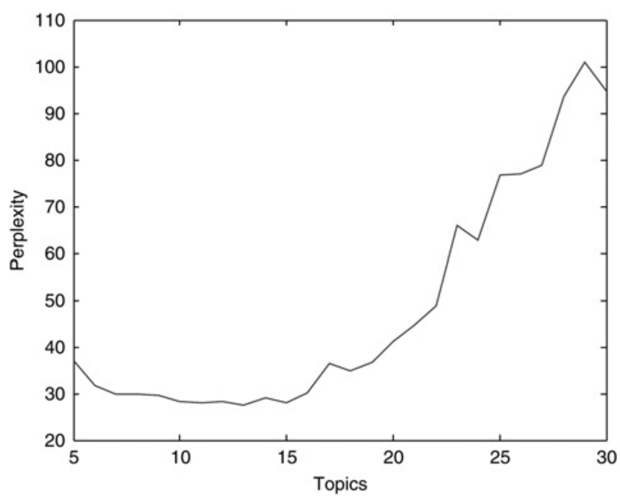

Figure 2 *Perplexity as a function of the number of topics*

Table 3 Obtained topics and probabilities from the LDA algorithm

TOPIC 1		TOPIC 2		TOPIC 3		TOPIC 4	
method	0.165	system	0.101	vehicle	0.265	incident	0.079
effect	0.087	service	0.074	vehicles	0.118	performance	0.067
benefit	0.073	systems	0.073	system	0.069	detection	0.065
analysis	0.072	transport	0.059	information	0.069	technique	0.061
methods	0.055	decision	0.054	transit	0.048	traffic	0.053
benefits	0.049	information	0.049	mobile	0.038	freeway	0.049
evaluation	0.047	framework	0.039	reserved	0.027	network	0.047
effective	0.047	support	0.039	automatic	0.025	management	0.047
result	0.037	environment	0.037	vision	0.025	conditions	0.047
methodology	0.032	assess	0.036	infrastructure	0.025	neural	0.040
intelligent	0.031	intelligent	0.032	approach	0.022	system	0.037
station	0.028	services	0.031	rights	0.021	real-time	0.033
potential	0.026	evaluate	0.030	function	0.021	techniques	0.033
effects	0.024	operations	0.023	changes	0.020	characteristics	0.031
accuracy	0.022	impacts	0.020	corridor	0.020	delay	0.031
alternative	0.022	technologies	0.019	motion	0.018	detector	0.027
effectiveness	0.021	analysis	0.018	moving	0.017	presented	0.027
experimental	0.021	weather	0.018	scenarios	0.017	strategies	0.023
TOPIC 5		TOPIC 6		TOPIC 7		TOPIC 8	
system	0.191	algorithm	0.187	transport	0.127	traffic	0.294
systems	0.134	algorithms	0.073	transportation	0.126	simulation	0.118
design	0.078	process	0.059	develop	0.089	signal	0.072
communication	0.060	present	0.057	system	0.066	results	0.047
present	0.051	develop	0.049	state	0.053	estimate	0.043
technology	0.046	approach	0.045	development	0.048	result	0.038
position	0.034	application	0.045	planning	0.041	parameters	0.034
highway	0.032	developed	0.044	region	0.039	estimates	0.032
message	0.028	sensor	0.043	architecture	0.038	intelligent	0.030
research	0.024	performance	0.041	intelligent	0.032	arterial	0.028
integrated	0.023	efficient	0.036	deployment	0.031	temporal	0.024
presented	0.020	complex	0.036	developing	0.028	roadway	0.023
implementation	0.020	applications	0.030	program	0.026	intersection	0.022
potential	0.019	transportation	0.026	issues	0.025	microscopic	0.020
communications	0.019	intelligent	0.022	management	0.023	spatial	0.019
global	0.017	scheme	0.022	process	0.022	empirical	0.017

