Energy Conversion and Management 153 (2017) 346–361

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Decomposing core energy factor structure of U.S. commercial buildings through clustering around latent variables with Random Forest on large-scale mixed data

Endong Wang

Construction Management, The University of Tennessee, EMCS 326B, Chattanooga, TN 37403, USA

ARTICLE INFO

Keywords:
Collinearity
Commercial buildings
Energy factor structure
Principal energy factors
Random Forest
Variable Clustering

ABSTRACT

Accurate selection of explanatory factors is critical for precise quantitative energy analysis (e.g. benchmarking and predicting) to support the sustainability strategy in commercial building sector. Nevertheless, the generic guiding information on factor selection lacks. This paper addresses the research gap to decompose building energy factor structure (i.e. the interaction structure among the factors which affect building energy performance) particularly at a nation level for factor selection. Specifically, an iterative approach is developed by integrating the technical strengths of Variable Clustering and Random Forest to remove collinearity, redundancy and nonrelevance. Based on a comprehensive source database extracted from a multiframe country-wide survey, the core energy factor space is revealed for 2779 commercial buildings in the U.S. In particular, 36 principal factors are identified to reliably explain building energy efficiency variations. These factors are of sufficient independence and heterogeneity which may benefit the development of parsimonious energy modeling frameworks. The robustness of the deciphered factor structure and the representativeness of the recognized critical factors are numerically validated. These acquired results can be useful for informed decision analysis and rational policy design in commercial building sector with a lighter data burden.

1. Introduction

Substantially reducing energy usage in commercial buildings through aimful efficiency initiatives is of paramount significance to the overall success of sustainability strategy in the U.S. given its grand share in the nationwide total building energy consumption. Per the latest report from the U.S. Energy Information Administration (EIA) [1], commercial building sector consumed 4.29 quintillion joules of primary energy, clarifying more than 38% of the aggregated amount attributed to all the extant U.S. buildings. In viewing of this fact, a wide range of conservation ordinances, retrofitting plans, renewable energy incentives at either local or federal level are strategically implemented [2,3]. Regardless of specific objectives and execution procedures, these efficiency programs often commonly require quantitative energy analysis on fault detection, usage projection, etc., to lay solid information bases for decision making or policy design. For instance, robust quantitative benchmarks [4–5] are usually desired for evaluating energy performance to locate the inefficiencies in commercial buildings for drawing sensible decisions during thermal retrofitting.

Statistical models which generate energy performance information essentially by the numerical analysis of available data counting on the scoped energy variables, appear more attractive in commercial sector over the thermodynamics-based approach [3,6–9]. This turns particularly true for a large-scale building energy investigation which typically occurs in energy efficiency domain, because statistical methods show more capability of accounting for the diverse energy aspects of occupants, structure, environment and geography [3]. From this angle, an extensive number of statistical procedures have been proposed for divergent efficiency upgrading purposes with varying complexities. These procedures typically range from the simple univariate statistical analysis (e.g. energy use intensity (EUI) based single-factor energy benchmarking) [10], to the multi-dimensional machine learning algorithms [9,11] (e.g., neural networks for forecasting) [7]. While these quantitative frameworks often have unequal purposes and complexities, the explanatory variables which characterize relevant factors and potentially affect energy performance are often commonly critical to them.

Nevertheless, the pool of energy factors that have been reported to possibly influence the efficiency of commercial buildings is extremely large due to the complicated interactions (either physical or non-physical) among energy systems and the surrounding built environments. Further, these documented factors are associated with distinct energy perspectives (e.g. the degree days regarding climate conditions,

E-mail address: Endong-Wang@utc.edu.

http://dx.doi.org/10.1016/j.enconman.2017.10.020
Received 9 June 2017; Received in revised form 3 October 2017; Accepted 7 October 2017

Nomenclature

Abbreviations

ACW	individual room air conditioners
AMI	advanced smart metering
ANN	artificial neutral network
ATT	attic
AWN	external overhangs or awnings
BLD	building shape
BOI	boilers inside the building
BUL	incandescent bulbs
CAP	electricity generation capability
CEN	census division
CFL	compact fluorescent bulbs
CHI	central chillers inside the building
COL	percent cooled
COO	energy used for cooking
COP	photocopiers
DAL	percent daylight
DAY	daylight harvesting
DIM	multi-level lighting or dimming
DRL	demand responsive lighting
ELH	electricity used for main heating
ELK	electricity for cooking
ELM	electricity for manufacturing
ELW	electricity for water heating
EMC	building automation system
ENR	energy management plan
EQG	equal glass on all sides
EVA	evaporative or swamp coolers
FAC	of a multibuilding complex
FAX	FAX machines
FKU	fuel oil/diesel/kerosene used
FLC	floor to ceiling height
FLU	fluorescent bulbs
FUR	furnaces that heat air directly
GEN	energy for electricity generation
GLS	percent of exterior glass
GOV	government owned
HAL	halogen bulbs
HEA	percent heated
HID	high intensity discharge bulbs
HTC	heat pumps for cooling
HTP	heat pumps for heating
HTV	heating ventilation
LAP	number of laptops category
LAT	number of laptops
LED	light-emitting diode bulbs
LNH	lit off hours category
LOH	lit when open category
LTE	percent of exterior lighted
LTN	percent lit off hours
LTO	percent lit when open
MAI	main heating equipment
MAL	main cooling equipment
MAN	energy used for manufacturing
MAT	regular HVAC maintenance
MLR	multiple linear regression
MLT	multiple monitors
MON	months in use
NFO	number of floors
NGU	natural gas used
NOA	number of businesses category
NOC	number of businesses
NWK	number of employees category
NWL	main cooling replaced
NWM	main heating replaced
NWR	number of employees
OCS	occupancy sensors
ONE	single activity in building
OOB	out-of-bag
OP4	open 24 h a day
OPE	open on weekend
OPF	open during week
OWC	owner occupied or leased to tenants
OWN	building owner
OWR	owner operate and maintain systems
PBA	principal building activity
PBS	building activity
PCC	number of computers category
PCM	computers used
PCT	number of computers
PKG	packaged heating units
PKL	packaged air conditioning units
PKT	lighted parking area
PLG	plug load control
PRU	bottled gas/liquid petroleum/propane
RCA	residential type central air conditioning
RDC	cooling reduced during 24-h period
RDH	heating reduced during 24-h period
RDL	lighting reduced during off hours
REB	plumbing system upgrade
REC	HVAC equipment upgrade
REE	electrical upgrade
REF	roof replacement
REG	census region
REL	reflective window glass
REN	window replacement
RER	structural upgrade
RES	insulation upgrade
RET	lighting upgrade
REV	renovated
REW	exterior wall replacement
RFC	number of compact refrigerators
RFG	count of closed-case refrigeration units
RFI	number of walk-in units
RFN	roof material
RFO	cool roof materials
RFP	number of open-case refrigeration units
RFQ	refrigeration
RFS	number of residential refrigerators
RFT	roof tilt
RFV	count of vending machines
RGI	number of ice makers
RGS	cash registers
RGT	number of cash registers
RS	random sampling
SCH	light scheduling
SEN	number of servers
SER	number of servers category
SKY	skylights or atriums
SQF	building area category
SQT	building area
SVM	support vector machine
TIN	tinted window glass
TVV	television or video displays
WAT	energy used for water heating
WIN	window glass type
WKC	weekly hours category