Continued

Table 3 Continued

TOPIC 5		TOPIC 6		TOPIC 7		TOPIC 8	
context	0.017	processing	0.021	public	0.019	emissions	0.017
wireless	0.016	sensors	0.019	regional	0.019	evaluated	0.014
TOPIC 9		TOPIC 10		TOPIC 11		TOPIC 12	
transportation	0.144	control	0.138	network	0.164	model	0.356
transport	0.142	driver	0.092	problem	0.116	models	0.121
system	0.087	system	0.087	dynamic	0.089	traffic	0.067
application	0.070	vehicle	0.067	networks	0.069	develop	0.056
systems	0.068	systems	0.065	problems	0.042	transportation	0.054
location	0.045	result	0.053	present	0.040	approach	0.039
intelligent	0.044	driving	0.052	computation	0.035	modelling	0.036
technologies	0.043	drivers	0.041	solution	0.033	developed	0.035
user	0.042	results	0.039	routing	0.033	transport	0.030
applications	0.040	safety	0.033	guidance	0.030	urban	0.029
advanced	0.025	conditions	0.022	assignment	0.030	intelligent	0.028
research	0.024	significant	0.021	structure	0.028	forecasting	0.025
experience	0.022	behaviour	0.020	optimal	0.027	research	0.024
function	0.020	tracking	0.019	presented	0.024	demand	0.017
requirements	0.019	camera	0.016	real-time	0.022	systems	0.016
quality	0.017	autonomous	0.016	optimisation	0.022	regression	0.015
commercial	0.016	dynamics	0.015	shortest	0.020	present	0.012
solution	0.015	human	0.015	computational	0.020	feasibility	0.012
TOPIC 13							
travel	0.205	–	–	–	–	–	–
predict	0.088	–	–	–	–	–	–
information	0.071	–	–	–	–	–	–
times	0.068	–	–	–	–	–	–
prediction	0.045	–	–	–	–	–	–
atis	0.045	–	–	–	–	–	–
result	0.042	–	–	–	–	–	–
traveller	0.040	–	–	–	–	–	–
choice	0.035	–	–	–	–	–	–
results	0.031	–	–	–	–	–	–
estimate	0.025	–	–	–	–	–	–
travel-time	0.021	–	–	–	–	–	–
future	0.020	–	–	–	–	–	–
statistical	0.019	–	–	–	–	–	–

Continued

Table 3 Continued

TOPIC 13							
propagation	0.019	–	–	–	–	–	–
congestion	0.018	–	–	–	–	–	–
estimated	0.017	–	–	–	–	–	–
rights	0.017	–	–	–	–	–	–

Each topic can be derived from its corresponding bag of words, leading to the following topic categorisation list:

- *Topic 1. Evaluation of effectiveness and benefits of ITS:* This topic deals with the analysis of implications and potential use of ITS including new challenges and opportunities, and the anticipation of user behaviour in the process of designing new ITS technologies.

- *Topic 2. Study of systems and tools supporting decision making and transportation planning:* It includes the development of models for travel demand decision and transportation planning modelling tools, technology integration and databases supporting archived data user services.

- *Topic 3. Automatic vehicle detection systems:* Vehicle detection appears to be one of the most promising areas in traffic surveillance and control. This concept entails the detection of vehicles and extraction of traffic parameters in real-time from images generated by video cameras overlooking a traffic scene.

- *Topic 4. Incident management:* Incident management includes emergency response deployment and rerouting to bypass the affected area, impact of freeway lane closures, and incident detection algorithms.

- *Topic 5. Mobile and wireless communication in ITS:* This topic covers intervehicle communication networks, in-vehicle communications), road-to-vehicle communications, wireless protocols, GPRS and third-generation systems.

- *Topic 6. Processing algorithms for ITS applications:* This topic is devoted to advanced processing algorithms for solving ITS problems, such as path problems in dynamic networks, origin–destination estimation and prediction, and image and video processing algorithms.

- *Topic 7. Advanced traffic management systems (ATMS):* They are focused on the development of ITS to improve safety and quality of service as well as the efficiency of existing roadway utilisation.

- *Topic 8. Traffic simulation:* A traffic simulation system consists of a traffic-flow simulation code, which is able to simulate traffic on a freeway network. It usually considers a microscopic representation (where each individual vehicle is represented) or a macroscopic model capturing traffic dynamics. The purpose of these systems consists of performing traffic-flow simulation for applications like traffic conditions prediction in real-time, traffic control and drivers' guidance, link travel time calculation and signal control strategy. This topic also covers issues like modelling driver behaviour under the influence of external factors.

- *Topic 9. Commercial Vehicles Operation (CVO):* This topic is focused on the impact of ITS on commercial vehicles and fleets for improving transportation safety and efficiency.

- *Topic 10. Advanced vehicles control systems (AVCS):* AVCS are based on systems that provide increased safety and/or control to the driver either by means of improving the information about the driving environment or by actively aiding the driver in the driving task. They include on-board autonomous intelligent cruise control systems, ABS and traction control systems, active suspension systems, vehicle stability systems, in-vehicle collision warning systems, etc.

- *Topic 11. Dynamic route selection algorithms:* Dynamic route selection problems are search problems for finding an optimal route from a starting to a destination point on a road map within a time limit. Since the time to traverse a link will depend upon traffic volume encountered on that link, link times are dynamic.

- *Topic 12. Models for traffic demand forecasting:* Traffic congestion is a major operational problem on ITS. Reducing congestion effects requires developing models that can accurately predict traffic demand.

- *Topic 13. Advanced travellers information systems (ATIS):* The function of ATIS is to assist travellers with planning, perception, analysis and decision making to improve the convenience and efficiency of travel.

The obtained results depict an exhaustive draw of ITS research areas. Some market-specific topics are suppressed compared to previous classification methods described in Section 2, like APTS or ARTS in [6] or emergency management and payment systems in [3–10], but new research topics are meanwhile identified. In fact, up to six new research areas have now been detected from the obtained 13 topics. For instance, specific topics like evaluation of effectiveness and benefits of ITS, mobile and wireless communications in ITS or models for traffic demand forecasting have not been previously established. They have emerged as the result of the incorporation of new technologies, and also due to the necessity of modelling and assessing their social and economical impact. The proposed categorisation should help researchers and practitioners to clarify their position for future work because the obtained 13 topics represent the most recent issues and emergent technologies in the ITS world.

The topic correlation matrix (Table 4) shows that topics are poorly correlated with each other, which means that topics are well defined and their scope is clearly delimited.

Nevertheless, it is impossible to achieve perfectly delimited topics with independent scopes. There is always some degree of overlap. Overlapping can be used to obtain several major paradigms attending to the topic similarity. For this purpose, a multivariate statistical technique like multidimensional scaling was used [33]. This analysis consists of projecting the works on a two-dimensional map, using the data from the correlation matrix as input data. Fig. 3 shows the obtained map. The obtained RSQ coefficient (0.93430) and Kruskal's stress (0.15374) suggest that goodness of fit is very acceptable (the RSQ coefficient is the squared correlation index R^2 that measures the model fit to the data, and its minimum acceptable score is 0.6

Table 4 Topic correlation matrix

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
T1	1.00	0.13	0.10	0.07	−0.05	−0.03	0.06	0.14	0.07	−0.01	−0.10	0.03	0.06
T2	0.13	1.00	0.01	0.00	0.12	−0.07	0.16	0.07	0.16	−0.07	0.03	0.10	0.04
T3	0.10	0.01	1.00	0.03	0.13	0.01	−0.06	0.10	0.07	0.13	−0.02	−0.02	0.08
T4	0.07	0.00	0.03	1.00	−0.05	0.14	−0.05	0.19	−0.02	0.05	−0.02	0.21	0.21
T5	−0.05	0.12	0.13	−0.05	1.00	0.05	0.21	−0.02	0.28	0.05	−0.03	−0.09	−0.08
T6	−0.03	−0.07	0.01	0.14	0.05	1.00	−0.04	0.04	0.01	0.03	0.34	−0.02	0.00
T7	0.06	0.16	−0.06	−0.05	0.21	−0.04	1.00	−0.06	0.12	−0.10	−0.06	−0.02	−0.07
T8	0.14	0.07	0.10	0.19	−0.02	0.04	−0.06	1.00	−0.01	0.08	0.03	0.25	0.15
T9	0.07	0.16	0.07	−0.02	0.28	0.01	0.12	−0.01	1.00	−0.01	0.04	−0.03	0.02
T10	−0.01	−0.07	0.13	0.05	0.05	0.03	−0.10	0.08	−0.01	1.00	−0.04	0.00	−0.10
T11	−0.10	0.03	−0.02	−0.02	−0.03	0.34	−0.06	0.03	0.04	−0.04	1.00	0.04	0.13
T12	0.03	0.10	−0.02	0.21	−0.09	−0.02	−0.02	0.25	−0.03	0.00	0.04	1.00	0.30
T13	0.06	0.04	0.08	0.21	−0.08	0.00	−0.07	0.15	0.02	−0.10	0.13	0.30	1.00

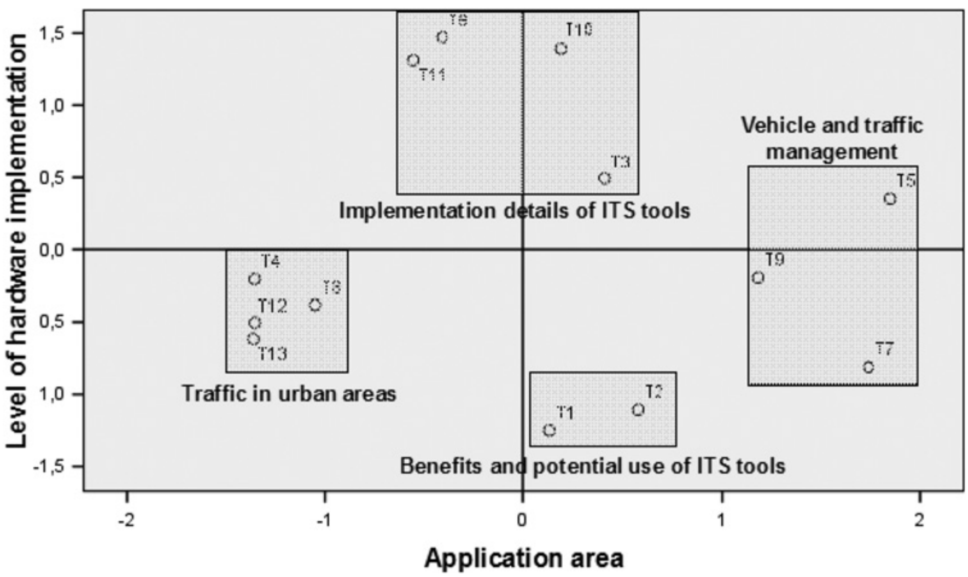

Figure 3 *Multidimensional scaling*

[34]) and that the map exhibits a good approximation of reality.

The proximities of topics on the map show their similarity and suggest a meaning for the axes. The horizontal axis is related to the application area. Topics located to the left part of the map (T4, T8, T12 and T13) are focused on the improvement of traffic in urban area. Topics located on the centre of the map consider ITS tools from two different perspectives, the implementation details represented by topics T3, T6, T10 and T11 at the top of the map and their benefits and potential use represented by T1 and T2 at the bottom of the map. Finally, topics located to the right part of the map are focused on vehicle and traffic management (T5, T6 and T7). The vertical axis is related to the level of hardware implementation, as it can be clearly deduced from topics in the upper half of the map, concerned with the implementation details of ITS tools (T3, T6, T10 and T11) or the mobile and wireless communication possibilities supporting traffic management (T5).

Comparing these results with previous classifications detailed in Section 2, notice that 'Benefits and potential use of ITS tools' were not considered from a technological perspective and the 'implementation details of ITS tools' were not considered from a market area perspective. Consequently, the map of Fig. 3 summarises a more complete vision of ITS.

5 Conclusion

The main contribution of this paper is a global view of ITS, which could be used by future researchers as a state of the art of methods, techniques and application areas. The analysis is intended to offer new perspectives into what is viewed as important to build upon, providing valuable insights into both what research is important and where the field of ITS is heading. The proposed methodology for producing this global view is based on the semantic analysis of keywords and abstracts of papers indexed by the ISI. As a difference to author or journal co-citation methods, semantic analysis is focused on the content of papers, so results are not biased by cites to irrelevant literature or recurrent cited papers. A total of 1147 papers have been analysed covering a large variety of topics related to ITS, including the most recent issues and emergent technologies. As a result of the analysis, 13 paradigms were obtained. Using a multidimensional scaling, they have been represented on a bidimensional map, illustrating the most related paradigms as well as the bridges among them.

Furthermore, this study also defines a starting point for other analyses aimed at a better understanding of the ITS field. This continuous analysis is considered necessary as ITS is an evolutionary field influenced by changes in technology, with changing services and support to end users.

6 Acknowledgments

The authors gratefully acknowledge support provided by the Spanish Ministry of Education and Science (project with reference DPI2007-60128) and the Consejería de Innovación, Ciencia y Empresa (Research Project with reference P07-TIC-02621).

7 References

[1] ANDRISANO O., VERDONE R., NAKAGAWA M.: 'Intelligent transportation systems: the role of third generation mobile radio networks', *IEEE Commun. Mag.*, 2000, **38**, (9), pp. 144–151

[2] TORAL S., VARGAS M., BARRERO F.: 'Embedded multimedia processors for road-traffic parameter estimation', *Computer*, 2009, **42**, (12), pp. 61–68

[3] MCQUEEN B., MCQUEEN J.: 'Intelligent transportation systems architectures' (Artech House, 1999)

[4] NATVIG M.K., WESTERHEIM H.: 'National multimodal travel information – a strategy based on stakeholder involvement and intelligent transportation system architecture', *IET Intell. Transp. Syst.*, 2007, **1**, (2), pp. 102–109

[5] TORAL S., VARGAS M., BARRERO F., ORTEGA M.G.: 'Improved sigma-delta background estimation for vehicle detection', *Electron. Lett.*, 2009, **45**, (1), pp. 32–34

[6] FIGUEIREDO L., JESUS I., MACHADO J.A.T., FERREIRA J.R., MARTINS DE CARVALHO J.L.: 'Towards the development of intelligent transportation systems'. Proc. 2001 IEEE Intelligent Transportation Systems, Oakland, CA, USA, 2001, pp. 1206–1211

[7] WANG F.-Y., MIRCHANDANI P.B., ZHENG N.: 'Advances and trends in research and development of intelligent transportation systems: an introduction to the special issue', *IEEE Trans. Intell. Transp. Syst.*, 2004, **5**, (4), pp. 222–223

[8] ZHENG N., KAWASHIMA H.: 'Advances and trends in research and development of vehicular electronics and safety: special section', *IEEE Trans. Intell. Transp. Syst.*, 2007, **8**, (1), pp. 106–107

[9] Artemis Strategic Research Agenda Working Group: 'Strategic research agenda' (Ideo Ltd, 2006, 1st edn. edn.)

[10] US Department of Transportation, Federal Highway Administration: 'Intelligent transportation systems benefits, costs and lessons learned' (Mitretek Systems, 2005)

[11] http://isi3.isiknowledge.com, accessed October 2009

[12] WOHLIN C.: 'An analysis of the most cited articles in software engineering journals – 2001', *Inf. Soft. Technol.*, 2008, **5**, pp. 3–9

[13] DING Y., CHOWDHURY G.G., FOO S.: 'Bibliometric cartography of information retrieval research by using co-word analysis', *Inf. Process. Manage.*, 2001, **37**, pp. 817–842

[14] SMALL H.: 'Co-citation in the scientific literature: a new measure of the relationship between two documents', *Essays Info. Sci.*, 1974, **2**, pp. 28–31

[15] CHENG C.H., KUMAR A., MOTWANI J.G., REISMAN A., MADAN M.S.: 'A Citation analysis of the technology innovation management journals', *IEEE Trans. Eng. Manage.*, 1999, **46**, (1), pp. 4–13

[16] CHEN C., PAUL R.J.: 'Visualizing a knowledge domain's intellectual structure', *Computer*, 2001, **34**, (3), pp. 65–71

[17] CALLON M., COURTIAL J.P., LAVILLE F.: 'Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry', *Scientometrics*, 1991, **22**, (1), pp. 153–205

[18] SALTO G., MCGILL M.J.: 'An introduction to modern information retrieval' (McGraw-Hill, 1983)

[19] CAI D., HE X., HAN J.: 'Document clustering using locality preserving indexing', *IEEE Trans. Knowl. Data Eng.*, 2005, **17**, (12), pp. 1624–1637

[20] SHI J., MALIK J.: 'Normalized Cuts and Image Segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (8), pp. 888–905

[21] NG A.Y., JORDAN M., WEISS Y.: 'On spectral clustering: analysis and an algorithm', in 'Advances in neural information processing systems 14' (MIT Press, 2001), pp. 849–856

[22] ZHA H., DING C., GU M., HE X., SIMON H.: 'Spectral relaxation for k-means clustering', in 'Advances in neural information processing systems 14' (MIT Press, 2001), pp. 1057–1064

[23] XU W., LIU X., GONG Y.: 'Document clustering based on non-negative matrix factorization'. Proc. Int. Conf. on Research and Development in Information Retrieval, 2003, pp. 267–273

[24] XU W., GONG Y.: 'Document clustering by concept factorization'. Proc. Int. Conf. on Research and Development Information Retrieval, 2004, pp. 202–209

[25] DEERWESTER S., DUMAIS S.T., FURNAS G.W., LANDAUER T.K., HARSHMAN R.: 'Indexing by latent semantic analysis', *J. Am. Soc. Inf. Sci.*, 1990, **41**, (6), pp. 391–407

[26] GRIFFITHS T.L., STEYVERS M.: 'Finding scientific topics'. Proc. National Academy of Sciences, USA, 2004, pp. 5228–5235

[27] BLEI D.M., NG A.Y., JORDAN M.I.: 'Latent Dirichlet allocation', *J. Mach. Learn. Res.*, 2003, **3**, pp. 993–1022

[28] HOFMANN T.: 'Unsupervised learning by probabilistic latent semantic analysis', *Mach. Learn. J.*, 2001, **42**, (1), pp. 177–196

[29] SUGUMARAN V., STOREY V.C.: 'Ontologies for conceptual modeling: their creation, use and management', *Data Knowl. Eng.*, 2002, **42**, pp. 251–271

[30] SHAFIEI M.M., MILIOS E.E.: 'Latent Dirichlet co-clustering'. Sixth Int. Conf. on Data Mining, ICDM'06, 2006, pp. 542–551

[31] MANNING C.D., SCHUTZE H.: 'Foundations of statistical natural language processing' (MIT Press, 1999)

[32] KOSKELA M., SMEATON A.F., LAAKSONEN J.: 'Measuring concept similarities in multimedia ontologies: analysis and evaluations', *IEEE Trans. Multimedia*, 2007, **9**, (5), pp. 912–922

[33] ROWLANDS I.: 'Patterns of author co-citation in information policy: evidence of social, collaborative and cognitive structure', *Scientometrics*, 1999, **44**, (3), pp. 533–546

[34] RENCHER A.C.: 'Methods of multivariate analysis' (A John Wiley & Sons, Inc. Publication, 2002, 2nd edn.)

8 Appendix: probability theory applied to document analysis

Representing the content of words and documents with probabilistic topics has one distinct advantage over the purely spatial representation of LSI. Each topic is individually interpretable, providing a probability distribution over a word that picks out a coherent cluster of correlated terms. Words are the only observable variables and they implicitly reflect the latent structure. Each topic is on the basis of the random variable θ that is sampled from a Dirichlet distribution $p(\theta;\ \alpha)$, where α is a hyper parameter. The topic z conditioned on θ and the word w conditioned on the topic and on ϕ (word distribution over topics) are sampled from multinomial distributions $p(z_n|\theta)$ and $p(w_n|z_n;\ \phi)$, respectively. The probability of a document can be computed as

$$p(w) = \int_\theta \left[\prod_{n=1}^{N} \sum_{z_n=1}^{k} p(w_n|z_n;\ \phi) p(z_n|\theta) \right] p(\theta;\ \alpha)\, \mathrm{d}\theta \qquad (2)$$

In [28], the probabilistic topic approach to document modelling in his probabilistic latent semantic indexing method (pLSI) is introduced. This model was extended by introducing a Dirichlet prior, calling the resulting generative model LDA [27]. In this case, α and ϕ are learnt by variational inference to maximise the log likelihood of the data. An easier implementation has been proposed in [26] introducing a simple modification to the model. A Dirichlet prior is introduced on the parameter ϕ, with hyper parameter β. Despite this modification, computation of the conditional probability $p(z|w)$ is still unmanageable. As a solution, they propose to approximate it by Gibbs sampling based on the following distribution:

$$p(z_i = j|z_{i-1}, w) \propto \frac{n_{-ij}^{(w_i)} + \beta}{n_{-i,j}^{(\cdot)} + w\beta} \frac{n_{-i,j}^{(d_j)} + \alpha}{n_{-i}^{(d_j)} + T\alpha} \tag{3}$$

This distribution represents the probability that word w_i should be assigned to topic j given all other assignments z_{-i}. The quantities $n_{-i,j}^{(w_i)}$ and $n_{-i,j}^{(\cdot)}$ represent the number of times the word w_i has already been assigned to topic j and the total number of words assigned to topic j, respectively. The quantities $n_{-i,j}^{(d_i)}$ and $n_{-i}^{(d_i)}$ represent the number of times the word w_i in the document d_i has already been assigned to topic j and the number of words in document d_i that are assigned to topic j. The hyper parameters α and β are computed using the method described in [26], that is, $\beta = 0.01$ and $\alpha = 50/T$.

Considering T topics, the probability of the ith word in a given document can be written as

$$P(w_i) = \sum_{j=1}^{T} P(w_i|z_i = j)P(z_i = j) \tag{4}$$

where z_i is a latent variable indicating the topic from which the ith word was drawn, and $P(w_i|z_i = j)$ is the probability of word w_i under the jth topic. $P(z_i = j)$ gives the probability of choosing a word from topic j in the current document, which will vary across different documents. Intuitively, $P(w|z)$ indicates which words are important to a topic, whereas $P(z)$ is the prevalence of those topics within a document. LDA combines (2) with a prior probability distribution to provide a complete generative model for documents.